

**Федеральное государственное бюджетное учреждение науки
Институт эволюционной физиологии и биохимии им. И.М.
Сеченова Российской академии наук (ИЭФБН РАН)**

На правах рукописи



Рузин Игорь Мартынович

Математические методы популяционной генетики

и их применение к вирусологии

1.5.8. Математическая биология, биоинформатика

Диссертация на соискание ученой степени *доктора физико-математических наук*

Санкт-Петербург – 2024

Содержание

Введение	5
Глава 1. Обзор литературы	12
1.1 Связь с остальными главами	12
1.2 Определение действующих факторов и оценка основных параметров эволюции из геномных данных	12
1.2.1 Механизм эволюции ВИЧ и оценка среднего коэффициента отбора	12
1.2.1 Оценка эффективного размера популяции	14
1.2.3 Оценка вероятности рекомбинации	15
1.3 Многолокусные модели эволюции популяций бесполок организмов и вирусов	16
1.3.1 Многолокусная эволюция с фиксированным коэффициентом отбора	16
1.3.2 Многолокусная эволюция с распределенным коэффициентом отбора	18
1.4 Многолокусная эволюция частично половых популяций	19
1.4.1 Эволюция с рекомбинацией и начальным генетическим разнообразием	19
1.4.2 Эффекты общей родословной (филогенеза) на эволюцию с рекомбинацией	19
1.5 Определение адаптивного ландшафта из геномных данных	20
1.5.1 Универсальный эволюционный след эпистаза	21
1.5.2 Обнаружение эпистатических пар локусов в одной отдельно взятой популяции: миссия невыполнима	22
1.5.3 Обнаружение эпистатических пар локусов методом трехлокусной корреляции	22
1.5.4 Оценка коэффициентов отбора отдельных локусов из геномных данных	23
1.6 Эволюционная роль признака	24
1.6.1 СПИД, как следствие адаптации ВИЧ к хозяину	24
1.6.2 Эволюционная роль латентного состояния ВИЧ	26
1.6.3 Роль рекомбинации в адаптации и оптимальная вероятность мутаций в системе полиовирус-мышь	28
1.7 Наперегонки с враждебным биологическим видом	29
1.7.1 Эволюция респираторного вируса в популяции под давлением гуморального иммунитета	29
1.7.2 Эволюция ВИЧ в хозяине под давлением клеточного иммунитета	30
1.7.3 Эволюционная стабильность ВИЧ в хозяине в присутствии дефектных интерферирующих частиц	32

1.7.4 Эволюционный конфликт ВИЧ между уровнями хозяина и популяции в присутствии дефектных интерферирующих частиц	33
Глава 2. Определение действующих факторов и оценка основных параметров эволюции из геномных данных	33
2.1 Механизм эволюции ВИЧ и оценка среднего коэффициента отбора	34
2.2 Оценка эффективного размера популяции	48
2.3 Оценка вероятности рекомбинации ВИЧ	59
Глава 3. Многолокусные модели эволюции популяций бесполок организмов и вирусов	69
3.1 Многолокусная эволюция с фиксированным коэффициентом отбора	70
3.2 Многолокусная эволюция с распределенным коэффициентом отбора	85
Глава 4. Многолокусная эволюция частично половой популяции	93
4.1 Эволюция с рекомбинацией и начальным генетическим разнообразием	94
4.2 Эффекты общей родословной (филогенеза) на эволюцию с рекомбинацией	102
Глава 5. Определение адаптивного ландшафта из геномных данных	116
5.1 Универсальный эволюционный след эпистаза	116
5.2 Обнаружение эпистатических пар локусов в одной отдельно взятой популяции: миссия невыполнима	128
5.3 Обнаружение эпистатических пар локусов методом трехлокусной корреляции	137
5.4 Оценка коэффициентов отбора отдельных локусов из геномных данных	147
Глава 6. Эволюционная роль признака	159
6.1 СПИД, как следствие адаптации ВИЧ к хозяину	159
6.2 Эволюционная роль латентного состояния ВИЧ.....	159
6.3 Роль рекомбинации в адаптации и оптимальная частота мутаций в системе вирус полиомиелита - мышь	185
Глава 7. Наперегонки с враждебным биологическим видом	197
7.1 Эволюция респираторного вируса в популяции под давлением гуморального иммунитета	197
7.2 Эволюция ВИЧ в хозяине под давлением клеточного иммунитета	210
7.3 Эволюционная стабильность ВИЧ в хозяине в присутствии дефектных интерференционных частиц	222
7.4 Эволюционный конфликт ВИЧ между уровнями хозяина и популяции в присутствии дефектных интерферирующих частиц	235

Заключение	248
Выводы	250
Указатель терминологии	251
Библиография	255

Введение

Актуальность темы диссертации

Многие вирусы (Вирус Иммунодефицита Человека (ВИЧ), вирус гепатита С, вирус гриппа, SARS-CoV-2) эволюционируют чрезвычайно быстро и, как следствие, обладают огромным генетическим разнообразием. Быстрая эволюция позволяет им уклоняться от естественного иммунного ответа, делать вакцины неэффективными и развивать устойчивость к противовирусным препаратам. Когда в 1980-х годах разразилась пандемия ВИЧ, ученые начали искать противовирусное лечение. Врачи прописывали инфицированным пациентам один антиретровирусный ингибитор за другим, но вирус демонстрировал невероятную генетическую гибкость, вырабатывая устойчивость ко всем лекарствам в течение нескольких дней или недель после начала терапии. Все это сопровождалось самой высокой скоростью эволюции в хозяине и популяции, когда-либо наблюдавшейся у вируса, даже в пациентах не получавших лекарств (рисунок 1.1).

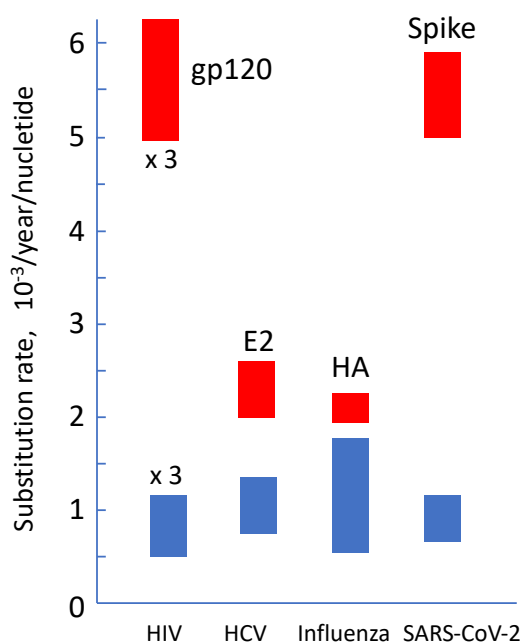


Рисунок 1.1. Скорость замещений в вирусах. Синие прямоугольники показывают интервалы медианных значений для наиболее быстро и наиболее медленно развивающихся подтипов ВИЧ, вируса гепатита С, вируса гриппа и SARS-CoV-2 для всего генома. Красные прямоугольники соответствуют внешним белкам вируса, на которые нацелены нейтрализующие антитела. Чтобы уместиться в график, данные для ВИЧ показаны в масштабе 1/3 от их фактических значений. Из обзора [1], входящего в список работ диссертации.

Причина такого беспрецедентного поведения, в то время, была неизвестна. Динамика размножения ВИЧ в хозяине и клетке была изучена плохо, а области вирусологии и популяционной генетики почти не пересекались. Стала очевидной необходимость применения

передовых методов математического моделирования из областей иммунологии, эпидемиологии, и генетики популяций. Хотя математические иммунологи и эпидемиологи быстро ответили на этот вызов, популяционные генетики не торопились, вероятно, потому, что количество эволюционирующих мест в ВИЧ геноме (локусов) слишком велико, а их эволюция слишком переплетена и сложна, чтобы классические простые модели могли бы быть достаточны.

В 1995 году появилась важная идея [2] о причине неудач антиретровирусной терапии, которая потребовала математического подтверждения. Для этого потребовалось использование методов популяционной генетики, которые ранее не применялись к вирусологии. В отличие от эволюционной биологии и областей знания, посвященных эволюции биосферы и происхождению видов, научная область «генетика популяций» рассматривает популяцию организмов (вирусов) одного вида с последовательностями ДНК (РНК), которые отличаются между индивидуальными организмами (вирусами). Предметом изучения является динамика изменения популяции во времени, то есть изменение этого набора геномов между последовательными поколениями, которое можно измерять экспериментально и предсказывать, в вероятностном смысле, с помощью математических моделей. Развитию математических методов популяционной генетики и ее применению в вирусологии и посвящена данная диссертация.

Степень разработанности темы исследований

На момент начала работы, модели стохастической эволюции, учитывающие естественный отбор, были хорошо развиты только для моделей геномов с одним и двумя локусами. К вирусологии они не применялись совсем. Хотя динамика ВИЧ в пациентах была понята, как стационарный процесс с быстрой ротацией инфицированных клеток, основные биологические факторы и параметры эволюции ВИЧ в нелеченных людях были неизвестны. Также неизвестны были эволюционная выгода латентности ВИЧ в отсутствие лекарств и причина постепенного истощения CD4 клеток приводящего к СПИДу. Хотя антигенная эволюция вируса гриппа на уровне популяции была неплохо изучена эмпирически, понимание того, как эпидемиологические, иммунологические, и вирусологические параметры вируса гриппа взаимодействуют, чтобы привести к определенной скорости эволюции, отсутствовало. Предпринимались попытки изучения эволюционной стабильности ВИЧ в присутствии его дефектных интерферирующих частиц. Однако, эти ранние модели делали произвольные допущения и не опирались на правильную вирусологию и генетику ВИЧ. Базы данных уже содержали огромное количество данных секвенирования геномов разных организмов, и десятки методов были разработаны для оценки эпистатических взаимодействий и коэффициентов отбора

из геномных данных. Но их предсказания плохо подтверждались прямыми экспериментами, из-за огромного шума в данных непонятного происхождения. Более того, математическая теория многолокусной эволюции полностью отсутствовала, из-за трудности описания стохастической динамики экспоненциально большого числа возможных последовательностей. Более подробно, состояние области и мотивация конкретных задач обсуждаются в *Обзоре литературы*. Базовые понятия генетики популяций даны в § 2.1.

Цели диссертации

1. Адаптировать модели популяционной генетики к вирусологии, чтобы определить факторы и параметры генетической эволюции патогенных вирусов из геномных данных (*главы 2, 6, 7*).
2. Разработать новые модели и математические методы популяционной генетики для учета многолокусной природы генома (*главы 3-5*).

Задачи диссертации

1. Оценить основные параметры популяции вируса у людей зараженных ВИЧ и пока нелеченых: эффективный размер популяции, вероятность рекомбинации, средний коэффициент отбора.
2. Разработать и решить аналитически реалистичную модель с большим количеством локусов (эволюционирующих оснований) в геноме. Дело в том, что эволюция разных мест генома не является независимой из-за их общей родословной, филогенеза. Этот эффект, который называется «генетическое сцепление», крайне осложняет математическое предсказание эволюции.
3. Предсказать, в общем виде, влияние генетической рекомбинации на адаптацию вируса для генома с большим числом локусов.
4. Решить проблему низкой надежности существующих методов определения адаптивного ландшафта из геномных данных.
5. Объяснить эволюционную выгоду латентного состояния ВИЧ для данного вируса.
6. Разработать модель, объясняющую медленное обеднение CD4 клеток в ВИЧ инфекции, и выразить время до наступления симптомов СПИД через параметры пациента в общем виде.
7. Выразить скорость эволюции антигенных участков вируса гриппа через иммунологические, эпидемиологические и вирусологические параметры.

8. Предсказать, разовьются ли вирусы, такие как ВИЧ и вирус полиомиелита, устойчивость к дефектным интерферирующим частицам, если те будут использованы в качестве анти-вирусной терапии.

Новизна

Впервые в вирусологии были использованы стохастические модели популяционной генетики с учетом естественного отбора. Средний коэффициент отбора, эффективная численность популяции и вероятность рекомбинации при типичной ВИЧ-инфекции были впервые оценены из геномных данных. Впервые биологически реалистичная модель многолокусной бесполовой эволюции, учитывающая эффекты генетического сцепления, была разработана, решена аналитически и протестирована с помощью компьютерной модели. Для описания многолокусной эволюции, разработан новый метод бегущей полудетерминистической волны со стохастическим фронтом. Впервые, в рамках многолокусной модели, показано, что включение даже редкой генетической рекомбинации в многолокусную модель частично компенсирует эффекты генетического сцепления, ускоряя адаптацию во много раз. Эволюция вируса гриппа, вызванная иммунным ответом населения, была впервые сведена аналитически к модели многолокусной эволюции, и ее скорость выражена, в общем виде, через иммунологические, вирусологические и эпидемиологические параметры. Разработан первый высокоточный метод определения адаптивного ландшафта из геномных данных, на фоне сильного шума из-за эффектов генетического сцепления. Предложено первое эволюционное преимущество латентности ВИЧ, подкрепленное моделью и данными. Предложена первая замкнутая модель прогрессирования ВИЧ инфекции к СПИД.

Теоретическая и практическая значимость

Полученные оценки базовых эволюционных параметров ВИЧ, такие как размер популяции вируса, оказались полезными при разработке коктейлей вирусных ингибиторов. А именно, быстрая эволюция штаммов ВИЧ устойчивых к лекарствам была объяснена из наличия этих мутаций еще до начала терапии. Создание тройных коктейлей решило эту проблему. Создана новая научная область: мультилокусная популяционная генетика. Разработан надежный метод оценки адаптивного ландшафта, применимый к любой адаптирующейся популяции, а не только к вирусам. Разработан метод прогноза эволюции респираторных вирусов, необходимый для разработки и обновления вакцин. Стало понятным, почему ВИЧ-инфекция в людях прогрессирует к СПИД, а в естественных хозяевах вируса, обезьянах — нет. Получены обнадеживающие результаты для лечения вирусов дефектными интерферирующими частицами.

Методы

Стандартные: Линейные и нелинейные обыкновенных дифференциальные уравнения. Уравнение Колмогорова. Численное моделирование методом Монте-Карло и Гиллеспи. Марковские цепочки (ветвящиеся процессы), включая процесс Райта-Фишера. Численное и аналитическое решение интегро-дифференциальных уравнений. Многоуровневое моделирование.

Оригинальные: Полудетерминистический подход для бегущей волны со стохастическим фронтом. Обобщение данного метода при наличии рекомбинации, с учетом общей родословной геномов. Метод определения коэффициентов отбора локусов из геномных данных, на фоне эффектов генетического сцепления. Метод многолокусных корреляторов для измерения эпистатических взаимодействий на фоне эффектов генетического сцепления. Метод моделирования на пересечении иммунологии, эпидемиологии и популяционной генетики для прогнозирования эволюции респираторных вирусов.

Положения, выносимые на защиту

1. Эффективный размер популяции у среднего нелеченного ВИЧ-инфицированного пациента составляет не менее миллиона активно-инфицированных клеток, поэтому естественный отбор в хозяине сильнее стохастических факторов, и невыгодные мутации, обеспечивающие устойчивость к монотерапии, существуют в популяции вируса уже до терапии.
2. Высокое разнообразие и скорость эволюции ВИЧ в хронической инфекции у нелеченных пациентов обусловлены мутациями ускользания в иммунологически важных областях и мутациями, возникающими для их компенсации. Порядок мутаций ускользания во времени определяется их положением в плоскости «стоимость-выгода». Средний коэффициент отбора компенсирующей мутации несколько меньше одного процента.
3. В популяции бесполок организмов или вирусов эффекты генетического сцепления вызывают потерю полезных аллелей и замедляют адаптацию во много раз пропорционально длине генома. Существует оптимальная вероятность мутации, при которой скорость адаптации оптимальна из за баланса между накоплением выгодных и невыгодных аллелей. Это предсказание позже подтверждено экспериментально на вирусе полиомиелита в мышах.
4. Частично компенсируя эффекты генетического сцепления, даже редкая рекомбинация позволяет ускорить адаптацию на порядки. В хронической нелеченной ВИЧ-инфекции вероятность рекомбинации примерно 1%. Наличие общих предков у гомологичных локусов

снижает эффективность рекомбинации из-за корреляций между геномами. В достаточно длинном геноме, даже очень частая рекомбинация не делает эволюцию далеко расположенных локусов независимой.

5. Низкая надежность существующих методов обнаружения эпистаза в образцах ДНК вызвана эффектами генетического сцепления. Усреднение по небольшому числу популяций и использование дополнительных локусов позволяет исключить этот эффект и достичь идеальной надежности обнаружения, по крайней мере, для геномов меньше сотни локусов.
6. Причиной СПИД, парадоксально, является адаптация ВИЧ к индивидууму. Эволюционное преимущество латентного состояния ВИЧ заключается в усилении передачи вируса между людьми.
7. Давление отбора, создаваемое иммунной памятью населения на респираторный вирус, выражается, в общем виде, через только два параметра: базовое число передачи в наивном населении и генетическое расстояние перекрестной реактивности антител.
8. В отсутствие эпистаза вирус не может развить устойчивость к своей дефектной интерференционной частице ни на уровне хозяина, ни на уровне популяции.

Апробация и публикации

По теме диссертации опубликовано 36 работ, в том числе 2 монографии [3, 4] и 34 статьи в рецензируемых журналах. Эти статьи состоят из 28 оригинальных статей [5-32], 4 обзоров или перспектив [1, 33-35] и 2 комментариев [36, 37]. И.М. Рузин был единственным, первым, последним автором или автором-корреспондентом в 30 из 36 этих работ.

Большинство этих статей опубликовано в журналах первой четверти, таких как *Cell*, *Cell Host & Microbe*, *Proceedings of the National Academy of Sciences USA*, *PLoS Pathogens*, *PLoS Computational Biology*, *Scientific Reports*, *Journal of Virology*, *Molecular Biology and Evolution*, и *Genetics*.

Результаты представлены на 40 международных конференциях, включая 12 важных докладов, перечисленных ниже:

- 2023** Лекция об антигенной эволюции вирусов в населении и эволюционных последствиях вакцинации от КОВИДа. NORDITA Program “Unifying the epidemiological and evolutionary dynamics of pathogens”, Stockholm, Sweden.
- 2023** Доклад о измерении адаптивного ландшафта из геномных данных, NORDITA Program “Unifying the epidemiological and evolutionary dynamics of pathogens”
- 2022** Доклад об эволюционной роли латентности ВИЧ, QLife Winter School “Quantitative Viral Dynamics Across Scales”, Paris, France

-
- 2019** Доклад об эволюции вируса гриппа, NORDITA program "From Molecular Basis to Predictability and Control of Evolution", Stockholm, Sweden.
- 2018** Доклад об обнаружении эпистаза в геномных данных, 3d IFOM Symposium "Evolution, Resistance, and Cancer", IFOM, Milan, Italy.
- 2015** Доклад об эволюционных моделях вирусов, 5th Interdisciplinary Symposium "Genomics and Microbial Communities", Institut de Biologie Paris-Seine, UPMC, Paris
- 2014** Доклад об эволюционной роли латентности ВИЧ, "Evolution of Drug Resistance", Kavli Institute of Theoretical Physics, University of California at Santa Barbara, CA
- 2014** Доклад об эволюционной роли латентности ВИЧ, "Workshop on From Within-Host Dynamics to the Epidemiology of Infectious Disease", Institute of Mathematical Biosciences, Columbus, Ohio
- 2011:** Доклад о моделях многолокусной эволюции, *Workshop on Microbial Evolution*, Kavli Institute for Theoretical Physics, University of California at Santa Barbara, Santa Barbara, CA, USA
- 2009:** Доклад об эволюции и иммунологии ВИЧ, *International 3rd Eastern European and Central Asia AIDS Conference*, Moscow, Russia
- 2006:** Доклад о моделях многолокусной эволюции, *Workshop on Mathematical Population Genetics*, Edinburgh, UK
- 1998:** Доклад об эволюции и иммунологии ВИЧ, *Meeting on HIV Dynamics and Evolution*, University of California at San Diego

Структура диссертации

Диссертация состоит из введения, главы с обзором литературы, шести глав с результатами, заключения, выводов, индекса терминов и списка литературы, включающего 498 различных источников. Работа изложена на 281 странице машинописного текста, иллюстрирована 13 таблицами и 105 рисунками.

Глава I. Обзор литературы

1.1 Связь с остальными главами

В целях удобства, обзор литературы разбит на разделы и подразделы по темам, с нумерацией вида 1.x и 1.x.y, где x и x.y — номера соответствующих глав и подразделов в *главах 2-7* содержащих результаты диссертации.

1.2 Определение действующих факторов и оценка основных параметров эволюции из геномных данных

1.2.1 Механизм эволюции ВИЧ и оценка среднего коэффициента отбора

По сравнению с другими вирусами, вирус иммунодефицита человека (ВИЧ) демонстрирует очень высокую скорость эволюции и высокий уровень генетического разнообразия [38]. У инфицированных людей ВИЧ вызывает персистентную инфекцию с сотнями генетически разнообразных геномных участков, которые различаются между пациентами. Мутации, препятствующие распознаванию вируса адаптивным иммунным ответом, мешают эффективности вакцин [39-42], в то время как другие мутации делают вирус устойчивым к ингибиторам репликации [43, 44].

С другой стороны, геномные данные содержат важную информацию о биологических факторах, действующих на вирус. О наличии функционального иммунного ответа, помимо иммунологических экспериментов, можно судить по данным генетического разнообразия. Доминирование несинонимичных замен в белке оболочки вируса свидетельствует о наличии отбора на разнообразие, действующего на этот поверхностный белок [45-47], вызванного гуморальным ответом и адаптацией ВИЧ к различным типам клеток.

На вирусы действуют целый ряд факторов эволюции, которые будут подробно обсуждаться в этой диссертации. Одним фактором является мутация, то есть случайная ошибка копирования генома. Другим важным фактором является естественный отбор различных типов, из которых главными являются два. Один из них, который называется «очищающий» или «отрицательный» отбор существует благодаря тому, что главный генетический вариант имеет большее среднее число потомков, чем миноритарный вариант, то есть он более «приспособленный». Среднее количество потомков способных оставить потомков, в популяционной генетике, называется «Мальтузианской приспособленностью» или просто

«приспособленностью». Если бы отрицательный отбор был единственным эволюционным фактором, действующим на популяцию, популяция в конечном итоге стала бы генетически однородной (мономорфной) и состоящей только из лучшего варианта: выживает наиболее приспособленный. Если более приспособленного варианта в начале меньше, и его количество увеличивается со временем, то говорят о «положительном отборе».

Принципиально другой тип естественного отбора — это отбор на разнообразие, который приводит к усилению генетического разнообразия. Он бывает, в свою очередь, двух видов. Первый возникает из-за изменения внешних условий со временем, таких как иммунный ответ организма (такой случай рассмотрен в §§ 7.1 и 7.2). Другой вид отбора на разнообразие — это эффект экологической ниши, когда разные геномные последовательности вируса лучше размножаются в разных клеточных тканях.

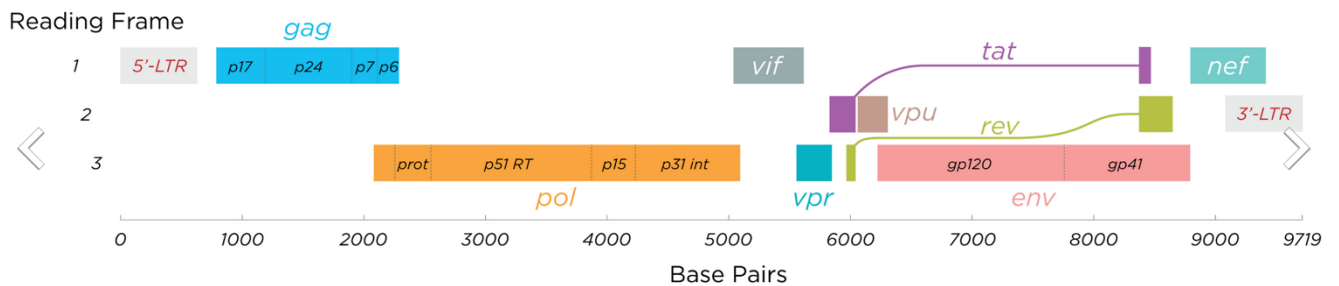


Рисунок 2. Геном ВИЧ.

Разные гены ВИЧ (рисунок 2) имеют разную степень генетического разнообразия (полиморфизма), тип естественного отбора и скорость эволюции. Ген оболочки вируса является наиболее быстро меняющимся и полиморфным геном, с вариациями от 3 до 5% внутри человека [41, 48, 49] и от 8 до 17% между людьми из одного и того же географического места [48, 50, 51]. Меньший, но все же высокий уровень разнообразия внутри индивидуума наблюдается у белка капсида, от 1 до 2% [52] и у белков полимеразы и протеазы, от 0,4 до 1% [53, 54]. В генетической изменчивости белка протеазы преобладают синонимичные и химически консервативные замены [53], что предполагает доминирующую роль очищающего отбора. Напротив, в эволюции оболочки преобладают несинонимичные замены, указывающие на наличие положительного отбора или отбора на разнообразие, вызванного иммунным ответом (глава 7).

Задача [6] состояла в том, чтобы разработать модель, которая объясняла бы происхождение быстрой эволюции и большого генетического разнообразия ВИЧ в организме хозяина, не принимающим антивирусных препаратов, и подтверждалась бы имеющимися геномными данными. Другой задачей было оценить среднее значение коэффициента отбора в типичном пациенте.

1.2.2 Оценка эффективного размера популяции

Еще в 1990-х годах велись споры о том, контролируется ли генетическое разнообразие ВИЧ у среднего нелеченного человека в основном дарвиновскими силами, такими как естественный отбор, или в основном стохастическими эффектами, такими как случайная мутация и случайный генетический дрейф из-за флуктуаций количества потомков одного провируса ВИЧ, встроенного в ДНК. Роль стохастических факторов критически зависит от эффективного размера популяции вируса в организме хозяина, а именно от числа мутаций на популяцию на нуклеотид, равное $N\mu$. Хотя количество клеток производящих вирус, $N = 10^7 - 10^8$ [55] намного больше, чем обратная вероятность мутации $1/\mu$ на основании, где $\mu = 0.4 \cdot 10^{-5}$ to $4 \cdot 10^{-5}$ [56], что согласуется с детерминистической эволюцией, есть довольно много сценариев, в которых эффективный размер популяции ВИЧ N_{eff} намного меньше, чем ее общий размер N [57]. Например, не все РНК-производящие клетки производят вирус, способный достигать других клеток. Из-за этого осложнения, оценки N_{eff} , полученные разными методами, варьировались в широких пределах: от 100 [57] до 10^8 [58].

Самый важный вопрос заключался в том, достаточно ли велик эффективный размер популяции ВИЧ, чтобы содержать мутации, которые придают вирусу устойчивость к антиретровирусному моноингибитору, то есть пептиду, чья способность связывать вирусный белок может быть нарушена заменой единственного нуклеотида. Положительный ответ на этот вопрос объяснил бы быстрое появление резистентности к любой монотерапии наблюдаемое у пациентов.

Трудность с оценкой эффективного размера популяции заключалась в том, что метод должен был быть устойчивым к предположениям модели популяции. Количество возможных моделей бесконечно. Нужно было отыскать сильный эффект, который бы выявил низкий эффективный размер популяции, независимо от дополнительных предположений модели. Им оказался эффект клональной интерференции.

Этот эффект был предсказан Фишером [59] и Мюллером [60], которые математически продемонстрировали, что, в небольшой популяции такой что $\mu N_{\text{eff}} < 1$, полезные мутации происшедшие в разных геномах мешают друг другу увеличиваться в числе и закрепляются в популяции только по одной из-за взаимной конкуренции. Кроме того, в небольшой популяции естественный отбор и генетический дрейф, действующие на один локус, влияют на эволюцию другого локуса (эффекты генетического фона и Хилла-Робертсона) [61, 62]. В один момент времени, эти эффекты можно наблюдать как сильную корреляцию между парами разных локусов: частоты четырех возможных генетических вариантов в двух локусах («гаплотипов») не мультипликативны по однолокусным частотам. Численная реализация этой идеи [5] рассмотрена

в § 2.2. Хотя стохастическую эволюцию двух локусов можно описать аналитически системой из четырех сцепленных уравнений Колмогорова [63, 64], данную практическую задачу проще решить численно.

1.2.3 Оценка вероятности рекомбинации

Эксперименты [65, 66] и модели с небольшим числом локусов [60, 62, 67-69] показали, что эволюция различных локусов, вообще говоря, не является независимой, даже если они не взаимодействуют биохимически, из-за эффектов генетического сцепления, включая клональную интерференцию и эффекты генетического фона, упомянутые выше и исследованные в *главе 3*. Эффекты генетического сцепления, происходящие из-за общей родословной геномов, замедляют адаптацию (увеличение средней приспособленности со временем), по сравнению с однолокусной моделью [59, 69, 70]. Величина этого эффекта сцепления, как показано в работах по бесполовым популяциям, возрастает с увеличением числа локусов [7, 71]. Предполагается, что половое размножение и рекомбинация возникли в ходе эволюции видов, чтобы противодействовать генетическому сцеплению и ускорить адаптацию [59, 60, 69, 71]. Это важное предсказание подтверждается исследованиями, в которых вычисляется скорость адаптации с рекомбинацией и без, с использованием двухлокусных [65, 68] и многолокусных моделей [7, 8, 10, 72-75]. При наличии рекомбинации, выгодные аллели, чтобы не вымереть из-за конкуренции между геномами, должны постоянно перескакивать на все более приспособленные геномы [76]. Модели популяции, учитывающие рекомбинацию, рассмотрены в *главе 4*.

Поскольку генетическое сцепление локусов сильно замедляет адаптацию генома (*глава 3*), многолокусные модели предсказывают, что даже редкая рекомбинация играет существенную ускоряющую роль [8, 10, 14, 74, 76] (*глава 4*). Как и некоторые другие вирусы, ВИЧ имеет эффективный механизм рекомбинации. Клетка, суперинфицированная несколькими штаммами ВИЧ, может производить вирусные частицы, содержащие гетерологичные пары РНК-геномов. При проникновении в новую клетку и обратной транскрипции геномной РНК в провирусную ДНК, белок-полимераза несколько раз переходит с одной матрицы на другую, создавая рекомбинантный ДНК-провирус. На момент начала работы, было неизвестно, какой процент инфицированных клеток инфицирован более чем один раз генетически удаленными штаммами и, следовательно, вносит вклад в вероятность рекомбинации на геном, r . Существовавшая тогда оценка r была основана на весьма упрощенной модели [77].

Для оценки этого важного параметра, были разработаны чувствительная мера рекомбинации и компьютерная модель эволюции [15]. Эти инструменты были применены к доступным генетическим данным по ВИЧ протеазе. Детали этой работы [15] описаны в § 2.3.

1.3 Многолокусные модели эволюции бесполой популяций

1.3.1 Многолокусная эволюция с фиксированным коэффициентом отбора

Фиксация выгодных аллелей (вариантов нуклеотидов увеличивающих приспособленность) в популяции замедляется за счет генетического сцепления [59, 70]. Позже, Хилл и Робертсон [78] показали, что отбор аллеля в одном локусе увеличивает генетический дрейф во втором сцепленном локусе, снижая эффективность отбора в обоих локусах. Затем Фельзенштейн [69] объяснил, что механизмы описанные Фишером [59], Мюллером [70] и Хиллом и Робертсоном [78], по сути, одинаковы. В результате всех этих эффектов, адаптация популяции сильно замедляется.

В 1990-х годах, популяционные модели все еще ограничивались одним или двумя локусами (основаниями) (§§ 2.1 и 2.2). Генетическое сцепление между большим количеством локусов, вызывающая взаимозависимость их эволюции, представляет собой серьезную проблему для аналитического решений. Если стохастическую эволюцию одного локуса можно описывать дифференциальным уравнением в частных производных второго порядка, уравнением Колмогорова [33], а для двух локусов нужно четыре сцепленных уравнения Колмогорова [63, 64], то для многолокусной системы потребовалось бы система сцепленных уравнений, экспоненциально большая по числу локусов. Поэтому количественные исследования этого эффекта либо рассматривают два (максимум три) локуса аналитически, либо используют псевдослучайное моделирование, либо делают радикально упрощающие предположения. Скорость эволюции в приближении двух и трех конкурирующих клонов, без учета множественных конкурирующих и вложенных клонов, была предсказана аналитически и проверена экспериментально [79-81].

Кроме клональной интерференции и генетического фона упомянутых выше, важным эффектом генетического сцепления является «храповик Мюллера» [69], а именно, накопление вредных мутаций усиленное генетическим сцеплением. Проще говоря, когда действует случайный генетический дрейф, наиболее приспособленный генотип в популяции в конечном итоге будет потерян, несмотря на действие естественного отбора; затем, следующий наиболее приспособленный генотип будет потерян, и так далее. Потеря наиболее приспособленных генотипов будет необратимой, если какой-либо другой процесс не воссоздаст особей с высокой приспособленностью. Одним из таких процессов является рекомбинация (*глава 4*), а другим выгодные мутации.

Преыдушие аналитические исследования храповика Мюллера рассматривали одну из двух возможных крайностей. В одном пределе, когда популяция очень мала, все особи почти все

время принадлежат к одному и тому же генотипу, и события фиксации новых аллелей хорошо разделены во времени [82, 83]. В противоположном пределе, когда популяция очень велика, распределение геномов по количеству вредных мутаций почти всегда близко к равновесному [84]. После того, как наиболее приспособленный генотип теряется из популяции, это равновесие быстро восстанавливается. Среднее время до потери наиболее приспособленного генотипа было рассчитано с использованием диффузионного приближения [85-87]. Промежуточный случай в широкой области параметров, в которой и находится большинство реальных популяций, рассмотрен не был.

В данной работе [7], разработана модель многолокусной эволюции, которая не включает рекомбинацию, но включает полезные и вредные мутации. Специально разработанный «полудетерминистический» подход основанный на методах статистической физики, описывает детерминистически все геномы, кроме самых приспособленных, немногочисленных геномов, которые описываются как миноритарная аллель, с помощью уравнения Колмогорова. Показано, что эволюционирующая популяция является уединенной волной (солитоном), бегущей по координате приспособленности, и ее скорость контролируется процессами на стохастическом фронте. Этот метод позволяет получить аналитически скорость либо понижения (храповик Мюллера), либо повышения приспособленности (адаптация), для произвольных значений параметров модели. Эта работа [7] и ее улучшенный вариант [11] рассмотрены в § 3.1. Эти результаты были подтверждены позже другими группами [12, 72, 88]; генеалогические свойства такой модели популяции были также изучены [89, 90].

1.3.2 Многолокусная эволюция с распределенным коэффициентом отбора

Выше предполагалось, что все мутации имеют эффект на приспособленность, положительный или отрицательный, но одинаковой абсолютной величины, «коэффициент отбора» s . В реальных организмах, значение s широко варьируется в зависимости от локуса. Случай экспоненциального распределения часто наблюдается в эксперименте для полезных мутаций [91-95] (§ 5.4). Априори не очевидно, что эта ситуация эквивалентна задаче с неким одинаковым «эффективным» значением s [96]. Ранние исследования Монте-Карло показали, что аппроксимация эффективного s может быть верной, когда распределение локусов по s затухает быстрее при больших s чем экспоненциальное [97]. Если хвост распределения затухает медленнее, приближение эффективного s не работает, и широкий диапазон значений s необходимо анализировать в явном виде. С этой целью были разработаны разные подходы. Две группы вычислили скорость эволюции предположив, что в любой момент времени только два или три клона (группы идентичных последовательностей) конкурируют друг с другом за раз [79, 81]. Эти авторы

предсказали, что скорость эволюции почти не зависит от размера популяции и гораздо меньше чем для одно-локусной модели (§ 2.1), на фактор порядка числа локусов. Результат же многолокусной модели с фиксированным значением s [7, 11, 12, 72] предсказывал логарифмически медленное возрастание скорости эволюции с размером популяции N и переход к одно-локусному значению при очень больших N (§ 3.1). Было непонятно, где и как результаты этих двух моделей сшиваются.

Данная работа ответила на этот вопрос [16]. Скорость эволюции была выведена аналитически для произвольно большого количества конкурирующих или вложенных друг в друга клонов с выгодными мутациями. Было использовано, параллельно, три метода, с весьма близкими результатами: оригинальный метод оптимальной мутации, Марковская цепь для вероятности фиксации аллеля [76] и метод «настраиваемой модели» [88]. Работа [16] объяснила результаты предыдущего моделирования [97] и, заодно, позволила сшить предсказания двух предыдущих теорий, упомянутых выше.

1.4 Многолокусная эволюция частично половых популяций

1.4.1 Эволюция с рекомбинацией и начальным генетическим разнообразием

Как уже упоминалось, накопление полезных мутаций в конечной популяции замедляется, если эволюционирующие локусы сцеплены в хромосоме [59, 70]. Это предсказание было подтверждено на моделях с двумя или несколькими локусами [68, 69, 78], численными исследованиями [98], а позже и более реалистичными много-локусными моделями в отсутствие рекомбинации [7, 11, 12, 72, 88] (*глава 3*). Было высказано предположение, что биологическая роль рекомбинации геномов родителей при репродукции заключается в противодействии неблагоприятному влиянию сцепления на прогрессивную эволюцию организмов и ускорении фиксации полезных мутаций. На основе этого эффекта в ряде работ рассматривалась эволюция пола в двухлокусных моделях [68, 99]. Храповик Мюллера занимал видное место в дискуссиях об эволюции пола [99-102]. Из-за математических трудностей, аналитический подход к реалистическим моделям с большим количеством локусов не применялся. Помимо математического вызова, также существовала биомедицинская необходимость предсказать эволюцию лекарственной устойчивости у ВИЧ-инфицированных, проходящих антиретровирусную терапию: ВИЧ имеет эффективный механизм рекомбинации геномов.

Чтобы заполнить этот пробел, работа [8], описанная в § 4.1, обобщила метод бегущей волны разработанный до этого для бесполок организмов [7] (*глава 3*), включив в него рекомбинацию. Рассмотрена начальная популяция с небольшим количеством выгодных аллелей. Скорость эволюции выражена аналитически (в общем виде) через размер популяции,

коэффициент селекции, число локусов, и вероятность рекомбинации.

1.4.2 Эффекты общей родословной (филогенеза) на эволюцию с рекомбинацией

Аналитический результат в § 4.1 основан на допущении, что аллели на гомологичных локусах разных геномов статистически не коррелируют между собой [8]. Последующий численный тест методом Монте-Карло [9] показал, что межгеномные корреляции, как и эффекты генетического сцепления между локусами, на самом деле, существуют, из-за общего происхождения (филогенеза) геномов, что сильно снижает ускоряющий эффект рекомбинации. Чтобы учесть эти корреляции, две последующие работы представленные в § 4.2 использовали более сложную аналитическую модель, учитывающую филогенетические связи [10, 14]. Этот метод выходит за рамки классов приспособленности и позволяет получить структуру семей и слияние генетических линий.

Эта работа была продолжена численными методами. Дополняя прежние результаты [10, 14], были обнаружены парадоксально сильные эффекты генетического сцепления между далеко расположенными локусами в присутствии сколь угодно частой рекомбинации [103]. Возможно, эти эффекты ответственны за появление новых штаммов SARS-CoV-2, «вариантов вызывающих озабоченность».

1.5 Определение адаптивного ландшафта из геномных данных

1.5.1 Универсальный эволюционный след эпистаза

Локусы генома, белков, РНК биохимически взаимодействуют с другими локусами. Это взаимодействие, «эпистаз», является вездесущим свойством биологических систем [104-107]. Эпистаз играет важную роль в эволюции популяций и наследуемости признаков. Учет эпистаза улучшает предсказания фенотипа по генотипу у различных видов, включая кур [108, 109], дрожжей [110-112], вирусов и бактерий [113-117] и различных растений [118-120]. Генетические исследования заболеваний человека сообщают об эпистатических взаимодействиях для большого количества пар локусов [121, 122], что указывает на важную роль эпистаза в патогенезе [123-130].

Примерами биологического сценариев, требующих количественных знания эпистаза, является адаптация вируса после передачи вируса новому хозяину (§ 2.1) или после распространения вируса между органами. Часто адаптирующийся организм или вирус проходит через промежуточные генетические варианты с пониженной приспособленностью, называемые

«долинами приспособленности» [116, 117, 131, 132]. Именно так патоген часто вырабатывает устойчивость к лекарствам или иммунному ответу [113-117, 132, 133] (§ 2.1).

Эпистатические локусы могут взаимодействовать как путем прямого биохимического связывания, так и удаленно (аллостерически) [133]. Последний тип взаимодействующих пар не могут быть обнаружены на основании простых структурных исследований. Примером служит пара мутаций E138K и M184I, придающая ВИЧ устойчивость к четырем одобренным FDA препаратам в фазе 3 клинических испытаний, что приводит к неэффективности лечения этими лекарствами. Пара 150L и A71V также связана с лекарственной устойчивостью ВИЧ [134, 135].

Чтобы предсказать такие эпистатические пары из генетической базы данных, нужно понять, как идентифицировать их отпечатки на генетическом уровне. В этой связи, огромные усилия приложены для определения взаимодействий между локусами на основе геномных данных, и десятки статистических методов разработаны с этой целью [136-140]. Все эти методы делают вывод об эпистазе, на основании измерения статистической ассоциации доли аллелей на разных локусах. Методы включают в себя регрессионный анализ [141, 142], байесовские методы [140, 143, 144] и статистику гаплотипов [145, 146]. Тем не менее, из многих предложенных взаимодействующих пар, очень немногие подтверждены независимыми экспериментами [107, 147].

Одной из причин этого затруднения является относительно небольшое число эпистатических пар локусов среди всех возможных. Действительно, количество эпистатических пар пропорционально длине генома, а количество всех возможных пар его квадрату. Другим важным препятствием к обнаружению эпистаза являются эффекты генетического сцепления, а также непрямые взаимодействия (§§ 5.2 и 5.3).

Кроме того, в литературе отсутствовал метод измерения взаимодействия для даже известных пар, который бы напрямую и количественно оценивал силу эпистаза без запутывания других параметров модели и переменных состояния популяции. Существующие методы обнаружения основывались на различных статистических маркерах, предназначенных для распознавания взаимодействующих пар качественно, без определения точной силы эпистаза [106, 107].

Чтобы заполнить последний пробел, в работе [24] была разработана и протестирована, аналитически и с помощью моделирования, мера эпистаза, которая зависит только от эпистатической силы (степени компенсации вредной мутации) и топологии эпистатической сети, но не от состояния популяции и других параметров модели.

1.5.2 Обнаружение эпистатических пар локусов в одной отдельно взятой популяции: миссия невыполнима

Статистическая корреляция между аллелями на разных локусах вызвана не только эпистазом. Общее происхождение аллелей в разных локусах, но расположенных на одном геноме, приводит к разнообразным эффектам генетического сцепления (§§ 3.1, 4.2), а также межлокусным корреляциям. Амплитуда этих эффектов увеличивается с числом локусов в геноме. Рекомбинация уменьшает эффекты сцепления между далекими локусами [8, 9, 15, 22, 74-77] и частично компенсирует эффекты эпистаза, что в условиях огромного размера популяции может привести к т.н. «квази-равновесию сцепления» (quasi-linkage equilibrium) [148-150]. Используя простейшие модели с двумя локусами, было показано, что эпистаз важен для эволюции рекомбинации [99, 151].

Аллельную ассоциацию, возникающую в результате сцепления, легко спутать с эффектами эпистаза. Этот эффект сцепления имеет случайный знак и возникает из-за случайных факторов: мутаций в случайных местах генома и генетический дрейфа. Эффекты сцепления становятся незначительными только в гигантских популяциях, экспоненциально увеличивающихся с числом локусов [7] (§ 3.1). Имея дело с реальным набором последовательностей ДНК или РНК, неясно, как отличить данный эффект от эпистатического взаимодействия и какой из двух доминирует в данном образце ДНК. Таким образом, обнаружение эпистаза по геномным данным остается очень сложной задачей [105-107].

Чтобы решить проблему повторяющейся неудачи в обнаружении эпистаза в образцах ДНК, нужно было найти ее причину. Работа [26] обсуждаемая в § 5.2 ее обнаружила, воспроизведя этот эффект в компьютерной модели.

1.5.3 Обнаружение эпистатических пар локусов методом трехлокусной корреляции

Как показано в предыдущем разделе, случайная ковариация аллелей из-за эффектов сцепления полностью скрывает отпечаток эпистаза в геномных образцах [26]. Это ограничение не может быть устранено никаким методом обнаружения, включая популярный подход основанный на приближении «квази-равновесия сцепления» (QLE) [148, 150]. Метод устранения аллельных корреляций из-за косвенного взаимодействия, основанный на этом приближении [148], не годится для устранения корреляций вызванных генетическим сцеплением, так как он просто постулирует их отсутствие. Поэтому он применяется к последовательностям разных видов животных, которые разошлись эволюционно миллионы лет назад, где это приближение имеет смысл, хотя бы, для наиболее недавних мутаций [152, 153].

Самый недавний общий предок текущей популяции удаляется от начального древнего предка по стохастической траектории в многомерном генетическом пространстве (§ 5.2). В результате, любая мера межлокусной корреляции или даже использование всего

филогенетического дерева дает только сильный шум со случайным знаком [26]. Единственный способ помочь обнаружению — усреднить частоты гаплотипов по многим независимым популяциям, развивающимся в сходных условиях. Однако, как оказалось, даже в очень коротком геноме из 40 локусов — а в некоторых вирусах их сотни и тысячи, не говоря уже о организмах с миллионами локусов — даже двухсот независимых популяций недостаточно, и половина предсказанных взаимодействий оказываются ложными [28].

Чтобы преодолеть эту трудность, в работе [28] разработан метод устранения этого гигантского шума, с помощью дополнительного условия на третьем локусе. Для демонстрации высокой надежности этого подхода даже при скромном количестве популяций, от 20 до 200, использованы два параллельных теста: Монте-Карло и аналитический тест. Результаты применены к геномным данным вируса гриппа, что позволяет объяснить происхождение пандемии «свиного гриппа» 2009 года [28].

1.5.4 Оценка коэффициентов отбора отдельных локусов из геномных данных

В дополнение к эпистазу, полное знание адаптивного ландшафта, то есть связи между генотипом и приспособленностью, включает в себя знание коэффициентов отбора мутаций на всех локусах. Они необходимы для предсказания эволюционной траектории популяции, например, при возникновении лекарственной устойчивости или иммунного ускользания вирусов (*глава 7*). Нет ничего удивительного в том, что огромные усилия были затрачены на разработку различных методов вывода коэффициентов отбора, не меньшие, чем на обнаружение эпистаза. Прямые лабораторные измерения коэффициентов отбора у вирусов и бактерий требуют трудоемких экспериментов [91-95]. Упрощенные аналитические модели и моделирование Монте-Карло позволяют оценить средние значения из геномных образцов в разные моменты времени [6, 15, 77, 154] (§§ 2.1 и 2.3). Однако, из-за генетического сцепления и эпистаза, оценка коэффициентов отбора для *конкретных* локусов является куда более сложной задачей. Например, коэффициенты отбора для локусов гена гемагглютинаина вируса гриппа A/H3N2 были получены путем подгонки детерминистической одно-локусной модели с и ее приблизительного обобщения на два локуса к долям аллелей измеренным в разные моменты времени [155]. Другая группа оценила параметры распределения эффектов мутаций в стационарном состоянии в предположении, что оно имеет вид гамма-распределения [156]. Общий метод оценки коэффициентов селекции, свободный от произвольных предположений и не искаженный эффектами сцепления, отсутствовал.

Толчком для развития метода послужило экспериментальное наблюдение, что распределение полезных мутаций по эффекту на приспособленность (DFE) часто имеет универсальную экспоненциальную форму [91-95]. Ранние попытки объяснить универсальную

DFE не использовали никаких эволюционных аргументов и основывались на формальной теории экстремальных значений [157-159]. Работа [29] показала, что экспоненциальная форма появляется с течением времени, как следствие процесса адаптации. Практическим продуктом этой теоретической работы стал метод измерения относительных значений коэффициентов отбора, независимый от эффектов сцепления.

1.6 Эволюционная роль признака

1.6.1 СПИД, как следствие адаптации ВИЧ к хозяину

У 99.5% нелеченых лиц, инфицированных ВИЧ, развивается синдром приобретенного иммунодефицита (СПИД) [160-162]. Демографические группы с высоким риском заражения включают мужчин, имеющих незащищенный секс с мужчинами, потребителей внутривенных наркотиков, повторно использующих иглы, и работниц сферы интимных услуг [163]. Симптомы СПИДа охватывают широкий спектр оппортунистических заболеваний у ранее здоровых людей, включая саркому Капоши, пневмонию, и микобактерии туберкулеза [164-167]. Непосредственной причиной приобретенного иммунодефицита у ВИЧ-инфицированных является истощение вирусом CD4 Т-клеток, сопровождающееся функциональными нарушениями, такими как низкая пролиферативная реакция лимфоцитов. Эти процессы вызывают появление симптомов СПИДа в течение 2–20 лет [164-166]. Истощение CD4 Т-клеток, создающее пробелы в их антигенном репертуаре, связано с тем фактом, что ВИЧ избирательно инфицирует профессиональные антигенпрезентирующие клетки, экспрессирующие белок CD4, включая хелперные CD4 Т-клетки, макрофаги и дендритные клетки. Низкий уровень CD4 Т-клеток в конце инфекции приводит к пробелам антигенного репертуара лимфоцитов, что вызывает СПИД [168, 169].

Два десятилетия назад, были разработаны сильнодействующие коктейли антиретровирусных препаратов, чтобы предотвратить или отсрочить этот исход и вернуть ВИЧ-инфицированных к нормальной жизни. Были предприняты активные международные усилия, чтобы сделать их широко доступными во всем мире, включая развивающиеся страны. Антиретровирусная терапия (АРТ) подавляет репликацию ВИЧ, что увеличивает количество CD4 Т-клеток в периферической крови и восстанавливает нормальную иммунную функцию [170, 171]. Эти результаты подтверждают, вне всякого сомнения, что иммунодефицит у нелеченых лиц вызван снижением количества хелперов, CD4 Т-клеток, и частичной потерей функции оставшимися хелперами. Дополнительным подтверждением является экспериментальное заражение нечеловекообразных обезьян патогенными штаммами Вируса Иммунодефицита Обезьян (родственным ВИЧ) или химерными обезьяньими/человеческими вирусами

иммунодефицита, которое приводит к быстрому истощению CD4 Т клеток и гибели животного в течение одного-двух лет [172-175].

В то же время, инфицирования вирусом CD4 клеток недостаточно для объяснения прогрессирования в СПИД, так как большинство CD4 клеток незаражены и могут делиться до бесконечности, так как они экспрессируют белок теломеразу. Число этих клеток-мишеней уменьшается во время острой фазы инфекции, а затем частично восстанавливаются примерно до половины нормального уровня [176-179]. Дальнейшая инфекция представляет собой устойчивое состояние с быстрым оборотом и заменой всех важных типов клеток. Гибель инфицированных клеток примерно через сутки полностью компенсируется поступлением вновь инфицированных клеток [170, 171, 180-183]. Основные клетки-мишени ВИЧ, CD4+ CCR5+ Т-клетки памяти, также постоянно пополняются, и время их обращения составляет несколько недель [180, 184, 185]. Равновесное состояние числа CD4 клеток почти точное, плюс минус флуктуации, но не совсем точное, с очень медленным снижением их числа. Время до появления симптомов СПИДа отрицательно коррелирует с вирусной нагрузкой^{84,185} и уровнем иммунной активации [186-189].

Существование точного стационарного состояния давно объяснено с помощью простых математических моделей типа «охотник-жертва» [181, 190, 191]. Однако все эти модели предсказывают идеальное стационарное состояние. Причины медленного снижения количества CD4 клеток на временных масштабах, на два порядка превышающих время оборота клеток-мишеней вируса, остаются неясными. Существование многих штаммов вируса иммунодефицита обезьян, которые не вызывают прогрессирования СПИДа у их естественных хозяев [192-196], прямо демонстрирует, что биологические факторы, вызывающие хроническую устойчивую инфекцию, отличаются от механизмов, вызывающих СПИД. Несмотря на различия, наблюдаемые между иммунным ответом на вирус иммунодефицита обезьян у его естественного хозяина и ВИЧ у людей [192, 197, 198], механизм прогрессирования заболевания к СПИДу до сих пор остается неизвестным.

Естественно предположить, что вирусы и их естественные хозяева взаимно адаптированы, а патогенные инфекции возникают после межвидового переноса, запускающего процесс адаптации вируса к новому хозяину. Эволюция ВИЧ внутри хозяина хорошо изучена [34] (§ 2.1). ВИЧ имеет самую быструю эволюцию из известных в природе даже среди РНК-вирусов со средней скоростью $3 \cdot 10^{-3}$ замен на нуклеотидную позицию в год и $1.5 \cdot 10^{-3}$ в своем оболочечном белке gp120 [199, 200]. Спайк SARS-CoV-2 является следующим по скорости вирусным белком, с частотой замен $3 \cdot 10^{-3}$ на нуклеотид в год [201].

Основываясь на этих фактах, в работе [202], обсуждаемой в § 6.1, было предположено, что постепенное уменьшение числа CD4 Т-клетках, приводящее к СПИДу, вызвано адаптацией вируса к хозяину. Разработана модель, объединяющая иммунологию с многолокусными

моделями эволюции (главы 3 и 4), предсказания которой гораздо лучше согласуются с экспериментальными данными, чем альтернативные объяснения популярные в литературе [202].

1.6.2 Эволюционная роль латентного состояния ВИЧ

Как сказано выше, у инфицированных людей ВИЧ реплицируется в CD4 T-лимфоцитах памяти (CCR5+). Большинство инфицированных клеток производят большое количество вируса и погибают примерно через день после заражения. Однако вирус, интегрированный в ДНК человека, провирус, также может находиться в состоянии покоя, называемом «латентное» [203, 204]. В этом состоянии, транскрипция провируса не происходит. Размер латентной популяции в зараженном человеке относительно невелик, около 1 на 10^6 CD4 T-клеток, и он не вносит большого вклада в общую вирусную нагрузку [205]. Однако, из-за латентной субпопуляции, ВИЧ пациенты должны оставаться на пожизненной антивирусной терапии, чтобы избежать рецидива. Как только пациент прерывает прием антивирусных препаратов, латентный вирус реактивируется, и ВИЧ поднимается до уровня, существовавшего до лечения, в течение нескольких недель. Следовательно, латентность является основным препятствием для полного удаления вируса из пациента [206].

В то время как латентность помогает ВИЧ сохраняться во время недавно созданной лекарственной терапии (около 2000), его эволюционная польза для вируса в течение миллионов лет естественной эволюции неясна. ВИЧ и другие лентивирусы эволюционировали миллионы лет в отсутствие антиретровирусных препаратов. Казалось бы, латентность должна быть только вредна для вируса, потому что вирус размножается активно, не прячется, а латентно инфицированные клетки не генерируют вирус и, таким образом, снижают уровень вируса у людей. Учитывая этот факт, популярна гипотеза о том, что латентность — это случайность эволюции, которая произошла несколько миллионов лет назад. Многие вирусологи считают, что механистической причиной латентного периода является инфекция CD4 T-клеток, находящихся в переходном состоянии между активированной и покоящейся фазами клеточного цикла [207-209]. Таким образом, латентность считается редким побочным эффектом, который проявляется только после начала иммунного ответа, активирующего CD4 T-клетки хозяина.

Однако исследование на макаках-резусах инфицированных ВИО, вирусом иммунодефицита обезьян, родственным ВИЧ, показало, что латентные клетки достигают высоких уровней очень рано, в течение нескольких дней после заражения [210], задолго до формирования адаптивного иммунного ответа на вирус [211, 212]. Другое экспериментальное исследование в культуре клеток [20] показало, что латентность встроена в регуляторную цепь гена ВИЧ, что также согласуется с другим набором данных, показывающим, что половина ВИЧ

инфекций в культуре приводит к латентному состоянию [213, 214]. Несколько других исследований продемонстрировали, что ауторегуляторная цепь Tat ВИЧ усиливает стохастические флуктуации в экспрессии провируса ВИЧ, которых достаточно, чтобы вызвать переключение в латентное состояние [215-217]. Вообще говоря, такой стохастический шум экспрессии генов вреден для вирусов или организмов [218, 219]. Таким образом, сохранение латентного состояния, встроенного в регуляцию гена провируса ВИЧ, предполагает неизвестное преимущество, которое перевешивает вред.

Возможное преимущество латентности заключается в том, что она может повысить шансы вируса на выживание в суровых условиях окружающей среды. Авторы [220] предложили подобное объяснение для лизогении бактериофага-λ [221] и персистенции бактерий [222]. В своем естественном хозяине, ВИЧ редко подвергается опасности горлышка бутылки. В отсутствие лечения он сохраняют высокую вирусную нагрузку $\sim 10^5$ частиц/мл плазмы крови и $\sim 10^8$ инфицированных клеток на одного хозяина в течение многих лет (вероятно, из-за подавления иммунного ответа за счет инфицирования иммунных хелперных CD4 клеток [223, 224].).

Однако существует фаза инфекции, когда вирусная нагрузка всегда низка, а именно, начальная инфекция слизистой оболочки. Многочисленные исследования показывают, что вероятность заражения ВИЧ при незащищенном половом акте очень мала. В парах, где один партнер заражен ВИЧ а другой нет, менее 1% незащищенных контактов приводит к системной ВИЧ-инфекции [225-227]. Также известно, что системная инфекция обычно запускается единственной последовательностью РНК в переданной дозе (инокуляте) [228, 229]. Использование обезьяньей модели вируса иммунодефицита обезьян выявило задержку примерно в 6 дней между временем экспериментальной инокуляции и экстраполированным началом системной инфекции [230, 231].

Таким образом, условия слизистой в первые 6 дней не позволяют вирусу расти [232]. Это связано с тем, что репликация ВИЧ, передаваемого обычно половым путем (у животных, также при зализывании ран [233]), начинается в слизистой ткани влагалища, рта или прямой кишки, бедной клетками, которые требуются вирусу для размножения [231]. Каждая активно инфицированная клетка живет примерно день [234], так что через 5 дней весь выживший вирус находится только в латентной форме. В данной работе [19], была проверена гипотеза о том, что латентность ВИЧ обеспечивает преимущество на этапе передачи вируса от хозяина к хозяину.

1.6.3 Роль рекомбинации в адаптации и оптимальная вероятность мутаций в системе полиовирус-мышь

РНК-вирусы быстро адаптируются к окружающей среде, что приводит к осложнениям заболеваний и их устойчивостью к противовирусным препаратам и иммунным ответом [235-237]. Поэтому понимание эволюционных факторов адаптации вируса имеет решающее значение для разработки противовирусных препаратов и вакцин, а также для понимания механизма вирулентности. Адаптация популяции происходит путем случайной мутации и последующего отбора новых полезных аллелей [235, 238, 239]. Высокая частота мутаций РНК-вирусов по сравнению с ДНК-организмами является одной из причин их эволюционной пластичности [238, 240]. В то же время большинство мутаций хорошо эволюционировавшего вируса ему приносят вред [29, 241, 242].

Как уже говорилось, в популяциях без рекомбинации полезные аллели находятся под эффектами генетического сцепления, включая клональную интерференцию [59, 60, 65, 67, 239, 243, 244] и накопление вредных аллелей (храповик Мюллера) [69, 245, 246]. Адаптация, то есть увеличение числа полезных аллелей, замедляется сцепленными с ними вредными аллелями [239, 247]. Экспериментальные и теоретические исследования показывают, что рекомбинация частично смягчает эти эффекты, комбинируя полезные мутации в одном и том же геноме (*глава 4*) и отфильтровывая вредные мутации [8, 74, 243, 244, 247-250].

Полиовирус представляет собой РНК-вирус с положительным смыслом, как мРНК в клетке, вызывающий полиомиелит [251, 252]. Как известно из экспериментов на культурах клеток, геном полиовируса подвергается частой рекомбинации, опосредованной переключением матрицы во время репликации [253-255]. Стадии рекомбинации полиовируса включают в себя преждевременное прекращение синтеза РНК, отделение комплекса полимеразы-РНК от матрицы, его повторную ассоциацию с другим геномом, а затем завершение репликации, в результате чего образуется химерный геном. В соответствии с этой схемой, гомология нуклеотидов между родительскими последовательностями важна для частоты рекомбинации и местоположения [253, 256].

Эффекты мутации и рекомбинации в сочетании с естественным отбором и случайным дрейфом интенсивно изучались [7-12, 14, 16, 65, 71, 73, 89, 90, 257] (*глава 4*). Синергический эффект мутации и рекомбинации также был исследован теоретически [74, 76]. Две теоретических работы на основе многолокусных моделей, вошедшие в данную диссертацию (§ 3.1), предсказали существование оптимальной вероятности мутаций, которая максимизирует скорость адаптации [7, 258].

Однако, взаимодействие этих факторов было мало изучено экспериментально, в частности, при острой вирусной инфекции. Также было неясно, действительно ли вероятность мутаций имеет оптимум, как предсказано теорией [7, 258], и если да, то зависит ли этот оптимум от рекомбинации. Чтобы восполнить этот пробел, группой вирусологов были проведены

эксперименты с 6 вариантами вируса, обладающими очень разной скоростью мутации и рекомбинации, как в культуре клеток, так и на мышах [22]. Затем эти данные были интерпретированы с помощью компьютерной модели [22] (§ 6.3).

1.7 Наперегонки с враждебным биологическим видом

1.7.1 Эволюция респираторного вируса в популяции под давлением гуморального иммунитета

Чтобы избежать иммунного распознавания хозяином, ранее перенесшим вирусную инфекцию, респираторные вирусы накапливают мутации в иммунологически-значимых областях генома, а именно, эпитопах антител [259]. Например, вирус гриппа ежегодно заражает 5-15% населения мира. Глобальная персистенция вируса из-за повторного заражения ранее инфицированных людей вызвана быстрой эволюцией антигенов (областей, связывающих антитела) в белке гемагглютинине, служащем вирусным рецептором для проникновения в клетку [259]. Доступная информация о вирусе гриппа довольно обширна и включает в себя его широкую циркуляцию в мире [260-262], генетическое картирование антител и штаммов вируса, молекулярную структуру и вирусологический цикл, а также коэффициент отбора для некоторых мутаций [259, 263-265].

Компьютерное моделирование в сочетании с анализом данных пролило свет на механизмы эволюции вируса и позволило прогнозировать краткосрочную эволюцию [262, 266-270]. Однако общая связь между эпидемиологическим, иммунологическим, и эволюционным поведением вируса была неясна. В работе [25] (7.1), разработана модель, которая находится на пересечении популяционной генетики, вирусологии, иммунологии и эпидемиологии. С ее помощью аналитически выведен эффективный коэффициент отбора создаваемый иммунной памятью населения, который постоянно гонит вирус вперед в антигенном пространстве. Проблема антигенной эволюции, таким образом, сведена к ранее решенной проблеме многолокусной эволюции с заданным коэффициентом селекции [11, 16] (главы 3 и 4).

Позже две другие группы получили похожие результаты, несмотря на другие предположения относительно механизма коллективного иммунитета [271, 272]. Тот же анализ, вероятно, применим к эволюции эпитопов белка Спайк в SARS-CoV-2 и может помочь с оптимизацией эффективности вакцинации против КОВИДа [1].

1.7.2 Эволюция ВИЧ в хозяине под давлением клеточного иммунитета

Несмотря на сильный иммунный ответ в нелеченом хозяине, ВИЧ персистирует годами. Вирус избегает полного клиренса и устанавливает хроническую инфекцию за счет комбинации двух

методов: путем уничтожения хелперных CD4 Т-клеток и, таким образом, снижения чувствительности иммунного ответа [174, 175, 224, 273, 274], а также за счет появления мутаций в антигенно-важных областях, эпитопах, распространяясь по инфицированным клеткам [275-277]. В результате, вирус становится частично устойчивым к иммунному ответу.

Уровень ВИЧ в хозяине контролируется цитотоксическими CD8 Т-клетками (CTL) [278-280]. Доказательством являются эксперименты по искусственному обеднению CD8 Т-клеток у обезьян, инфицированных ВИО, что вызывает быстрый всплеск вирусной нагрузки [212, 281]. Быстрая эволюция ВИЧ в эпитопах [275-277] также демонстрирует, что цитотоксические клетки, которые распознают специфические последовательности РНК, функциональны и оказывают давление отбора на изменение вируса. Это «антигенное ускользание» препятствует разработке вакцины ²⁷¹ и, как упомянуто выше, является вероятной причиной адаптацией и патогенеза вируса [6, 202] (§ 2.1 и § 6.1).

Передаваемый штамм вируса является мишенью многих клонов CTL [278-280], каждый из которых распознает короткую последовательность из 8-10 аминокислот, представленную на поверхности клетки молекулами МНС-I. Мутации в эпитопах CTL возникают в течение нескольких недель после заражения и продолжают медленно накапливаться на протяжении всей хронической инфекции. В отсутствие иммунного ответа, эти мутации снижают скорость репликации вируса, т.е. они имеют цену для вируса) [275-277]. Важно отметить, что не все эпитопы-мишени ускользают, и что скорость ускользания снижается после нескольких месяцев заражения. Какие параметры определяют скорость ускользания в данном эпитопе, каков порядок ускользания, какие эпитопы ускользают, а какие нет, исследовалось экспериментально [282, 283]. Математическое моделирование было предпринято для изучения поздних ускользающих мутаций [284, 285] и влияния распределенного давления отбора из-за иммунного ответа [286, 287]. В этих работах, особое внимание уделялось двум параметрам: цене мутации, Δf , и количеству активных эпитопов, n . Цену ускользающей мутации можно легко оценить по скорости ее реверсии, наблюдаемой при передаче индивидууму с несовпадающим HLA-типом, чьи эпитопы расположены в других местах генома [229, 275, 276, 288]. Часто наблюдаемое появление компенсаторных мутаций за пределами мутировавшего эпитопа (что и создает эволюцию ВИЧ в пациентах [6], § 2.1), также указывает на конечную цену мутации ускользания [289, 290]. Эксперименты показывают, что цена мутации распределена в широком диапазоне [277, 291, 292].

Однако Δf и n — не единственные параметры, важные для динамики антигенного ускользания. Как было показано на ВИЧ-инфицированных людях и ВИО-инфицированных животных, ускользающий мутант не полностью предотвращает распознавание эпитопа, и

«коэффициент частичного ускользания» Δr меняется от мутации к мутации [282, 293]. Этот факт ранее учтен не был.

В работе [18] (§ 7.2), предыдущая модель [284] переработана, чтобы включить новый параметр, Δr . Показано, что параметры Δr , Δf и n вместе определяют скорость и порядок антигенного ускользания на разных локусах, и предсказания сопоставлены с ранее полученными клиническими данными [286, 294]. Предсказания этой модели объясняют происхождение отрицательной корреляции между Δr и Δf наблюдаемой в клинически-важных мутациях в гене полимеразы [295] и позволяют оценить диапазон Δr из опубликованных экспериментальных данных [296-298].

1.7.3 Эволюционная стабильность ВИЧ в хозяине в присутствии дефектных интерферирующих частиц

Если геном вируса подвергается обширным делециям во время цикла репликации, в нем будут отсутствовать элементы, необходимые для его дальнейшей репликации. Для успешной репликации, он должен быть дополнен вирусом дикого типа, заражающим ту же клетку, который будет производить недостающие элементы. Такие, условно реплицирующиеся варианты вируса называются «дефектными интерферирующими частицами» (DIP) [299, 300].

Многие патогены человека и животных имеют естественные DIP, включая вирус везикулярного стоматита [301], вирус мышиноного лейкоза [302], вирус гриппа [303], вирус саркомы Рауса [304] и вирус лихорадки Денге [305, 306]. Подверженная ошибкам репликация РНК-вирусов порождает множество дефектных мутантов [299, 300]. В вирусе мышиноного лейкоза, мутации внутри DIP приводят к экспрессии новых белков, которые усиливают вирусную цитотоксичность или иммунный ответ на инфицированную клетку [302, 307-309]. Другие DIP препятствуют репликации дикого типа или ослабляют вирулентность в животных моделях [310-314]. DIP много раз предлагались для использования в качестве терапевтических агентов и агентов переноса генов [310, 312, 315-320]. DIP могут возникать спонтанно или создаваться искусственно [30, 302-305, 318, 321]. Они также могут служить живой вакциной, вызывающей широко-нейтрализующий ответ против вируса [321].

Загадочно то, что несмотря на большие объемы данных секвенирования вариантов ВИЧ и ВЮ, а также в отличие от многих других семейств вирусов, природные лентивирусные DIP не обнаружены. В то время как способность образовывать и распространять DIP различается у вирусов [322-324], врожденного блока для образования больших делеций в геноме ВИЧ не существует [325]. Также нет каких-либо известных препятствий для мобилизации субгеномных ВИЧ, что используется при создании ретровирусных векторов для генной терапии из упаковывающих клеточных линий [326].

Отсутствие естественных лентивирусных DIP может быть связано с эволюционным ускользанием ВИЧ от его вторичного паразита. В долговременных клеточных культурах, в течение 5 лет, что соответствует нескольким сотням пассажей, наблюдалась коэволюция как ВИЧ дикого типа, так и DIP [322, 327]. В ходе этой эволюционной гонки, вирус дикого типа смог приобрести устойчивость к DIP [328-330]. Совместная эволюции других вирусов и DIP обсуждается в обзоре [317]. В то время как динамическая стабильность различных DIP были рассмотрены в большом количестве работ [315, 331-334], эволюционная стабильность лентивирусных DIP была недостаточно изучена.

В работе [17] обсуждаемой в § 7.3, разработана модель динамики ВИЧ и DIP, включающая в себя биологический уровень клетки и уровень хозяина. С ее помощью, коэффициент отбора для ускользания ВИЧ в хозяине выражен, в общем виде, через параметры модели на уровне клетки, и его знак проверен. Положительный коэффициент отбора предсказывает эволюцию штаммов ВИЧ устойчивых к DIP. В отличие от предыдущего многоуровневого моделирования вирусных инфекций [334-337], эта двухмасштабная модель учитывает конкретную вирусологию ВИЧ.

1.7.4 Эволюционный конфликт ВИЧ между уровнями хозяина и популяции в присутствии дефектных интерферирующих частиц

В § 7.3 показано, что любая одиночная мутация ускользания ВИЧ вредит ему больше, чем помогает, подавляя репликацию DIP [17]. Иными словами, цена мутации ускользания для ВИЧ больше, чем ее выгода. Таким образом, при эволюции в хозяине, ВИЧ не может ускользнуть от DIP (по крайней мере, в отсутствие эпистаза, не включенного в эту модель эволюции).

На уровне популяции хозяев, эволюционная стабильность ВИЧ в присутствии DIP оставалась неясной. В литературе этот вопрос не освещался, за исключением одной работы с довольно упрощенной вирусологией [334]. Хотя устойчивый мутант ВИЧ не распространяется в организме хозяина, в котором уже присутствует дикий тип ВИЧ и DIP, но если он все же появится (скажем, в результате большой флуктуации), он смог бы, в принципе, затем распространиться по людям получающим DIP.

В работе [21] (§ 7.4), эта возможность анализируется с фокусом на демографические группы с высоким риском инфицирования ВИЧ [338-340]. Предыдущее эпидемиологическое моделирование показало, что DIP следует за ВИЧ и автоматически концентрируется в группах высокого риска где превалентность ВИЧ велика [334]. Следовательно, эти группы высокого риска представляют собой ключевую группу населения, в которой могут возникать устойчивые к DIP мутанты.

Чтобы понять, могут ли устойчивые к DIP мутанты ВИЧ распространяться в популяции, двухуровневая модель введенная в § 7.3 была надстроена до третьего биологического уровня, а

именно, уровня населения [21]. Затем параметры адаптивного ландшафта были выражены через два безразмерных параметра модели, на уровне клетки: асимметрию экспрессии геномной РНК DIP и ВИЧ, и отношение числа капсидов ВИЧ к числу геномов.

Глава 2. Определение действующих факторов и оценка основных параметров эволюции из геномных данных

Генетическая эволюция вируса в организме животного-хозяина является результатом взаимодействия множества различных факторов [341-343]: случайной мутации, естественного отбора, включая эпистаз, случайного генетического дрейфа, генетического сцепления, рекомбинации, распространения между особями и инфицированными органами со случайной выборкой вируса-основателя. Все эти сложности приводят к необходимости использования математических и вычислительных моделей, чтобы изучать систему в упрощенном виде. Основная проблема, стоящая перед математическим моделированием реальных систем, заключается в том, что заранее неизвестно, какие из множества существующих биологических факторов влияют на поведение системы и должны быть включены в модель. Уместность приближений не может быть проверена до тех пор, пока анализ или расчет не завершены; она зависит от вопросов на которые должна ответить модель и от области параметров модели.

Таким образом, разработчик моделей сталкивается с временным парадоксом. С одной стороны, в начале необходимо ввести все приближения, чтобы сделать модель решаемой и проверяемой данными. С другой стороны, обнаружение подходящего набора приближений («модели») является конечной целью такого исследования. Динамическое взаимодействие между теорией и экспериментом разрешает этот временной парадокс [181]. Разработчик выбирает несколько основных экспериментальных наблюдений которые он собирается объяснить или подгонять и подбирает простейшую модель для их интерпретации. Когда получено первое совпадение между данными и предсказаниями модели, разработчик ищет ту часть выбранных экспериментальных данных, которая противоречит предсказаниям модели. Как только такие противоположные свидетельства обнаруживаются, разработчик меняет первоначальные предположения, одно за другим, пока разногласие не устранено. Цикл повторяется до тех пор, пока все интересующие разработчика экспериментальные данные не будут согласованы с предсказаниями модели. Результат считается рабочей моделью, подлежащей публикации и дальнейшей экспериментальной проверке, которая либо подтвердит модель, либо опровергнет, либо потребует дальнейшей доработки.

Настоящая глава представляет три работы, где данная стратегия использовалась чтобы определить доминирующие эволюционные факторы, действующие на популяцию ВИЧ в инфицированных людях не принимающих лекарств и оценить основные эволюционные параметры вирусной популяции по доступным геномным данным.

2.1 Механизм эволюции ВИЧ и оценка среднего коэффициента отбора

Чтобы понять какие главные факторы отвечают за эволюционную динамику ВИЧ, в работе [6] проанализированы последовательности белка протеазы ВИЧ из инфицированных лиц, ранее не получавших ингибиторы протеазы [53]. Генетически разнообразные (полиморфные) основания интерпретированы как локусы, подвергающиеся медленной адаптации в течение хронической инфекции. Чтобы объяснить высокую частоту повторения полиморфных локусов, наблюдаемую в индивидуальных людях, передача одной или двух последовательностей вируса между людьми также включена в модель (узкое место передачи). Сначала предполагается, что наиболее приспособленная последовательность одинакова для всех людей. При инфекции из единственного источника, такая модель может объяснить высокую частоту повторения. Однако этот теоретический результат оказывается весьма чувствительным к предположению о почти полном отсутствии коинфекции из разных источников, что не соответствует имеющимся в литературе данным. Поэтому начальное упрощение отменено и учтен тот факт, что наиболее приспособленная последовательность вируса сильно различается у разных людей, из-за индивидуальных различий в подтипах главного комплекса гистосовместимости класса I (MHC-I). Постулировано, что ускользающая мутация, не распознаваемая иммунной системой (§ 7.2), быстро фиксируется в популяции, что меняет знак коэффициентов отбора на многих других, компенсирующих локусах. Это приводит к росту числа аллелей, компенсирующих цену ускользающей мутации для вируса. Это предположение было позже подтверждено в многочисленных экспериментах (ссылки в §§ 5.1 и 7.2).

2.1.1 Разнообразие последовательностей в гене протеазы: анализ данных

Набор 265 последовательностей гена протеазы был получен в пациентах, ранее не принимавших ингибиторы протеазы [53]. Для этого 265 клонов провирусной ДНК были выделены из клеток 13 пациентов и подвергнуты 70 циклам вложенной ПЦР-амплификации.

Консенсусная последовательность была определена [6] для каждого человека и каждого основания как наиболее распространенный вариант нуклеотида (аллель) из 4х возможных. Консенсус различался между пациентами. Общий консенсус для всего набора людей был идентичен консенсусу подтипа В в базе данных LANL, который имеет длину 297 оснований и начинается с консервативной последовательности ССТСАgАТСАСТСТТ (PQITL), где g обозначает переменное молчащее основание.

Эти данные были отфильтрованы следующим образом [6].

(i) Консенсусные последовательности у пациентов с номерами 06 и 12 имели стоп-кодоны, указывающие на дефектный вирус; образцы от этих людей были исключены из анализа.

(ii) У оставшихся 11 человек 6% последовательностей содержали делеции, вставки или стоп-кодоны; эти последовательности также были отфильтрованы, поскольку основное внимание в работе уделялось точечным мутациям, которые не делают вирус неспособным к репликации. Оставшийся набор данных состоял из 213 последовательностей провирусной ДНК.

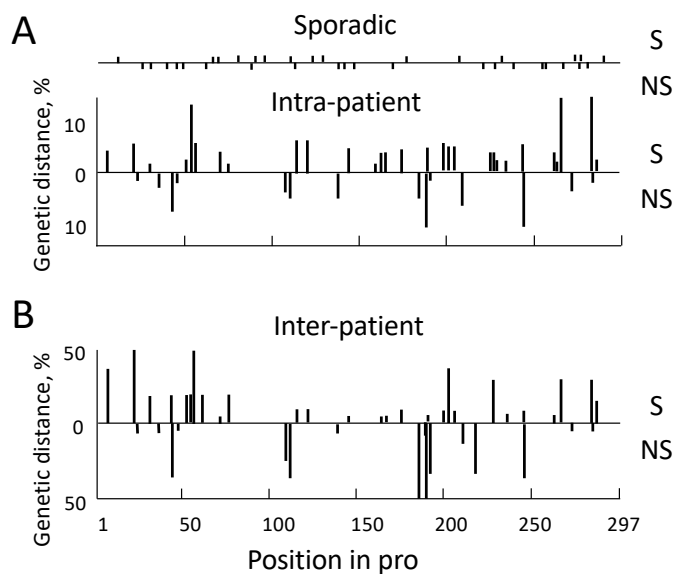


Рисунок 2.1. Внутрипациентные (A) и межпациентные (B) генетические расстояния, усредненные по пациентам, в разных положениях гена протеазы ВИЧ по данным [53]. Верхняя и нижняя гистограммы на каждом рисунке соответствуют синонимичным и несинонимичным участкам соответственно. Точки на верхней горизонтальной линии на панели а показывают положения спорадических мутаций, наблюдаемых только один раз в наборе данных. По работе [6].

(iii) 25% замен были трансверсиями относительно консенсуса, определяемыми как мутации, которые заменяют нуклеотид на нуклеотид с сильно отличающейся биохимической структурой, такой как $A \rightarrow C$ или $A \rightarrow T$. Трансверсии игнорировали, заменяя их на консенсусный вариант. В результате в каждом основании было только два аллеля: либо A/G, либо C/T. (Модель, учитывающая трансверсии, потребовала бы включения еще двух параметров, скорости прямой и обратной мутации для трансверсий, которые малы и были известны в то время с плохой точностью.)

(iv) Чтобы исключить возможные ошибки во время раундов ПЦР, аллели, присутствующие в единственном экземпляре во всей базе данных, игнорировались. Доля «спорадических» мутаций была 0.06%, что соответствует частоте ошибок ПЦР 1 на 10^5 оснований на цикл [344, 345].

В отфильтрованной базе данных каждое основание генома было классифицировано как «полиморфное» или «консервативное». Для каждого полиморфного основания и каждого

индивидуума была рассчитана доля миноритарных аллелей по отношению к общему консенсусу. Для каждого человека и каждого основания внутрипациентное генетическое расстояние (гетерозиготность) рассчитывалась как доля случайно выбранных пар последовательностей, которые различаются по этому основанию. Генетическая дистанция между пациентами рассчитывалась таким же образом, но для пар, взятых у разных людей.

Оба генетических расстояния могут быть выражены в терминах частоты миноритарного аллеля по отношению к любой эталонной последовательности, такой как консенсус базы данных или наиболее приспособленная последовательность. Обозначим через f_{ki} частоту аллелей в основании i для индивидуума k . Тогда внутрипациентное расстояние для каждого пациента, T_{ki}^{intra} , и межпациентные генетические расстояния для пары индивидуумов, $T_{k_1k_2i}^{inter}$, могут быть записаны как

$$\begin{aligned} T_{ki}^{intra} &= f_{ki}(1 - f_{ki}) \\ T_{k_1k_2i}^{inter} &= f_{k_1i}(1 - f_{k_2i}) + f_{k_2i}(1 - f_{k_1i}) \end{aligned} \quad (2.1)$$

Как можно показать из этих формул, внутрипациентные и межпациентные генетические расстояния изменяются между 0 и 1/2 и 0 и 1 соответственно. Внутрипациентное расстояние максимально, когда популяция имеет состав 50:50 по двум аллелям, $T_{ki}^{intra} = \frac{1}{2}$, $f_{ki} = \frac{1}{2}$. Расстояние между пациентами всегда больше или равно, чем среднее расстояний внутри двух пациентов. Расстояние между пациентами максимально, $T_{k_1k_2i}^{inter} = 1$, когда две популяции вирусов однородны и состоят из противоположных аллелей: либо $f_{k_1i} = 0$, $f_{k_2i} = 1$, либо $f_{k_1i} = 1$, $f_{k_2i} = 0$. Оба генетических расстояния не зависят от выбора эталонной последовательности: правые части (2.1) инвариантны к замене $f_{ki} \rightarrow 1 - f_{ki}$. Эта инвариантность делает внутрипациентную генетическую дистанцию гораздо лучшим параметром для сравнения теории с экспериментом, чем частота замен. Действительно, как показано ниже, наиболее приспособленный вариант и вариант консенсуса не обязательно совпадают.

Оба типа расстояний, усредненные либо по людям, либо по их парам соответственно, показаны для разных участков на рисунке 2.1. Внутрипациентное расстояние показано для отдельных пациентов на рисунке 2.2. Генетическое расстояние было усреднено по трем группам оснований, для каждого индивидуума отдельно: (i) основания, полиморфные в индивидууме, (ii) основания, полиморфные у любого индивидуума, и (iii) все основания гена протеазы (табл. 2.1). Поскольку множественные замены в пределах одного и того же кодона очень редки, все мутации можно классифицировать на синонимичные (без изменения аминокислоты) или несинонимичные (с изменением аминокислоты) с отдельным рассмотрением этих двух классов (таблица 2.1). Когда два миноритарных аллеля были обнаружены в одном и том же кодоне одной

и той же последовательности, они считались принадлежащими к двум разным последовательностям.

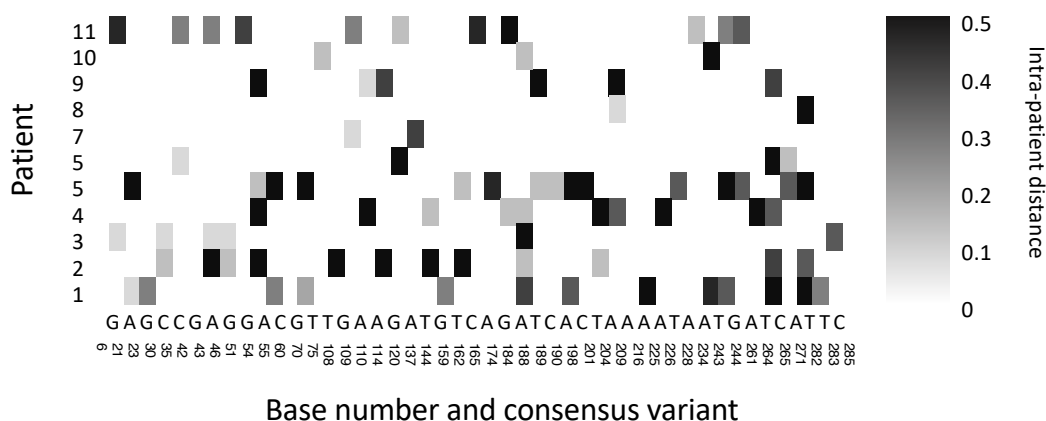


Рисунок 2.2. Диаграмма генетической дистанции внутри пациента T_{ki}^{intra} у разных пациентов на разных полиморфных основаниях гена протеазы ВИЧ. Генетическая дистанция внутри пациента указывается степенью затемнения, как показано на шкале справа. Буквы и цифры под диаграммой обозначают консенсусные варианты нуклеотидов и положения в гене протеазы соответственно. По работе [6].

Основные выводы из анализа этих данных следующие:

- (i) Генетический полиморфизм внутри человека сосредоточен на редких основаниях: 47 из 297 оснований.
- (ii) Расположение полиморфных локусов отличается между индивидуумами: в среднем такой локус появляется только у $\sim 16\%$ индивидуумов.
- (iii) Полиморфные основания очень гетерозиготны: $\langle T_{ki}^{intra} \rangle = 0.27$ на полиморфное основание, при максимуме 0.5.
- (iv) Синонимичные и несинонимичные полиморфные основания имеют одинаковое разнообразие.
- (v) Однако синонимичные полиморфные основания встречаются в два раза чаще, чем несинонимичные.

Таблица 2.1. Генетическое разнообразие гена протеазы ВИЧ по данным [53]. Внутрипациентное генетическое расстояние в протеазе T усреднено по одной из трех показанных групп оснований для каждого инфицированного, а затем усреднено по инфицированным.

Тип мутации	Среднее внутрипациентное расстояние ^a , T (%)			Число полиморфных оснований
	Индивидуальное полиморфное основание ^b	Любое полиморфное основание ^c	Любое основание ^d	
Синонимичные	29.5	4.7	0.23	31
несинонимичные	21.9	4.3	0.72	16
Синонимичные и несинонимичные	27.5	4.5	0.05	47
Спорадические синонимичные ^c			0.07	16
Спорадическая несинонимичная ^c			0.12	23
Спорадические синонимичные и несинонимичные ^c			0.23	39

^a Количество полиморфных оснований

^b Основания, которые изменчивы только у данного индивидуума.

^c Основания, которые полиморфны у какого-то человека.

^d Все основания в протеазе.

^e Замены представленные в единственной копии в базе данных и поэтому исключенные из анализа, поскольку могут быть ошибкой PCR. Обратите внимание на разницу в относительном количестве синонимичных оснований между спорадическими и неспорадическими заменами.

2.1.2 Модель 1: детерминистическая адаптация вируса у индивидуума

Ниже анализируется семейство из четырех моделей. Во всех моделях эволюция внутри хозяина рассматривается как детерминистический процесс с очищающим отбором и мутацией как единственными эволюционными факторами, в то время как стохастические факторы, такие как случайный характер мутации и случайный генетический дрейф, игнорируются [3]. Это ограничение будет снято в § 2.2. Преобладание синонимичных замен (§ 2.1.1) предполагает доминирующую роль очищающего отбора, а не отбора на разнообразие. Геномный тест, представленный в § 2.2, демонстрирует большой размер популяции и небольшую роль стохастических эффектов в хронической нелеченной ВИЧ-инфекции. Эффекты сцепления и эпистаз также не учитываются (им посвящена *глава 3*), а эволюция каждого основания рассматривается независимо от других (однолокусная модель). Предполагается, что наиболее приспособленная штамм вируса одинаков для всех людей. Это и другие предположения обсуждаются в оригинальной работе [6] и § 2.1.7. Цель настоящего исследования состояла в том, чтобы объяснить наблюдения (i)-(v).

Первое объяснение этих фактов таково: слегка вредные («мутантные») аллели на локусе (основании) генерируются у некоторых особей в результате случайной мутации и передаются по цепочке (рисунок 2.3). Из-за случайной выборки во время передачи, заражающая вирусная последовательность может содержать или не содержать мутантный аллель на данном локусе. Если человек заражен мутантным вирусом (человек А на рисунке 2.3), вирус постепенно адаптируется и возвращается к «дикому типу». В процессе адаптации, когда пропорции мутантного и дикого типа в популяции сопоставимы, локус очень гетерозиготен. Чтобы выяснить, успешно ли эта модель объясняет этот факт, процесс изучался в два этапа: внутри хозяина (Модель 1) и передача между хозяев (Модель 2). Выходные данные Модели 1 использовались в качестве входных данных для Модели 2.

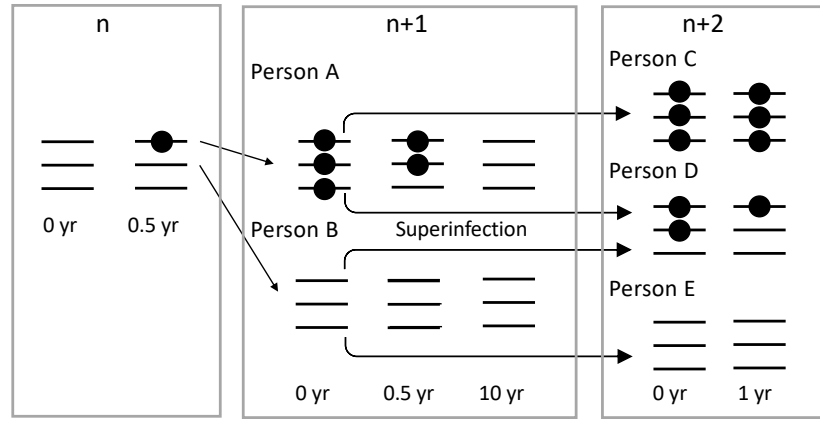


Рисунок 2.3. Модель эволюции основания по цепи передачи в условиях очищающего отбора (Модель 1). Цифры вверху обозначают циклы передачи. Кружками обозначены мутанты. Мутантная основа спонтанно появляется у человека n , который заражает человека А мутантом, а человека В - вариантом дикого типа. Человек А передает мутант по цепочке человеку С, после чего его/ее собственная популяция вируса медленно возвращается к дикому типу. У человека D, стабильно коинфицированного парой последовательностей, полиморфных по этому основанию, отбор быстро устраняет мутантный вирус. По работе [6].

Каждый полиморфный локус может иметь один из двух аллелей: А/С или G/Т. Приспособленность варианта (число репродукции) определяется как среднее количество клеток, инфицированных одной инфицированной клеткой до ее гибели. Один из двух аллелей лучше приспособлен и называется «диким типом». Локус характеризуется тремя параметрами: относительной разницей приспособленности между двумя аллелями, коэффициентом отбора s , и вероятностями прямой и обратной мутации, μ_r и μ_f . Оценки частоты мутаций ВИЧ составляют $4 \cdot 10^{-5}$ для замен $A \rightarrow G$ и $C \rightarrow T$ и в 5–10 раз ниже для обратных замен [56]. Коэффициент отбора s сильно различается между разными основаниями; его среднее число рассматривалось как подгоночный параметр.

Детерминистические уравнения для этой модели имеют вид ОДУ

$$\frac{dn_1}{dt} = (1 - \mu_r)\kappa_1 n_1 + \mu_f \kappa_2 n_2 - \frac{n_1}{t_{rep}} \quad (2.2)$$

$$\frac{dn_2}{dt} = (1 - \mu_f)\kappa_2 n_2 + \mu_r \kappa_1 n_1 - \frac{n_2}{t_{rep}} \quad (2.3)$$

Здесь n_1 и n_2 — количество невыгодных (мутантных) и выгодных (дикого типа) аллелей соответственно; κ_1 и κ_2 — их коэффициенты репликации; t_{rep} — время цикла репликации; μ_f и μ_r — вероятность (скорость) прямой и обратной мутации соответственно. Предполагается, что среднее время гибели клеток (цикл репликации) t_{rep} одинаково для двух вариантов; гибель клетки представляет собой случайный пуассоновский процесс (для рассматриваемой долговременной эволюции ни одно из предположений не имеет значения). По определению, коэффициент отбора s представляет собой относительную разность приспособленности, $s =$

$\kappa_2/\kappa_1 - 1$, которая является постоянной и удовлетворяет двойному неравенству $\mu_{f(r)} \ll s \ll 1$. Он измеряется экспериментально по экспоненциальным темпам роста.

Уравнения (2.2) и (2.3) можно свести к одному уравнению, введя долю мутантов в популяции (долю мутантов) $f = n_1/(n_1 + n_2)$. Дифференцируя f по времени с использованием (2.2) и (2.3), получаем

$$\frac{df}{dt'} = -sf(1-f) + \mu_f(1-f) - \mu_r f + O(\mu s) \quad (2.4)$$

где $O(\mu s)$ — поправка второго порядка, которой можно пренебречь, $t' \equiv \kappa_1 t$ — безразмерное время в единицах циклов репликации $1/\kappa_1 = t_{rep}$. Ниже обозначение t' заменено на t с учетом этих единиц. Решение уравнения (2.4) для общего начального условия $f(0) = f_0$ дает

$$f(t|f_0) = \tilde{\mu}_f + \frac{f_0(1 + \tilde{\mu}_f + \tilde{\mu}_r) - \tilde{\mu}_f}{f_0 - \tilde{\mu}_f + (1 - f_0 + \tilde{\mu}_r)e^{-st}} + O[(\tilde{\mu}_{f(r)}/s)^2] \quad (2.5)$$

с новыми обозначениями $\tilde{\mu}_f \equiv \mu_f/s$, $\tilde{\mu}_r \equiv \mu_r/s$.

Известно, что менее 1% половых контактов приводит к системной ВИЧ-инфекции [225-227], что подразумевает наличие очень небольшой дозы, вызывающей инфекцию. Ранняя популяция вируса через несколько дней после заражения обычно генетически однородна, что предполагает передачу одной, реже двух последовательностей вируса. В этом разделе предполагается передача одного вируса. Ко-передача двух вирусов будет рассмотрена ниже в § 2.1.4.

Случаи мутантной реверсии (адаптации вируса) и накопления мутантов задаются начальными условиями $f_{rev}(t) = f(t|1)$ и $f_{acc}(t) = f(t|0)$. Две этих функции изображены на рисунке 2.4 для реалистичного набора параметров модели. Полупериод адаптации, t_{50} , определяется условием $f_{rev}(t_{50}) = 1/2$. Используя формулу (2.5) при $f_0 = 1$ получается $t_{50} = (t_{rep}/s)\log(s/\mu_r)$. В пределе больших t , таких как $t - t_{50} \gg t_{rep}/s$, мутантная частота $f(t)$ сходится к стационарному значению $f_{ss} = \tilde{\mu}_f = \mu_f/s \ll 1$.

Предположим, заражающая популяция является 100% мутантной. Поскольку персистирующая ВИЧ-инфекция включает в себя множество циклов инфицирования клеток, а размер популяции велик (§ 2.2), варианты дикого типа будут иногда возникать из-за мутаций и постепенно расти в числе, пока популяция не станет почти полностью дикого типа (рисунок 2.4, верхняя часть). Конечная популяция будет находиться в стационарном состоянии с небольшой долей мутантов из-за баланса между отбором и вредными мутациями [58, 342]. Время до 50% состава, обозначенное как t_{50} , обратно пропорционально s и прямо пропорционально времени на цикл заражения. Время цикла примерно равно среднему продуктивному периоду инфицированной клетки перед ее гибелью, который в ранних исследованиях реакции на лекарство оценивался примерно как $t_{rep} \approx 2$ дня [55, 170, 171, 346] (позже оценка была уточнена

до 1 дня). В интервале времени $t = t_{50} \pm 2/s$, где время t измеряется в единицах t_{rep} , гетерозиготность популяции T будет превышать 0.015. Поскольку средняя нелеченная инфекция длится 10 лет, среднее время выборки находится посередине, $\langle t \rangle t_{rep} \approx 5$ лет [347]. Результат подгонки значения t_{50} для случая, когда нуклеотиды А или С соответствуют дикому типу, дает коэффициент отбора $s = 0.008$ [6]. Для инфицированных, отобранных ранее или позже 5 лет, значение s будет соответственно выше или ниже. Когда G или T представляют собой дикий тип, вероятность мутаций выше, и подгонка дает $s = 0.006$.

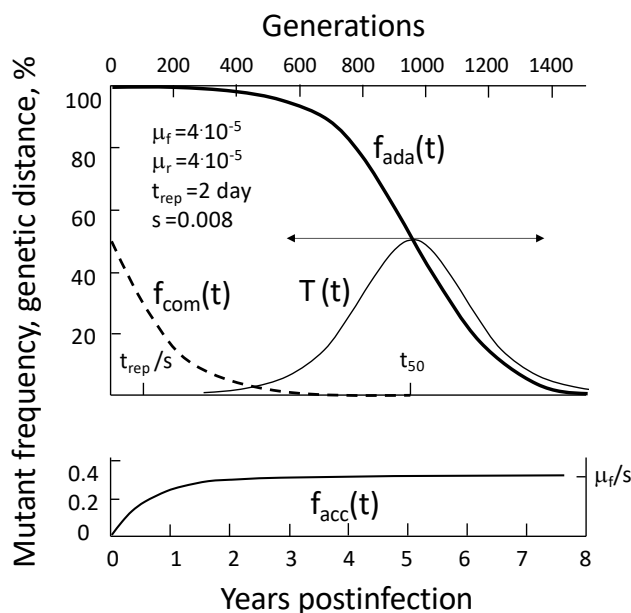


Рисунок 2.4. Зависимость частоты мутаций от времени в разных исходных популяциях. Исходно чисто мутантный (адаптация, толстая линия на верхней панели), чисто дикий тип (накопление, нижняя панель) или 50% мутантный (конкуренция роста, пунктирная линия). Более тонкая кривая показывает внутрипациентную генетическую дистанцию $T(t)$ во время адаптации. Горизонтальная полоса — это временной интервал, в течение которого база будет классифицироваться как переменная (от 5% до 95%). Показаны параметры: скорость прямой (дикий тип \rightarrow мутант) и обратной мутации, μ_f и μ_r , соответствующие либо А, либо С дикому типу; время цикла репликации, t_{rep} , и коэффициент отбора, s . Полупериод адаптации, t_{50} , определяется уравнением $t_{50} = (t_{rep}/s)\log(s/\mu_r)$. По работе [6].

Модель адаптации вирусов естественным образом объясняет большинство экспериментальных особенностей, обсуждавшихся выше:

- (i) Среднее полиморфное основание очень гетерозиготно, потому что находится в процессе адаптации.
- (ii) Вариабельные основания встречаются редко в геноме и различаются у разных людей, потому что они имеют маленькое значение s , потому что человек может быть инфицирован или не быть инфицирован мутантным вирусом, а также потому, что большинство пациентов отбираются за пределами узкого окна около t_{50} (рисунок 2.4).
- (iv) Синонимичные и несинонимичные разнообразные основания имеют примерно одинаковое разнообразие, потому что они предварительно выбраны на основе одинаково малых s .

(v) Синонимичных полиморфных оснований больше чем несинонимичных (таблица 2.1), потому что они, как правило, имеют меньшие s .

2.1.3 Цепочка одноклоновой передачи

Остается задача объяснить (сопоставить) частоту появления переменного узла у особей, 16% (табл. 2.1). Для этого в модель необходимо включить эволюцию по цепочке передачи (рисунок 2.3). Предположим, что каждый человек заражает следующего человека в фиксированное время t^* с момента своего собственного заражения. Инфицирование происходит с помощью одной последовательности вируса, образец которой выбирается случайным образом и может быть либо мутантным, либо диким типом в представляющем интерес полиморфном основании. Тогда частота полиморфных оснований у индивидуума с номером $n + 1$ в цепочке передачи равна вероятности f_n^* того, что вирус, переданный от человека с номером n , является мутантным. Среднее значение частоты мутаций у человека n в момент времени t определяется Марковской цепочкой (ветвящийся процесс)

$$f_n(t) = f_n^* f_{rev}(t) + (1 - f_n^*) f_{acc}(t) \quad (2.6)$$

Для самосогласованности f_n^* должно быть равно средней частоте мутаций у исходного индивидуума в момент времени t^* после заражения.

$$f_n^* = f_{n-1}(t^*) \quad (2.7)$$

Вместе уравнения (2.6) и (2.7) полностью описывают эволюцию по цепи передачи. Если передача происходит рано и до времени адаптации, так что $t^* < t_{50}$ & $|t^* - t_{50}| \gg t_{rep}/s$, вероятность f_n^* медленно изменяется с n , и уравнения (2.6) и (2.7) можно упростить до

$$\frac{df_n^*}{dn} = -\frac{f_n^* - f_\infty^*}{n_{eq}} \quad (2.8)$$

где

$$f_\infty^* \equiv \frac{f_{acc}(t^*)}{1 - f_{rev}(t^*) + f_{acc}(t^*)} \quad (2.9)$$

$$n_{eq} \equiv \frac{1}{1 - f_{rev}(t^*) + f_{acc}(t^*)} \quad (2.10)$$

Решение уравнения (2.8) при общем начальном условии $f^*(0) = f_0^*$ получаем

$$f_n^* = f_\infty^* \left(1 - e^{-\frac{n}{n_{eq}}} \right) + f_0^* e^{-\frac{n}{n_{eq}}}$$

Таким образом, вероятность мутации f_n^* растет с каждой передачей n , пока не достигнет значения плато f_∞^* , которое представляет устойчивое состояние передачи (рисунок 2.5а). Если время передачи t^* короче полупериода адаптации, $t^* < t_{50}$, значение плато f_∞^* намного выше, чем

стационарная мутантная частота в отдельных μ/s . Если, кроме того, скорость прямой мутации намного выше скорости обратной мутации, f_{∞}^* может быть близка к 100%.

Таким образом, при передаче одного вируса любой основе с малым коэффициентом отбора со временем станет очень разнообразным. Качественно говоря, сильное накопление мутантов по цепи передачи является комбинированным эффектом узкого места стохастической передачи, делающего начальную популяцию генетически однородной в интересующем участке, и короткого времени между передачами, не дающего достаточно времени для естественного отбора. действовать до следующего события передачи.

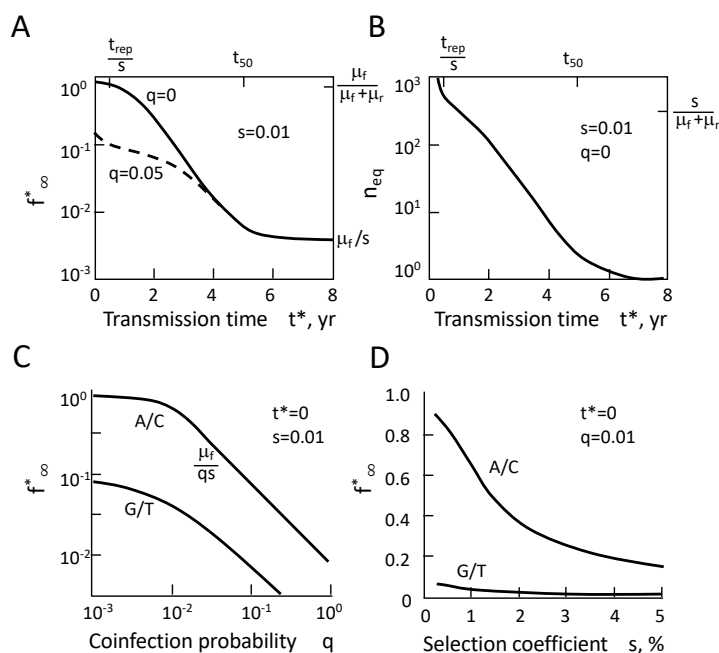


Рисунок 2.5. Предсказания модели цепной передачи. (а) Вероятность присутствия мутантного основания в инокуляте в устойчивом состоянии цепи, f_{∞}^* , в зависимости от времени передачи, t^* , уравнение (2.9). Сплошная линия, строго единичное заражение; пунктирная линия – коинфекция из двух источников с вероятностью $q = 0.05$. (б) Количество циклов, необходимых для достижения устойчивого состояния цепи, n_{eq} , в зависимости от времени передачи, t^* , уравнение (2.10). (с) Вероятность мутантной базы в течение очень короткого времени передачи, $t^* \rightarrow 0$, по сравнению с вероятностью двойной инфекции, q . Верхняя и нижняя кривые соответствуют А/С и G/Т дикого типа соответственно. (д) Та же вероятность в зависимости от s для фиксированного значения $q = 0.01$. Параметры $t_{rep} = 2$ дня; $\mu_f = 4 \cdot 10^{-5}$, $\mu_r = 4 \cdot 10^{-6}$ для дикого типа А/С и наоборот для G/Т. По работе [6].

Количество передач для достижения устойчивого состояния n_{eq} и f_{∞}^* линейно пропорциональны друг другу (рисунок 2.5а). Несмотря на то, что средний интервал передачи короткий, $t^* t_{rep} < 1$ год, для объяснения высокого наблюдаемого значения $f_{\infty}^* \sim 0.1$ требуются сотни циклов передачи. Таким образом, в рамках этой модели накопление мутантов происходило задолго до начала пандемии ВИЧ, включающей эндемическую и энзоотическую инфекцию человека у хозяина-примата.

2.1.4 Модель 3: коинфекция из независимых источников

Приведенные выше прогнозы в решающей степени зависят от предположения, что каждый человек всегда заражается из одного источника. Теперь предположим, что человек за короткое время и до дальнейшей передачи инфицирован двумя вариантами вируса, полученными от двух разных инфицированных людей, одним мутантным и другим диким типом (пациент D на рисунке 2.3). Этот сценарий суперинфекции подтверждается наблюдениями за обезьянами и людьми в течение первых недель после заражения [348, 349]. Тогда оба генетических варианта хорошо представлены в исходной популяции, и естественный отбор между двумя вирусами начинает действовать сразу, не дожидаясь появления новых мутаций (см. кривую $f_{com}(t)$ на рисунке 2.4). В результате мутантный вирус почти исчезнет в организме хозяина, прежде чем будет передан дальше. Такое редкое событие коинфекции может прервать длинную цепочку передачи мутантных аллелей. Таким образом, даже редкие коинфекции могут остановить накопление мутантов.

Предположим, что пара вирусных геномов может быть передана от двух разных людей. Вероятность этого события обозначается q . Вероятность одногеномной передачи составляет $1 - q$. Тогда уравнение (2.6) заменяется более сложным

$$f_n(t) = [(1 - q)f_n^* + q(f_n^*)^2]f_{rev}(t) + [(1 - q)(1 - f_n^*) + q(1 - f_n^*)^2]f_{acc}(t) + 2qf_n^*(1 - f_n^*)f_{com}(t) \quad (2.11)$$

где $f_{com}(t) \equiv f\left(t\left|\frac{1}{2}\right.\right)$ в (2.5) – частота мутаций в ростовой конкуренции двух штаммов (рисунок 2.4). Значение f_n^* соответствует тому же условию согласованности (2.7). В уравнении (2.11) два источника инфекции считаются статистически независимыми (эпидемиологически удаленными). Стационарное значение инокулята f_∞^* находится из квадратного уравнения

$$q(f_{acc} + f_{rev} - 2f_{com})(f_\infty^*)^2 - [(1 + q)f_{acc} + (1 - q)(1 - f_{rev}) + 2q\left(\frac{1}{2} - f_{com}\right)]f_\infty^* + f_{acc} = 0$$

где f_{acc} , f_{rev} и f_{com} оцениваются при $t = t^*$, и выбирается меньший из двух корней, поскольку больший больше 1 и не имеет биологического смысла. Значение f_∞^* как функция времени передачи t^* показано при фиксированной вероятности коинфекции $q = 0.05$ (рисунок 2.5а).

Демпфирующее влияние редких коинфекций на генетическое разнообразие особенно сильно проявляется при коротких t^* , т. е. именно тогда, когда предсказана очень высокая вероятность появления мутанта в инокуляте f_∞^* (рисунок 2.5а). Даже 5% коинфекций, $q = 0.05$, снижают значение f_∞^* на порядок по сравнению с чистой одновирусной передачей (рисунок 2.5в). Передача нескольких вариантов вируса из одного источника, как и ожидалось, не сильно влияет на накопление мутантов в цепи передачи.

2.1.5 Оценка вероятности коинфекции ВИЧ

Таким образом, прогнозируемое разнообразие гена протеазы критически зависит от вероятности коинфекции q , которую необходимо оценивать по независимым экспериментальным данным. Из-за разницы в частоте мутаций, вероятность мутантного основания в переданном вирусе f_{∞}^* намного меньше для оснований с нуклеотидами дикого типа G или T, чем для оснований с A или C (рисунок 2.5с). Если бы время наблюдения у пациентов было фиксированным, f_{∞}^* было бы равно наблюдаемой частоте у отдельных лиц, 16% (табл. 2.1). Поскольку время выборки сильно варьируется между пациентами, и ожидается, что каждое основание будет различаться в узком временном окне $t = t_{50} \pm \frac{2}{c}$, можно ожидать, что f_{∞}^* будет больше 16%.

Оценка q получена в обсуждаемой работе [6] при следующих предположениях. Изменение времени репликации вируса t_{rep} относительно невелико, и им можно пренебречь [170]. Время до СПИДа равномерно распределяется между 2 и 14 годами [347], и предполагается, что время тестирования изменяется от 1.3 до 8.7 лет. Поскольку окончательная оценка устойчива к плотности распределения s , предполагается, что она является равномерной в интервале $s \in [0.005, 0.05]$ (более реалистичное распределение будет обсуждаться в § 3.4). При этих предположениях наблюдаемая экспериментальная частота переменных участков у особей, составляющая 16%, требует $q < 0.01$ (рисунок 2.5). Это соответствует $f_{\infty}^* \sim 1$ для дикого типа A или C и $f_{\infty}^* \sim 0.1$ для дикого типа G или T (рисунок 2.5с).

Таким образом, влияние даже редких коинфекций на прогнозируемую частоту переменных оснований очень сильное: должно произойти менее 1% коинфекций из независимых источников, чтобы объяснить наблюдаемое разнообразие вирусов. Только 10% коинфекций уменьшат прогнозируемое количество переменных участков в 10 раз (см. рисунок 2.5с). Фактически, в странах, где не преобладает ни один подтип ВИЧ, коинфекции с различными подтипами ВИЧ встречаются в 10–15% случаев [350, 351], не считая случаев межсубтипной рекомбинации, где коинфекция имела место в прошлом [352, 353]. Эффект защиты от суперинфекции у ВИО-инфицированных животных наблюдается, но является только частичным [348]. Если предположить, что оба подтипа присутствуют в популяции в равных количествах, то получается $q = 0.2—0.3$ и даже больше, что составляет от 20 до 30 раз выше предказанного потолка $q < 0.01$. Уровень коинфекции, наблюдаемый в группах высокого риска, намного превышает 1%, что полностью опровергает идею накопления вредных аллелей из-за узкого места передачи. Необходимо другое объяснение.

2.1.6 Модель 4: индивидуальная изменчивость дикого типа из-за подтипов МНС

Вероятным объяснением является наличие иммунного ответа и антигенной эволюции (§§ 5.1 и 5.2). В то время, когда была создана оригинальная работа, роль иммунного ответа у ВИЧ-инфицированных людей горячо обсуждалась между иммунологами и вирусологами. Обсуждаемая работа [6] предположила, что ответ цитотоксических Т-клеток (CTL) вносит существенный вклад в борьбу с вирусом. Вскоре после ее публикации, эта точка зрения была подтверждена прямыми экспериментами по обеднению CD8 Т-клеток в экспериментально зараженных животных [211, 212, 354].

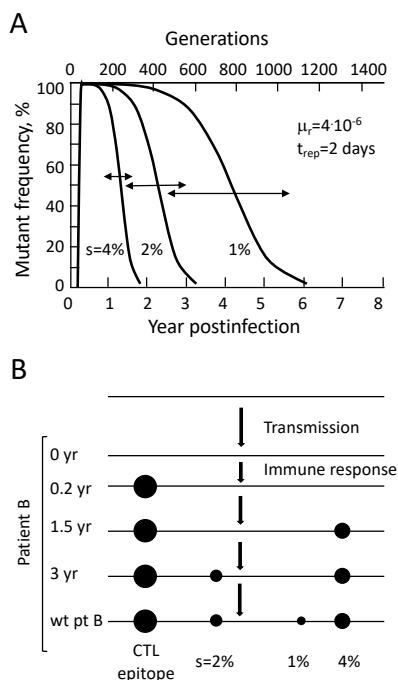


Рисунок 2.6. Модель эволюции ВИЧ, обусловленной иммунными ускользящими мутациями и компенсаторными мутациями. Быстрое начальное переключение доминантных эпитопов переопределяет дикий тип для ряда оснований, которые затем эволюционируют в сторону нового дикого типа со скоростью, обратно пропорциональной их коэффициентам отбора. (А) Зависимость от времени частоты мутаций на трех локусах, связанных с эпитопом. Три коэффициента отбора после переключения эпитопа показаны рядом с соответствующими кривыми. (В) Эволюция индивидуальной консенсусной последовательности в трех локусах. пт, пациент. По работе [6].

ВИЧ-специфические лимфоциты, в том числе CD8 Т-клетки и CD4 Т-клетки, распознают короткие участки из 8-9 аминокислот (эпитопы), расположенные повсюду в геноме [355]. По аналогии с хорошо изученными вирусами, такими как LCMV и грипп, было сделано предположение [6], что протеаза быстро накапливает антигенные ускользящие мутации в эпитопах CTL, отменяя их распознавание вирусом, а клетки иммунной памяти гарантируют, что эти изменения станут постоянными. Как только антигенный ускользящий мутант фиксируется, наиболее приспособленные аллели для ряда других оснований переопределяются благодаря эпистатическим взаимодействиям (глава 5). Эти основания теперь становятся мономорфно мутантными и начинают адаптироваться к новому дикому типу, как описано в модели 1. Некоторые замены не меняют аминокислот, и для них естественный отбор действует на уровне

мРНК или транскрипции. Позже данная модель была фактически подтверждена подробными наблюдениями за антигенным ускользанием в эпитопах CTL [18, 280, 286, 291, 356]. Динамика такого ускользания рассмотрена в § 7.2, а эпистаз в *главе 5*.

Расположение эпитопов CTL в геноме определяется индивидуальной иммуногенетикой, а именно шестью подтипами HLA, каждый из которых существует в сотне вариантов [355]. Это объясняет, почему перекрывание переменных оснований у пациентов лишь частичное. Это происходит среди пациентов, которые имеют некоторые общие подтипы HLA. Аллель, являющийся более приспособленным для одного человека, в котором он вне CTL эпитопов, для другого является менее приспособленным, после начала иммунного ответа, так как он случайно попадает в какой-нибудь CTL эпитоп. В такой модели, частота переменных оснований отражает степень перекрытия подтипов HLA и частоту эпитопов для каждого подтипа HLA в геноме. Всего у каждого пациента активны 20-30 эпитопов [18, 280, 286, 291, 356].

Большинство ускользающих мутаций в эпитопах фиксируются быстро, в течение нескольких месяцев после инфицирования. Эти фиксированные мутации снижают репродукцию вируса, что приводит к мутациям компенсации вне эпитопов, где адаптация происходит с разной скоростью (рисунок 2.6). Фиксация компенсаторных локусов с большей степенью компенсации (большей s) происходит раньше, а с меньшей s - позже. В период от месяцев до лет средние s переменных локусов уменьшаются во времени, замедляя средний темп эволюции (§ 3.4). Происходит постепенный переход от несинонимичных к синонимичным переменным локусам, поскольку вторые, как правило, имеют меньшие s . В то же время, средняя гетерозиготность любого переменного локуса остается на том же высоком уровне (таблица 2.1). Оболочечный ген ВИЧ *env*, который связывают нейтрализующие антитела, проявляет противоположный эффект: соотношение числа синонимичных к несинонимичным локусам снижается на ранних стадиях инфекции, признал отбора на разнообразие [45, 46]. Динамика ускользания антигена от ответа антител подробно обсуждается в § 5.1.

Предложенная модель эволюции протеазы как следствие иммунного ускользания с компенсацией похожа на эволюцию мутаций, придающих устойчивость к лекарственному лечению, которые состоят из первичных ускользания, и вторичных компенсирующих. Например, в работе [357], после 32-60 недель лечения группы пациентов индинавиром, ингибитором протеазы, индивидуальный консенсус по протеазе изменился в 10 локусах на пациента, четверть из которых были синонимичны (разница с данными в таблице 2.1 происходит, вероятно, из-за разницы во времени после инфекции).

2.1.7 Заключение

Таким образом, показано, что распределение аллелей в относительно консервативном гене ВИЧ между пациентами и внутри них согласуется с реверсией вредных мутаций с небольшими коэффициентами отбора $s \sim 0.005$. Происхождение этих вредных аллелей исследовано с использованием семейства моделей. Модель которая объясняет главные факты предполагает, что большинство этих оснований компенсируют стоимость ранних мутаций «антигенного ускользания», снижая распознавание вируса иммунным ответом. Более поздние экспериментальные работы эту модель подтвердили [18, 280, 286, 291, 356].

Приведенный выше анализ использовал приближения: (i) игнорирование трансверсий в данных, (ii) однородное распределения вируса и клеток, (iii) независимые источники коинфекции. Уместность этих приближений, в данном контексте, была объяснена в оригинальной работе [6]. Кроме того, в анализе не учитывались стохастические факторы эволюции вируса в хозяине, а также эпистаз, генетическое сцепление и рекомбинация. Эти важные эффекты детально обсуждаются в § 2.2 и главах 3-5.

2.2 Оценка эффективного размера популяции

Случайные мутации и естественный отбор представляют собой основные факторы модели эволюции с одним локусом (рисунок 2.7а). Дополнительным фактором, который не учитывался выше в § 2.1, является случайный генетический дрейф, происходящий из-за случайной вариации числа потомков между генетически идентичными вирусами. Как хорошо известно в популяционной генетике [33], относительная важность этих трех факторов зависит от размера популяции, т.е. количества особей или зараженных клеток, обозначенного N . Главные факторы эволюции существенно разнятая отличаются между тремя широкими интервалами значений N с приблизительными границами, заданными обратным коэффициентом отбора $1/s$ и обратной вероятностью мутаций на локус $1/\mu$:

(i) В самых больших популяциях, $N \gg 1/\mu$, преобладают мутация и естественный отбор, а случайным генетическим дрейфом можно пренебречь. Поскольку многие мутационные события усредняются, мутация действует как детерминистический фактор, и вся динамика почти детерминистическая, с небольшими флуктуациями. Этот случай был рассмотрен выше в § 2.1.

(ii) В наименьших популяциях, $N \ll 1/s$, можно, наоборот, пренебречь естественным отбором и применить стохастические селективно-нейтральные модели [343]. Полиморфная (генетически разнообразная) популяция меняет свою генетическую композицию только благодаря случайному генетическому дрейфу. Мутация случается редко и важна только в

генетически однородных популяциях, где ее роль заключается в создании новых аллелей и переходе популяции в полиморфное состояние.

(iii) Широкий промежуточный интервал $1/s \ll N \ll 1/\mu$ имеет смешанные свойства в зависимости от его полиморфизма: случайный дрейф контролирует популяции с низким полиморфизмом, более полиморфные популяции контролируются естественным отбором, а однородные популяции остаются однородными до мутации, что остается редким событием [33].

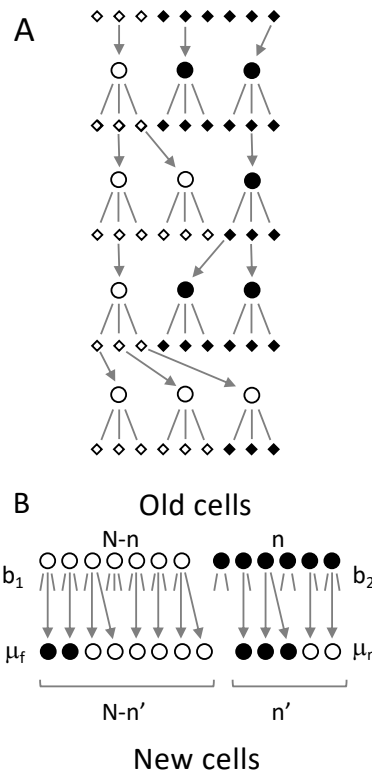


Рисунок 2.7. Три фактора эволюции стохастической однолокусной двухаллельной модели. А. Случайный дрейф генетического состава из-за выборки заражающих вирионов. Кружки обозначают инфицированные клетки; маленькие ромбы представляют собой свободные вирусные частицы (вирионы). Два генетических варианта вируса показаны черным и белым цветом. В. Модель вирусной популяции, включающая факторы эволюции: случайный дрейф, естественный отбор и мутация. Показаны два последовательных поколения инфицированных клеток. Линии, исходящие из кругов, обозначают вирионы, производимые инфицированными клетками, некоторые из которых (стрелки) заражают новые клетки. Клетка, зараженная мутантным вирусом (черный кружок), оставляет меньше инфекционного потомства, чем клетка дикого типа (белый кружок). По работе [5].

Другим важным биомедицинским вопросом был вопрос о том, достаточно ли велика эффективная численность ВИЧ-популяции, чтобы содержать устойчивые к лекарственным средствам мутации к антиретровирусным моноингибиторам до применения лекарства (§ 2.1). Среднее количество таких мутаций равно $N(\mu/s)$, что можно оценить из баланса мутации и отрицательного отбора (рисунок 2.4, внизу), где s цена такой мутации. Положительный ответ, $N(\mu/s) > 1$, мгновенно объяснил бы быстрое появление резистентности к любой монотерапии, чем так прославился ВИЧ [2]. Рассмотренная ниже работа [5] ответила и на этот вопрос.

В работе [5], детерминистическая модель с одним локусом обобщена с учетом стохастических эффектов и на случай с двумя локусами. Затем разработан тест на маленькую популяцию и применен к опубликованной базе данных последовательностей ВИЧ для гена протеазы *pro* [53] и гена оболочки *env* [358]. Анализ демонстрирует эффективный размер популяции порядка $10^5 - 10^6$ инфицированных клеток или более, что и отвечает на оба заданных вопроса.

2.2.1 Однолокусная модель стохастической эволюции

Исходная модель такая же, как и в § 2.1, но теперь она включает стохастический фактор генетического дрейфа. Популяция вируса представлена N клетками (рисунок 2.7b), инфицированными либо вирусом дикого типа, либо мутантным вирусом. Каждая клетка производит некоторое количество инфекционных вирусных частиц, а затем умирает. Среднее число потомков между двумя вариантами (аллелями) незначительно различается на относительную величину s , что создает естественный отбор (рисунок 2.7b). Чтобы учесть случайный генетический дрейф, N вирионов, заражающих следующее поколение инфицированных клеток, выбираются случайным образом из гораздо большего числа произведенных вирионов. Общее количество зараженных клеток остается равным N . При заражении новой клетки, аллель может, с малой вероятностью μ , мутировать в противоположный генетический вариант. Частота мутаций у ВИЧ находится в интервале $\mu = (0.4 - 4) \cdot 10^{-5}$ на основании за инфекционный цикл, точное значение зависит от пары нуклеотидов и направления замены [56]. Прямое обобщение той же модели для двух локусов и четырех вариантов (гаплотипов) приведен ниже в § 2.2.3.

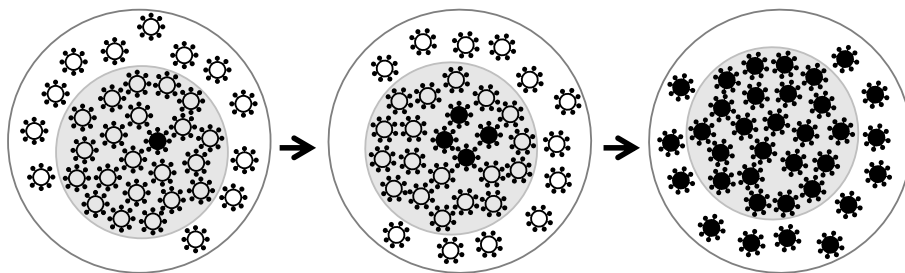


Рисунок 2.8. Эффективная численность населения может быть намного меньше, чем численность населения по переписи. Полезные вирусные мутанты (черные) возникают в эффективной субпопуляции вируса (N_{eff} , серый круг) и постепенно распространяются на всю переписную популяцию (белый круг). По работе [35].

Частота менее приспособленного аллеля в популяции незначительно меняется между последовательными поколениями из-за комбинированного эффекта (i) естественного отбора, (ii) случайного дрейфа из-за случайного выбора заражающих вирионов и (iii) случайной мутации.

Из-за последних двух факторов, эволюционная динамика системы, вообще говоря, не может быть описана детерминистическими ОДУ (2.2) и (2.3). Следовательно, эволюционную динамику нельзя предсказать с уверенностью, даже если известны начальные условия с бесконечной точностью. Тем не менее, возможно и полезно вычислить вероятность обнаружения мутантной частоты в интервале значений $[f, f + df]$ в данный момент времени t . Простейшим способом изучения эволюции является моделирование описанного выше процесса с помощью генератора псевдослучайных чисел (моделирование Монте-Карло). Более общим аналитическим методом является уравнение Колмогорова [67, 343], представляющее частный случай марковского процесса. Уравнение Колмогорова является случаем уравнения Фоккера-Планка, используемого в статистической физике. Его можно решить в различных конкретных ситуациях и проверить с помощью моделирования методом Монте-Карло [см. обзор [33]]. Настоящий параграф основан на компьютерном моделировании.

2.2.2 Три режима эволюции

Как уже упоминалось во введении, ведущие факторы эволюции различаются между тремя широкими интервалами численности популяции N с границами $1/s$ и $1/\mu$ [33]. Например, отбор пренебрежимо мал, если $N \ll 1/s$ (нейтральный предел). Случайный генетический дрейф и случайность времени мутации являются небольшими поправками, если $N \gg 1/\mu \sim 10^5$ (детерминистический предел). Между этими двумя пределами находится широкий интервал размеров популяции, $1/s \ll N \ll 1/\mu$, со смешанными свойствами. В этом режиме случайный дрейф доминирует в динамике слабополиморфных популяций, тогда как высокополиморфные популяции являются квазидетерминистическими, с доминирующей силой естественного отбора. Характеристическое число копий, называемое «стохастическим порогом», равно $1/s$.

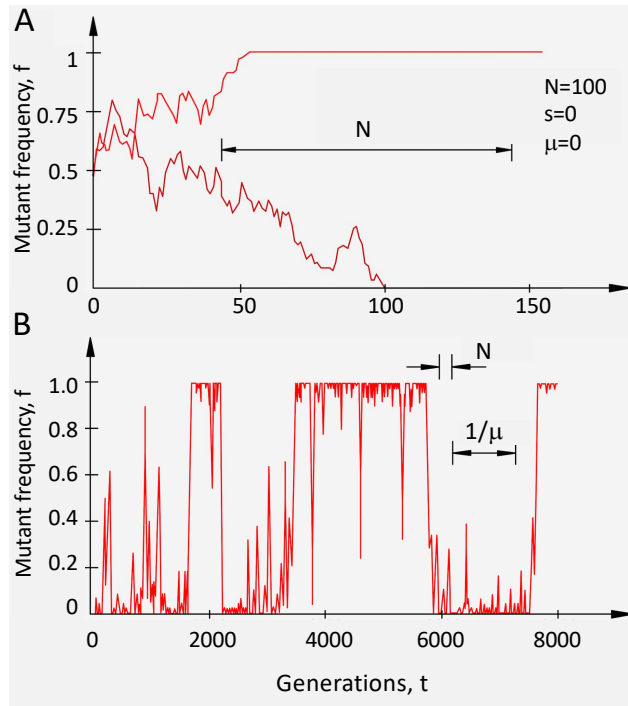


Рисунок 2.9. Зависимость частоты мутантов от времени в нейтральном пределе, $N \ll 1/s$, полученная методом Монте-Карло. А. Два репрезентативных цикла Монте-Карло в эксперименте с конкуренцией роста, $f(0) = 1/2$. В. Опыт накопления/адаптации в долговременном масштабе, $f(0) = 0$. Последняя временная зависимость была получена при $\mu = 10^{-3}$ и $N = 50$, а затем масштабирована по горизонтальной оси, чтобы соответствовать показанным значениям μ и N . По работе [5].

При заданных начальных условиях, типичная стохастическая траектория мутантной частоты может быть смоделирована по правилам модели, указанной выше. В конце концов, популяция достигает динамического устойчивого состояния, и ее статистические свойства более не меняются. То, как происходит этот переход, зависит от интервала N . Рассмотрим три вида начальных условий: (i) 100% дикий тип, (ii) 100% мутант, (iii) 50–50%. Соответствующие мысленные эксперименты, определяемые этими начальными условиями — это накопление вредных мутантов, адаптация (увеличение средней приспособленности со временем) и эксперимент по конкуренции роста.

В детерминистическом пределе, $N \gg 1/\mu$, динамика в трех «экспериментах» может быть рассчитана из (2.2) и (2.3) (рисунок 2.4). Стационарная частота мутантов равна μ/s . Динамика накопления мутантов и стационарность подтверждается прямым моделированием в этом режиме (рисунок 2.10с). Время перехода в стационарное состояние определяется как $1/s$ для накопления и конкуренции. Адаптация занимает больше времени, $t_{50} = (1/s)\log(s/\mu) \gg 1$.

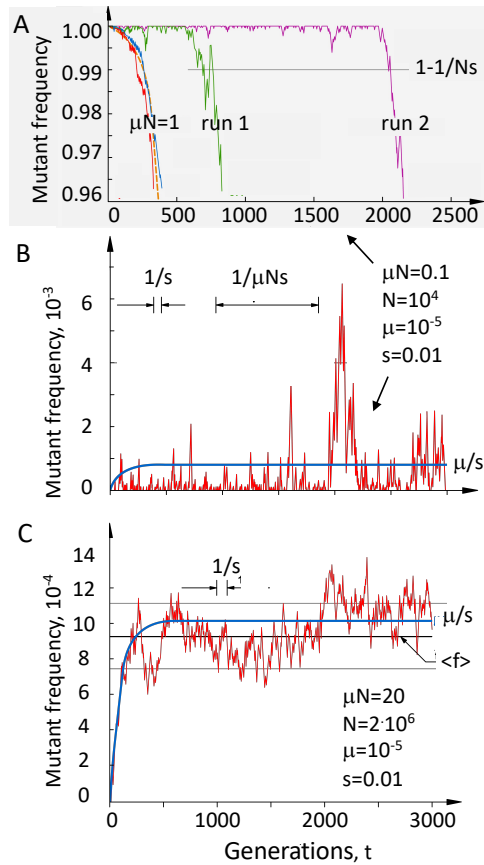


Рисунок 2.10. Моделирование зависимости частоты мутаций от времени при размерах популяций $N \gg 1/s$. (A) Начальная часть адаптационного эксперимента в режиме отбор-дрейф, $1/s \ll N \ll 1/\mu$. Показаны два репрезентативных цикла Монте-Карло (фиолетовые и зеленые кривые). Для сравнения показаны два прогона при $N = 1/\mu$ (синяя и красная кривые). Они очень близки к детерминистическому результату (штриховая кривая). (B) Эксперимент по накоплению мутантов в режиме отбор-дрейф. (C) Эксперимент по накоплению в квазидетерминистическом режиме, $N \gg 1/\mu$. (A-C) Гладкие кривые соответствуют пределу $N = \infty$. По работе [5].

В противоположном нейтральном пределе $N \ll 1/s$, когда естественный отбор не важен, конкуренция роста выглядит как рваная кривая, сильно зависящая от реализации (рисунок 2.9а). С равной вероятностью, один из двух аллелей вымирает из-за случайного дрейфа, в среднем, на шкале времени $t \sim N$. Затем, мутация возвращает вымерший аллель в популяцию, но затем он снова вымирает. После многих попыток, одному аллелю, наконец, удастся захватить всю популяцию, другой аллель вымирает, и так далее. Полученная временная зависимость переключается между двумя мономорфными состояниями в случайные моменты времени (рисунок 2.9b).

Промежуточный интервал, $1/s \ll N \ll 1/\mu$, имеет общие черты с обоими крайними режимами (рисунок 2.10). Моделирование конкуренции роста мало чем отличается от детерминистического случая (рисунок 2.4), за исключением небольших колебаний. Однако фиксация полезного аллеля сильно задерживается по сравнению с детерминистическим временем t_{50} , а случайная задержка подчиняется статистике Пуассона со средним временем задержки $\sim 1/(\mu Ns)$ (рисунок 2.10а). Задержка вызвана временем ожидания для создания одной

копии полезного аллеля, $1/(\mu N)$, и малой вероятностью s новой субпопуляции, достигающей стохастического порога [6]. Существует приблизительный критический размер $1/s$, выше которого случайный дрейф слабее естественного отбора. Если и как только клон достигнет такого размера, он будет расти быстро и «почти детерминистическим» образом (рисунок 2.10а). Подобные шкалы времени и размера клона появляются в детерминистическом накоплению мутантов (рисунок 2.4 внизу). Как и в нейтральном режиме при малых размерах популяции, большую часть времени популяция остается генетически однородной.

2.2.3 Двухлокусная модель и тест неравновесия по сцеплению

Как видно из вышеприведенных примеров, роль стохастических факторов зависит от размера популяции N . Для оценки его эффективного значения у среднего нелеченого пациента, следующей задачей являлось проведение следующего теста геномных данных с использованием двух-локусной модели .

На ранних стадиях инфицирования, популяция ВИЧ однородна или почти однородна генетически из-за узкого места передачи и ранней селекции [349, 358-360]. Сильно разнообразными основаниями являются те, которые адаптируются к хозяину [6] (§ 2.1). Выбираются два разных основания (локуса), и все геномы в популяции классифицируются в четыре группы (гаплотипы): ab , Ab , aB и AB , где строчные и прописные буквы обозначают менее приспособленные и более приспособленные аллели соответственно. Исходная популяция состоит из однородного гаплотипа ab . Конечная популяция равномерно AB . Гаплотипы Ab , aB хорошо представлены лишь в процессе перехода. Когда популяция мала, $N\mu \ll 1$, и, при ограниченном размере выборки, один из четырех классов гаплотипов будет пустым, поскольку адаптация на каждом локусе начинается в случайное время (рисунок 2.10а).

Смоделированные частоты гаплотипов моделируются в двухлокусной модели как функция времени, когда оба коэффициента отбора одинаковы. В детерминистическом пределе $N\mu \gg 1$, все четыре гаплотипа хорошо представлены в едином временном интервале (рисунок 2.11а).

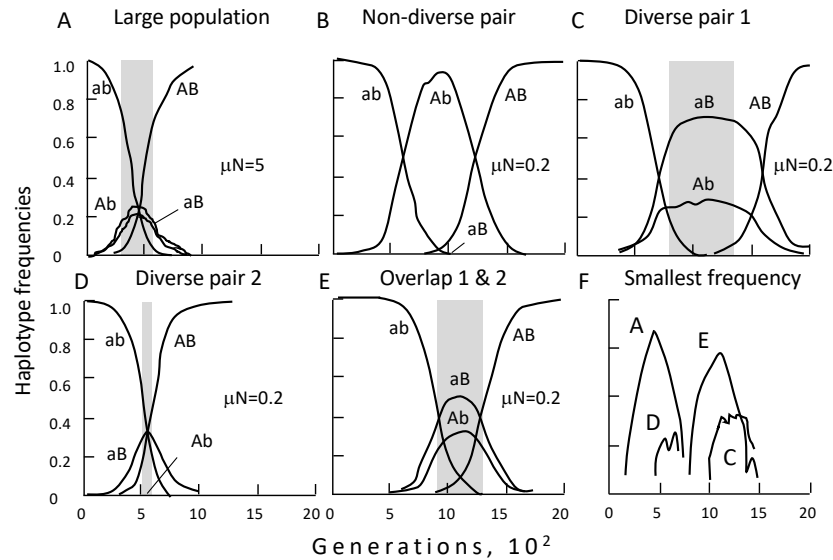


Рисунок 2.11. Компьютерное моделирование адаптации на двух участках в режиме отбор-дрейф, $1/s \ll N \ll 1/\mu$. А–Е. Частоты четырех гаплотипов. А и В обозначают первый и второй локус, соответственно, А и а обозначают дикый тип и мутант. Коэффициент отбора s одинаков для двух локусов. Значения параметров: все панели получены при $s = 0.1, N = 5000$ и перемасштабированы по оси времени, чтобы соответствовать $s = 0.01$. Частота мутаций $\mu = 10^{-3}$ на (а) и $4 \cdot 10^{-5}$ на (В–F). Прогоны, подобные показанному на (В), встречаются чаще всего, случай (С) встречается реже, а случаи (D) и (E) встречаются реже всего. Серым цветом показан временной интервал, в котором оба локуса пары имеют частоту мутаций в интервале 25-75%. (F) Временная зависимость наименьшей частоты гаплотипов для четырех прогонов в случаях (А) и (С–E). По работе [5].

Однако если популяция мала, $N\mu \ll 1$, два локуса обычно адаптируются в разные случайные моменты времени, даже при одинаковых коэффициентах отбора (рисунок 2.10b). Почти одновременная адаптация может происходить, с небольшой вероятностью, по одному из двух следующих сценариев. В Сценарии I, две мутации на двух локусах генерируют два гаплотипа, Ab и aB. Случайно, размер обоих клонов достигает стохастического порога $1/s$ примерно в одно и то же время. Затем третья мутация в одном из двух клонов порождает новый гаплотип АВ, которому впоследствии повезет достичь стохастического порога (рисунок 2.11c). В Сценарии II, мутация в ab создает либо Ab, либо aB, которому повезет вырасти выше стохастического порога. Затем, очень скоро после того, как это произойдет, вторая мутация в этом новом клоне породит гаплотип АВ (рисунок 2.11d). В любом сценарии, в любой момент времени, имеется не более трех хорошо представленных гаплотипов одновременно (рисунок 2.11c и d). Четвертый гаплотип может быть хорошо представлен только в том случае, если сценарии I и II перекрываются, что имеет очень малую вероятность (рисунок 2.11e).

Для проверки гена протеазы [53] на этот эффект «отсутствующего гаплотипа», среди его геномов из нелеченных пациентов были найдены все пары оснований с частотой мутаций в интервале 25–75%. Для каждой пары, гаплотипы были подсчитаны, в соответствии с консенсусным или антиконсенсусным аллелем по каждому основанию. (Тест не зависит от консенсуса.) В трех парах оснований из четырех исследованных, присутствовали все четыре

гаплотипа (таблица 2.2). Тест был повторен на последовательностях белка оболочки *env* от нелеченых пациентов из работы [358]. И опять таки, четвертый гаплотип отсутствовал только в двух парах из шести (таблица 2.2). Следовательно, эффективная популяция ВИЧ достаточно велика, чтобы находиться в детерминистическом режиме или на его границе, $\mu N \geq 1$.

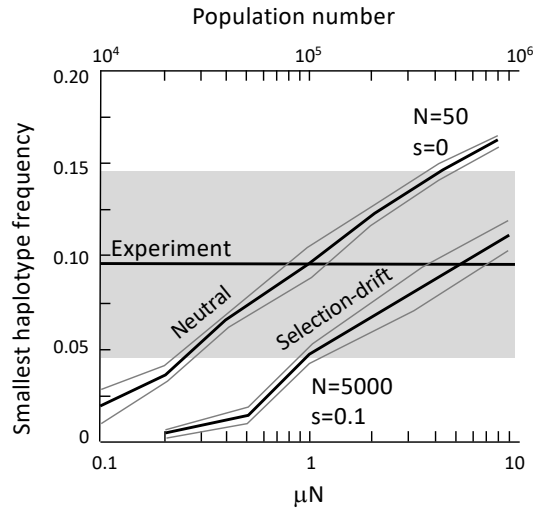


Рисунок 2.12. Предсказанная зависимость средней частоты наименее представленного гаплотипа от численности популяции. Переход от детерминистического $N \gg 1/\mu$ к стохастическому $N \ll 1/s$ режиму показан для двух значений коэффициента отбора: $s = 0.1$ и $s < 10^{-5}$. Для усреднения используются только прогоны моделирования, в которых генетический состав на каждом участке находится в интервале 25–75% во временном интервале (рисунок 2.10с, d), при этом разные прогоны взвешиваются соответственно длине этого интервала. временной интервал (т. е. тот же критерий отбора, что и в эксперименте). Толстые линии — это среднее значение; тонкие линии показывают область достоверности 95%. Зависимость от μN получена путем варьирования μ при фиксированной численности популяции, $N = 5 \cdot 10^3$ для режима отбора-дрейфа и $N = 50$ для нейтрального режима. Числа популяции, показанные на верхней горизонтальной оси, соответствуют фиксированной частоте мутаций $\mu = 10^{-5}$. Толстая горизонтальная линия и заштрихованная полоса представляют собой экспериментальное среднее значение и область достоверности 95%, полученную из данных в таблице 2.2. По работе [5].

Чтобы получить количественную границу размера популяции снизу, была рассчитана средняя доля наименее распространенного гаплотипа, для разных размеров популяции N , и проведено сравнение с наблюдаемыми данными (таблица 2.2). Зависимость от времени доли наименее представленного гаплотипа для прогонов Монте-Карло типичных для различных сценариев (рисунок 2.11а–d) показано на рисунке 2.11е. Затем эта величина усреднялась по 100 прогонам моделирования при разных N (рисунок 2.11 и его подпись).

Соответствующая экспериментальная оценка была получена путем объединения данных по генам протеаз и оболочечных генов (таблица 2.2). Было использовано значение $\mu = 10^{-5}$, которое представляет собой логарифмическое среднее между скоростью $A \rightarrow G$ или $C \rightarrow T$ и скоростью противоположных переходов [56]. На графике зависимости результата от размера популяции, размер популяции определенно ($p = 0.05$) больше, чем $9 \cdot 10^4$ инфицированных клеток. Наиболее вероятное значение больше, чем $5 \cdot 10^5$ инфицированных клеток (рисунок 2.12).

Таблица 2.2. Распределение последовательностей среди четырех гаплотипов для нескольких сильно различающихся пар оснований в геноме ВИЧ.

Номер оснований	Пара аллелей				Размер выборки
	а-а	а-к	к-а	к-к	
Ген протеазы					
21, 174	1	6	6	1	14
21, 201	4	3	3	4	14
114, 209	0	6	14	3	23
174, 201	2	5	5	2	14
Ген оболочки					
-17, -2	3	4	3	5	15
-2, 3	0	6	4	5	15
-17, 3	1	6	3	5	15
3, 10	0	4	5	6	15
-2, 10	2	4	3	6	15
-17, 10	2	5	3	5	15

Три образца из 14, 23 и 15 геномов были изолированы из трех ВИЧ-инфицированных. Последовательности белков протеазы и оболочки (область V3) были получены в работах [53] и [358] соответственно. Номера доступа в GenBank для [358] были M84240–M84314. Буквы «а» и «к» в первой строке обозначают антиконсенсус и консенсус выборки. Номера оснований в первом столбце для протеазы *pro* представляют собой стандартные положения нуклеотидов; для оболочки *env* это позиции кодонов, отсчитываемые от короны GPG петли V3 (первому G присвоен номер 0).

2.2.4 Влияние рекомбинации на оценку размера популяции

ВИЧ имеет механизм рекомбинации [361, 362], так что отсутствующий четвертый гаплотип мог быть получен путем рекомбинации, даже если популяция невелика. Рекомбинация перераспределяет уже существующие аллели между последовательностями. Чтобы создать разнородную пару локусов, две мутации должны генерировать гаплотипы Ab и aB в популяции ab (рисунок 2.10с). Затем рекомбинация может генерировать гаплотип AB из гаплотипов Ab и aB . Все четыре гаплотипа будут присутствовать одновременно, только если это событие произойдет достаточно рано, до того, как клон ab вымрет (рисунок 2.10е). По сути, в модели с двумя локусами рекомбинация эквивалентна второй мутации, $Ab \rightarrow AB$ или $aB \rightarrow AB$. Таким образом, графики, представленные на рисунке 2.11, описывают и случай рекомбинации, за исключением того, что частоту мутаций μ следует заменить ее эффективным значением, включающим рекомбинацию

$$\mu_{\text{eff}} = \mu + \frac{1}{8} \langle 4f_{Ab}f_{aB} \rangle \rho L f_{\text{coinf}} \quad (2.12)$$

где ρ — скорость рекомбинации на основании за цикл репликации, L — расстояние между двумя локусами вдоль генома, f_{Ab} и f_{aB} — частоты одиночных мутаций, f_{coinf} — доля коинфицированных клеток среди продуктивно инфицированных клеток, $\langle \dots \rangle$ обозначает усреднение по реализациям, а $1/8$ представляет собой объединенную вероятность того, что пара

провирусов в клетке представляет собой Ab и aB , гетерозиготную пару геномов РНК, упакованных в вирион, и полученный рекомбинант представляет собой AB , а не ab .

Параметры в формуле (2.12) оцениваются следующим образом. Средняя скорость рекомбинации составляет $\rho = 4 \cdot 10^{-5}$ на основании за цикл репликации [361]. Среднее расстояние между локусами пары в геноме в таблице 2.2 составляет $L = 71$ основание. Параметр f_{coinf} на момент работы [5] был неизвестен и оценивался косвенно, следующим образом.

Скорость оборота Т-клеток у человека и обезьян показывает ежедневную замену 2–5% Т-клеток у хронически инфицированного человека, с текущим количеством клеток CD4 200/мл крови, что составляет 10^9 клеток в день [180, 361]. Среднее количество вируспродуцирующих клеток составляет $4 \cdot 10^7$ клеток [363]. Средняя продолжительность жизни такой клетки оценивается в 1-2 дня [55, 170, 171]. Таким образом, вероятность инфицирования клетки составляет менее 3%. Если предположить, что инфицированные клетки выбраны случайным образом, и зная, что время до экспрессии Nef и дерегуляции рецептора CD4, в течение которого клетка остается перmissive для вируса после заражения, не более 12 часов, получается, что доля дважды инфицированных клеток среди вируспродуцирующих клеток не превышает

$$f_{\text{coinf}} = \left(\frac{(1 - 2) \text{ day}}{0.5 \text{ day}} \right) \left(\frac{3\%}{2} \right) = (3 - 6)\%$$

даже при отсутствии суперинфекционной резистентности. Подставив значения ρ , L и f_{coinf} в уравнение (2.12) и оценивая среднее значение $\langle 4f_{Ab}f_{aB} \rangle$ из рисунка 2.10 как 0.5, мы получаем что рекомбинация увеличивает эффективную частоту мутаций менее чем в два раза. Поэтому рекомбинация существенно не меняет оценки эффективного размера популяции. Более точная оценка дважды инфицированных клеток на основе генетических данных (§ 2.3) показывает, что, на самом деле, f_{coinf} еще меньше, (1 – 2)%.

2.2.5 Заключение

Эта работа показала, что эффективный размер популяции у среднего нелеченного ВИЧ-инфицированного человека велик, а эволюция отдельных оснований в основном детерминирована. Рассмотрение выше было основано на четырех приближениях, надежность которых была тщательно проверена в обсуждаемой работе [5]. Самое главное, что эта работа впервые показала, что популяция достаточно велика чтобы содержать резистивные аллели на одном локусе, еще до начала лечения.

Данный результат сильно отличается от предыдущих оценок [57], где был применен «тест нейтральности» Тажимы [364] к последовательностям гена оболочки *env* [358]. Этот тест

измеряет зависимость числа полиморфных локусов от размера образца и сравнивает с результатом нейтральной теории. Автор [57] получил очень малую среднюю популяцию ВИЧ, $N_{\text{eff}} = 100 - 1000$ и заключил, что естественный отбор не играет значения в эволюции ВИЧ в нелеченых пациентах. Недостатки теста Тажимы в том, что предсказание не является статистически отличным от предсказаний альтернативных моделей, тестируемая зависимость очень медленная, а предсказанная теорией статистическая ошибка порядка среднего значения. На самом деле, естественный отбор сильно уменьшает количество полиморфных оснований, что в этом тесте неправильно интерпретируется, как небольшой размер популяции.

Оценка $N_{\text{eff}} \sim 10^6$ зараженных клеток была позже подтверждена независимыми методами, такие как измерение задержки фиксации полезных аллелей [154] и степени гетерозиготности близких областей генома, сцепленных с полезными аллелями [35, 365]. Нелинейное накопление генетического разнообразия также указывает на присутствие эффектов естественного отбора и большую численность популяции [366].

Вывод об относительно слабой роли стохастических эффектов в стационарной популяции ВИЧ ограничен эволюцией отдельных оснований внутри нелеченого хозяина, если предположить, что такая двух локусная модель применима. На самом деле, из-за многолокусной природы эволюции ВИЧ, стохастические эффекты всегда важны для наиболее приспособленных геномов, которые определяют скорость адаптации вируса (*глава 3*). Тем не менее, оценка эффективного размера популяции в среднем нелеченом пациенте $\geq 10^6$ никаких сомнений больше не вызывает [367].

2.3 Оценка вероятности рекомбинации ВИЧ

Целью следующей работы [15] была оценка эффективной вероятности рекомбинации генома ВИЧ с другим геномом, у нелеченых пациентов. Для оценки вероятности рекомбинации r по геномным данным, использовалась компьютерную модель, включающую следующие факторы: мутация, естественный отбор с распространенным коэффициентом, случайный генетический дрейф, генетическое сцепление и рекомбинация. Значение r для типичного нелеченных инфицированного человека было оценено используя смоделированные последовательности, геномные данные из пациентов, а также две чувствительные меры генетического сцепления.

2.3.1 Методы

Модель населения

Модель, реализованная в виде алгоритма Райта-Фишера [15] <https://github.com/irouzine/Strong-linkage-in-sex>, представляет собой обобщение двухлокусной модели из § 2.2 на произвольное большое количество локусов. Эволюционные факторы — естественный отбор, случайная мутация со средней частотой μL на геном и случайный генетический дрейф [33, 368]. Рассмотрим популяцию из N бинарных последовательностей $[K_i]$, например, 0111010101, где каждый эволюционирующий локус с номером $i = 1, 2, \dots, L$ является положением нуклеотида или аминокислоты. Рассмотрен случай двух возможных аллелей: $K_i = 0$ или $K_i = 1$. В реальной жизни, бинарное предположение справедливо в среднесрочной эволюции, когда для большинства локусов наблюдаются только два варианта в течение периода времени. Геном предполагается длинным, $L \gg 1$. В случае ВИЧ, каждый геном соответствует провирусу ВИЧ, интегрированным в клеточную ДНК. Для хронической ВИЧ инфекции, смена поколений происходит за 1 день [234, 369].

В каждом поколении, все особи умирают (вирусы исчезают) и заменяются идентичными им потомками, так что общее число геномов остается постоянным. Число потомков генома является случайным, а среднее значение определяется относительной приспособленностью $\exp(W)$ которая зависит от количества полезных аллелей

$$W = \sum_{i=1}^L s_i K_i$$

В каждом поколении, все особи заменяются их потомством, численность которого случайна и рассчитывается с использованием алгоритма сломанной палочки (полиномиальное распределение), учитывающего как естественный отбор, так и генетический дрейф, следующим образом. В частности, интервал $[0, N]$ разбивается на N интервалов, длина которых пропорциональна среднему числу потомков соответствующих отдельных геномов, и на всем интервале генерируются N случайных точек. Количество точек, попадающих в интервал i , равно номеру генома i . Таким образом, общая популяция остается постоянной между поколениями и равна N .

Более простой версия распределения числа потомков - это модель клеточного деления, в которой число потомков равно 0 или 2 с разными вероятностями, что создает тот же средний естественный отбор и тот же генетический дрейф, что и полиномиальное распределение.

Если мутации или рекомбинации не произошло, геном потомства идентичен родительскому. Геном, состоящий из аллелей 00000..., имеет наименьшую возможную приспособленность, равную 1, по определению. Частота мутаций на локус есть $\mu = 3 \cdot 10^{-5}$ [56]. Эпистаз отсутствует (*глава 5*). Вредные мутации игнорируются, что является хорошим приближением, когда популяция далека от стационарного состояния (§ 2.1, рисунок 2.4 внизу).

Значения s_i выбираются в соответствии с заданным случайным распределением, и их значения фиксируются до начала прогонов моделирования. Рассматриваются две формы распределения: экспоненциальная и степенная

$$\frac{1}{s_0} \exp\left(-\frac{s}{s_0}\right), \quad a \frac{s_0^a}{(s + s_0)^{a+1}}$$

где s_0 и a — входные параметры модели. Основное внимание было уделено экспоненциальному распределению, поскольку оно часто наблюдается в эксперименте (§ 5.4). Предполагается, что начальная популяция однородна и несет менее приспособленный аллель на каждом локусе.

Перед случайной выборкой потомства, из популяции случайным образом отбирают rN пар геномов (rN коинфицированных клеток). Каждая пара геномов подвергается рекомбинационным перекрестам в M случайно выбранных местах, создавая два дочерних генома с чередующимися родительскими сегментами. Один из двух рекомбинантов заменяет один родительский.

При экспоненциальной форме распределения s , существует 6 параметров модели: N, L, μ, r, M, s_0 . Три из этих параметров известны из эксперимента: $\mu = 3 \cdot 10^{-5}$ [56], $M \approx 10$ [370], $L \approx 2000$ [15]. Четвертый параметр — эффективное население $N = 10^5 - 10^6$ [5, 154] (§ 2.2). Для распределения по степенному закону, есть дополнительный параметр, a . Наиболее чувствительные параметры r и s_0 были получены подгонкой результатов к экспериментальным данным из геномных образцов из работы [15].

Меры корреляции аллелей

Для оценки этих двух параметров, были использованы две следующие статистики корреляции между аллелями на парах локусов (в популяционной генетике, «неравновесие сцепления», LD). При моделировании и в геномных данных из пациентов, были определены «очень полиморфные» пары локусов, как пары где аллельная доля f , для обоих локусов, находится в промежутке $0.25 < f < 0.75$. Первая статистика LD была определена как $1 - \langle f_{AB}^{LRH} / (f_A f_B) \rangle$, где f_{AB}^{LRH} — частота наименее представленных гаплотипов для локусов А и В, $f_A f_B$ — произведение частот соответствующих аллелей, а $\langle \dots \rangle$ означает усреднение по «очень полиморфным» парам. Второй статистикой LD была доля «очень полиморфных» пар, для которых $f_{AB}^{LRH} < 0.04$.

Таблица 2.3. Наблюдаемые количества, рассчитанные по последовательностям ВИЧ полимеразы^a

Пациент # Последовательность ^b	Размер выборки	# полиморф. оснований	% в моделировании ^c	Наблюдаемо полиморфны	% в моделировании	Очень полиморф. пары $f > 25\%$	% в моделировании	Ср. LD для очень полиморф. пар	95 % довер. интервал парам	Доля очень разн. пар с 4-м тапп. < 4%
# 1 1	20	63	0.28	63	0.28	11	0.05	0.43	0.09	0.09
2	30	72	0.32	28	0.12	14	0.06	0.75	0.05	0.6
3	17	49	0.22	49	0.22	11	0.05	0.8	0.05	0.31
# 2 1	30	46	0.2	9	0.04	4	0.02	0.75	0.33	0.33
2	12	21	0.09	21	0.09	5	0.02	0.7	0.28	0.3
# 3 1	18	39	0.17	39	0.17	8	0.04	0.55	0.14	0.18
2	43	71	0.31	36	0.16	5	0.02	0.53	0.25	0.2
# 4 1	38	76	0.34	44	0.19	13	0.06	0.67	0.09	0.44
2	22	70	0.31	70	0.31	11	0.05	0.37	0.09	0.07
Average \pm SD	26	56 ± 19	0.25 ± 0.08	40 ± 19	0.16 ± 0.09	9 ± 4	0.04 ± 0.02	0.62 ± 0.15		0.28 ± 0.17

^a Геномные данные получены экспериментальной группой работы [15]. Показанные результаты сравниваются с прогнозами моделирования для оценки параметров модели (рисунок 2.14).

^b Для каждого пациента, показаны 2-3 временные точки.

^c Для сравнения с моделированием, где моделируются L локусов, количество различных оснований, наблюдаемых в полимеразе, экстраполируется на весь геном путем деления количества различных оснований на общее количество оснований в последовательностях *pol* (1100 нуклеотидов), умноженных на длину генома ВИЧ (9700 нт) и делением на длину генома ВИЧ, считающуюся переменной (2000 нт).

Данные пациента

У 4 пациентов, не получавших лечения, были отобраны девять наборов последовательностей генов ВИЧ-полимеразы, в 2-3 временных точках для каждого пациента. Полное число последовательностей на образец варьировалось от 12 до 42, в среднем 25. Образцы плазмы были получены от взрослых, хронически инфицированных пациентов, включенных в клинические исследования в Клиническом центре НИИ в Бетесда, Мэриленд, США. Все пациенты дали информированное согласие в письменной форме. Для получения каждой последовательности, секвенировали единственный геном, как описано в работе [371]. В методе отсутствует рекомбинация, связанная с анализом. Длина последовательности варьировалась от 1200 до 1250 нуклеотидов. Программное обеспечение Clustal X (DNASTAR-MEGALIGN или MEGA) использовалось для выравнивания. После сравнения с консенсусом полимеразы (1100 оснований), каждая последовательность была преобразована в бинарную строку. Поскольку приведенные выше показатели LD не зависят от того, какой аллель является более приспособленным, результаты инвариантны по отношению к выбору консенсусной

последовательности. Размер выборки, количество локусов и два показателя LD приведены в таблице 2.3.

2.3.2 Моделирование адаптации к ВИЧ

На рисунке 2.13 показаны наблюдаемые величины эволюции для набора значений $r = 0.01$ и $s_0 = 0.005$, оцененных в следующем подразделе из экспериментальных данных. Как предсказывают бесполое и половое модели [7, 8, 12, 14, 72, 73, 76], распределение приспособленности (рисунок 2.13А) представляет собой уединенную волну с фиксированным профилем, движущуюся к более высокой приспособленности. Средняя скорость адаптации (производная по времени логарифмической приспособленности) составляет 0.0023/день, что в 2 раза больше бесполого значения при $r = 0$. Таким образом, только 1% рекомбинации ускоряет адаптацию в два раза.

Частоты выгодных аллелей для нескольких оснований $f_i(t)$ были смоделированы, как функции времени t (рисунок 2.13В). Детерминистическая динамика, описываемая моделью с одним локусом (§ 2.1), показана пунктирными линиями. Все траектории имеют сильную стохастическую составляющую, с двумя типами поведения. На некоторых локусах, $f_i(t)$ колеблется вокруг монотонной траектории, предсказанной моделью с одним локусом. На других локусах, полезные аллели падают до низкого уровня, через 50-200 поколений, где они поддерживаются только новыми мутационными событиями.

Четыре наблюдаемые меры разнообразия на рисунке 2.13С следующие:

- (i) Средняя доля выгодного аллеля $\langle f \rangle$
- (ii) Доля оснований i с долей миноритарного аллеля $\min[f_i(t), 1 - f_i(t)] > 0.04$, именуемые «наблюдаемо полиморфными»
- (iii) Доля оснований с долей миноритарного аллеля больше 0.25, «очень полиморфные»
- (iv) Гетерозиготность $T = 2\langle f(1 - f) \rangle$, усредненная по «наблюдаемо полиморфным» основаниям.

Примерно через 500 дней, эти наблюдаемые достигают стационарного состояния, с некоторыми колебаниями. Средняя высота их плато не далека от предсказания одно-локусной модели, с широко варьируемым значением s [6] (§ 2.1). Это справедливо как для экспоненциального, так и для степенного распределения s (при $a > 1$). При изменении скорости рекомбинации на два порядка, $0.01 < r < 1$, средняя гетерозиготность T остается между 0.25 и 0.33, что близко к наблюдаемому у пациентов значению (таблица 2.1). Доля наблюдаемо полиморфных локусов в геноме весьма чувствительна как к r , так и к s_0 и может использоваться для оценки s_0 из данных о последовательности (см. ниже).

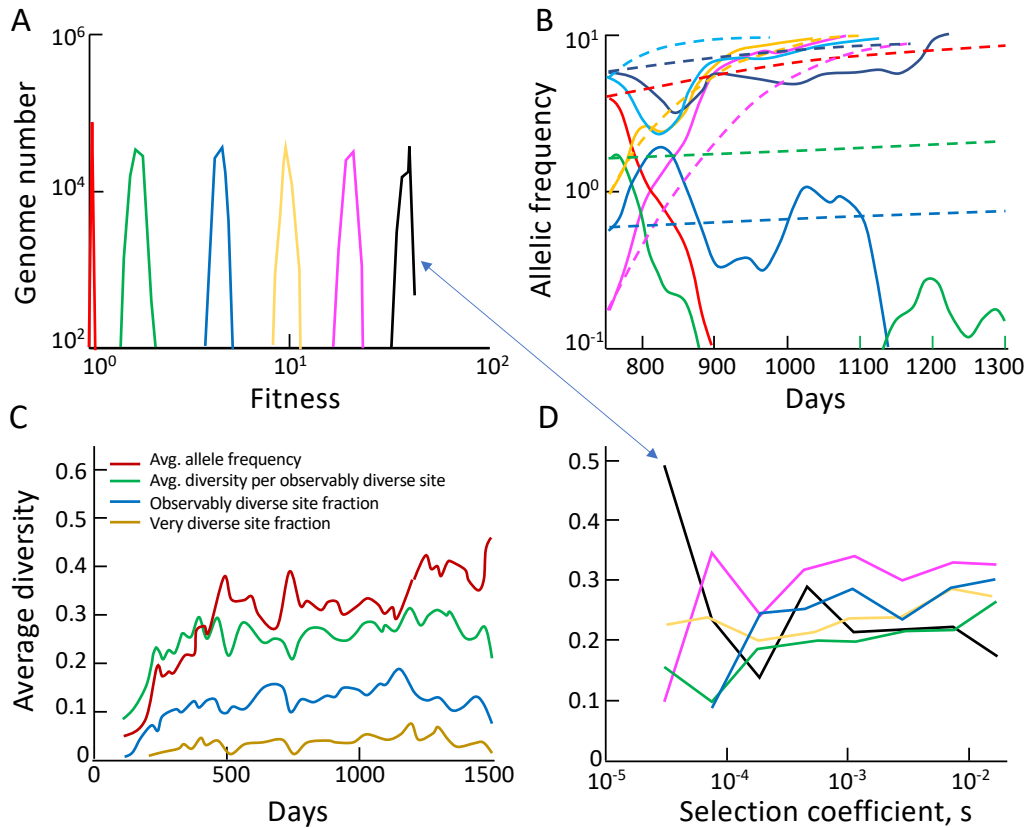


Рисунок 2.13. Моделирование адаптации вируса при хронической ВИЧ-инфекции. (А) Распределение популяции по приспособленности: бегущая волна. Распределение показано с интервалом в 300 дней: 1-й день (красный), 300-й день (зеленый), 600-й день (синий), 900-й день (желтый), 1200-й день (пурпурный), 1500-й день (черный). (В) Динамика выгодных аллелей на высоких частотах. Траектории, показанные для 8 случайно выбранных оснований с частотой выгодных аллелей на локус, $f > 0.1$ на 750-й день: моделирование (сплошные), предсказание модели с 1 локусом (пунктир). Значения s на показанных участках: 0.02 (голубой), 0.004 (черный), снова 0.004 (красный), $4.5 \cdot 10^{-4}$ (зеленый), 0.01 (желтый), 0.01 (синий), 0.01 (пурпурный). (С) Величины, характеризующие среднее разнообразие населения. Доля «наблюдаемо полиморфных» локусов, определяемых как наличие более 0.04 аллеля меньшинства (синие), доля «очень полиморфных» локусов, определяемых как наличие более 0.25 аллеля меньшинства (коричневые), средняя частота выгодных аллелей на локус $\langle f \rangle$ (темно-красный), гетерозиготность $\langle 2f(1-f) \rangle$, усредненное только по «наблюдаемо полиморфным» основаниям (зеленый). (D) Средняя гетерозиготность наблюдаемо полиморфных оснований, классифицированных в соответствии с их коэффициентом отбора s для 10 равноудаленных моментов времени с интервалом в 300 дней, те же времена и цвета, что и в (А). Параметры в (А-D) оцениваются из последовательностей гена полимеразы ВИЧ в типичных пациентах (рисунок 2.14): $r = 0.01$, $s_0 = 0.005$, $M = 10$, $L = 2000$, $N = 10^5$, $\mu = 3 \cdot 10^{-5}$. По работе [15].

Среднее разнообразие «наблюдаемо полиморфных» оснований, сгруппированных по их коэффициенту отбора s , показано на рисунке 2.13D для пяти моментов времени. Зависимость от s слабая, показывая, что основания с большими s и малыми s вносят примерно одинаковый вклад в чистое разнообразие в любое время. Эта особенность не может быть объяснена моделью независимых локусов (§ 2.1) и свидетельствует о том, что аллели с малым s сцеплены с аллелями с большим s , и естественный отбор происходит на уровне целого генома а не локуса. Эта ситуация соответствует началу адаптации, начиная с самой неприспособленной последовательности. По мере приближения популяции к наиболее приспособленной последовательности, динамика на локусах с разными s начинает сильно отличаться (§ 5.4).

2.3.3 Оценка скорости рекомбинации и среднего коэффициента отбора

Наблюдаемые, показанные на рисунке 2.13, описывают эволюционный процесс для общей информации. Однако они не подходят для основной цели: оценки эффективной вероятности рекомбинации (коинфекции) r . Обычно r оценивается по выборкам в разные моменты времени, на основе обнаружения генетических кроссоверов [77]. Преимущество этого метода заключается в необходимости только одновременного образца. Он основан на двух мерах корреляции (неравновесия по сцеплению, LD), чувствительных к вероятности рекомбинации r , а именно:

- i) Доля наименее представленных гаплотипов (§ 2.2), усредненная по парам «очень полиморфных» оснований и нормализованная к ее предсказанному нулевому значению LD
- ii) Доля «очень полиморфных» пар оснований, в которых отсутствует хотя бы один гаплотип в выборке из четырех. Девять наборов последовательностей гена полимеразы, который занимает 11% генома ВИЧ, были выделены из плазмы 4 нелеченых пациентов с 2-3 временными точками на пациента (таблица 2.3, § 2.3.1).

Значения LD, рассчитанные на основе данных пациентов, были подогнаны к двум мерам LD, оцененным на основе 11% длины смоделированного генома, используя r , как подгоночный параметр. Используя подогнанное значение s_0 (ниже), две смоделированные меры LD, построенные как функция скорости рекомбинации r , были сопоставлены с диапазоном LD в данных пациентов (таблица 2.3, рисунки 2.14А и В). Быстрое уменьшение обоих показателей LD с r демонстрирует их чувствительность к этому параметру. Интервал r , который соответствует данным, составляет от 0.05 до 0.15 с центром в $r = 0.01$.

Оценка чувствительна к количеству кроссинговеров на геном M . Приведенные выше оценки предполагали $M = 10$, на основе независимых данных [370]. Уменьшение M в 3 раза увеличивает предсказываемую LD, что увеличивает конечную оценку r в 2–3 раза (рисунок 2А, Б).

Значение s_0 , используемое на рисунке 2.14А-В, было оценено путем сопоставления чистого изменения приспособленности вируса, предсказанного с помощью моделирования, как функции s_0 , с экспериментальным значением (рисунок 2.14D). Экспериментальный прирост приспособленности был оценен на основе дополнительной информации следующим образом. Большой класс моделей динамики ВИЧ, включающий популяции восприимчивых, инфицированных и иммунных клеток, предсказывает устойчивое состояние [19, 223, 224, 372-375]. Во всех этих моделях, количество инфицированных клеток фиксируется вблизи порога активации иммунных клеток, а количество восприимчивых клеток обратно пропорционально средней приспособленности вируса. Это предсказание является общим для моделей, в которых

иммунный ответ контролирует ВИЧ. Последний факт экспериментально доказан во многих исследованиях [211, 212].

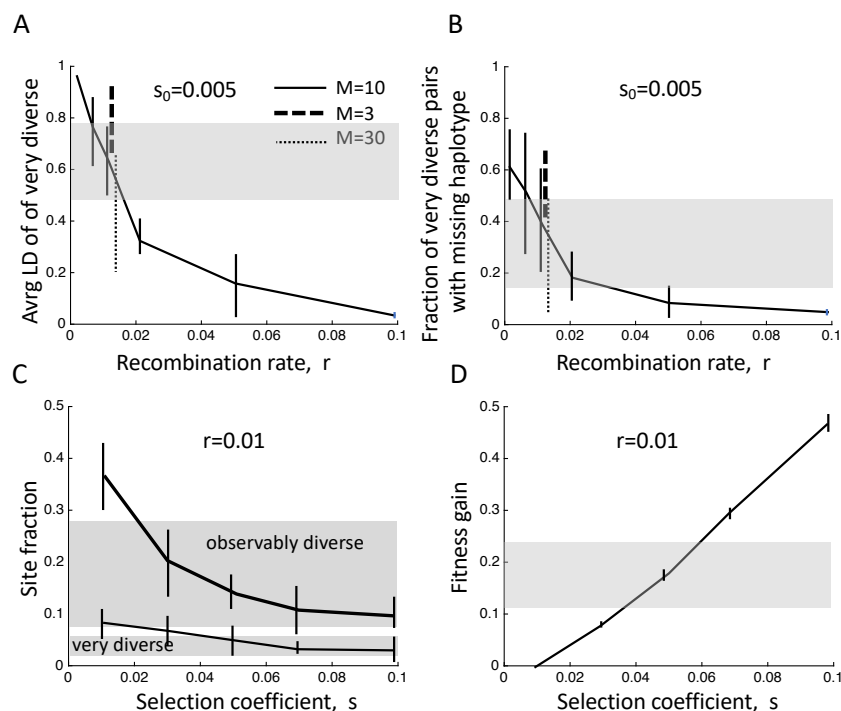


Рисунок 2.14. Оценка параметров модели хронической ВИЧ-инфекции. Скорость рекомбинации r и средний коэффициент отбора s_0 оцениваются путем сравнения четырех наблюдаемых величин в данных с их предсказанными значениями при моделировании. Серые области показывают среднее \pm одно стандартное отклонение величин, рассчитанных по данным гена *pol* у 4 пациентов (таблица 2.3). При моделировании каждое количество рассчитывается на 1500-й день для 11% генома. Показано среднее значение по 16 случайным прогонам, а планки погрешностей показывают стандартное отклонение от прогонов. Результаты показаны для одного размера популяции, $N = 10^5$, и 3 значений числа кроссоверов, $M = 3$ (пунктирная линия), 10 (сплошная линия) и 30 (пунктирная линия). Результаты для разных M немного сдвинуты по горизонтали для ясности. (А) Неравновесие по сцеплению (LD) «очень разных» пар для $s_0 = 0.005$. LD определяется как $1 - \langle f_{AB}^{LRH} / f_A f_B \rangle$, где f_{AB}^{LRH} — частота наименее представленного гаплотипа для пары оснований А и В, $f_A f_B$ — произведение частот одного локуса или значение f_{AB}^{LRH} находится в отсутствии сцепления для $s_0 = 0.005$. (В) Другой показатель LD: доля очень полиморфных пар с частотой наименее представленного гаплотипа ниже 0.04 для $s_0 = 0.005$. (С) Доля оснований, которые «наблюдаемо полиморфны» (жирная линия) и «очень полиморфны» (тонкая линия) для $r = 0.01$. (D) Средняя приспособленность населения для $r = 0.01$. Остальные фиксированные параметры показаны в легенде к рисунку 2.13. По работе [15].

Значение s_0 , используемое на рисунке 2.14А-В, было оценено путем сопоставления чистого изменения приспособленности вируса, предсказанного с помощью моделирования, как функции s_0 , с экспериментальным значением (рисунок 2.14D). Экспериментальный прирост приспособленности был оценен на основе дополнительной информации следующим образом. Большой класс моделей динамики ВИЧ, включающий популяции восприимчивых, инфицированных и иммунных клеток, предсказывает устойчивое состояние [19, 223, 224, 372-375]. Во всех этих моделях, количество инфицированных клеток фиксируется вблизи порога активации иммунных клеток, а количество восприимчивых клеток обратно пропорционально средней приспособленности вируса. Это предсказание является общим для моделей, в которых

иммунный ответ контролирует ВИЧ. Последний факт экспериментально доказан во многих исследованиях [211, 212].

Следовательно, если приспособленность вируса медленно увеличивается в процессе адаптации, стационарное количество CD4 медленно снижается, что может быть связано с прогрессированием заболевания до СПИДа [202] (§ 6.1). Перед полностью развившимся СПИДом устойчивое количество Т-клеток CD4 падает с 500 до 5 – 50 клеток на 1 мл крови. Это уменьшение соответствует увеличению приспособленности между 10 и 100, что определяет среднюю оценку $s_0 = 0.005$ для обоих рассмотренных распределений s , экспоненциального и степенного ($a = 2$).

Другой метод оценки s_0 заключается в сопоставлении смоделированных фракций «наблюдаемо полиморфных» и «очень полиморфных» участков с данными пациента для гена полимеразы (рисунок 2.14С), где s_0 больше 0.003. Обоиими методами, получается $s_0 = 0.003-0.005$.

Основное предположение приведенного выше моделирования состоит в том, что начальная популяция равномерно менее приспособлена. Добавление малой доли полезных аллелей в начале прогона чтобы проверить это предположение слабо влияет [15] на конечные оценки r и s_0 . При малых скоростях рекомбинации $\sim 1\%$, большинство начальных аллелей быстро вымирают из-за случайного дрейфа. При более высоких $r > 5\%$, начальные аллели очень важны и доминируют в эволюции [8-10, 14]. Представленные здесь результаты сосредоточены на экспоненциальном распределении s , поскольку оно часто наблюдается в эксперименте [91-95] (§ 5.4). Однако, похожие оценки получаются и для степенного распределения $as_0^a/(s + s_0)^{a+1}$, если $a \geq 2$. Для $a < 1$, данные вообще не могут быть подогнаны, потому что очень редкие оснований полиморфны; крайне маловероятно, что этот случай является репрезентативным для реального распределения. Заметим, что фактор эпистаза не включен во все эти модели. На самом деле, эпистаз у вирусов существует [250, 376-378], включая ВИЧ [249, 379], и может повлиять на наблюдаемые значения LD (*глава 5*).

2.3.4 Заключение

Таким образом, вероятность рекомбинации у среднего нелеченного ВИЧ-позитивного человека составляет $r \approx 1\%$. Это значение слишком низкое, чтобы компенсировать эффекты генетического сцепления. В частности, средняя скорость адаптации меньше, чем предсказывает

одно локусная модель, в 4 раза [15] (рисунок 2.3А и В). Тем не менее, даже при столь редкой рекомбинации, адаптация ускорена в 2 раза, а мера генетического сцепления в 1.5 раза меньше, чем при бесполом размножении, $r = 0$ (рисунок 2.14А и В). Похожие оценки для этих двух параметров, $r \approx 0.02$ и $s_0 \approx 0.005$, были получены другим методом [77], используя две упрощенных модели. В целом, эти результаты подчеркивают важное значение полового размножения для адаптации, даже если доля особей участвующих в рекомбинации, очень мала.

Глава 3. Многолокусные модели эволюции популяций бесполой организмов и вирусов

Реальные популяции имеют большое количество полиморфных локусов. Два человека, случайно выбранные с разных концов планеты, отличаются друг от друга на миллионы оснований ДНК. Многие РНК-вирусы с их коротенькими геномами, такие как ВИЧ и вирус гриппа, имеют сотни полиморфных локусов. Многолокусный контекст требует совсем другой, более общей аналитической теории.

Ранние работы в области популяционной генетики содержали теоретические аргументы [59, 69, 70, 78, 100] и экспериментальные доказательства [66] того, что в отсутствие рекомбинации родительских геномов, присущей половому размножению, или на близких локусах при ее наличии, процессы эволюции на разных локусах сильно взаимодействуют, так что модель с одним локусом (§ 2.1), вообще говоря, не применима. Количественные результаты, полученные в рамках одного локуса, хотя и полезны в некоторых частных случаях, не могут быть применены к эволюции с несколькими локусами из-за генетического сцепления (совместного наследования): генетическая информация на локусах, не разделенных рекомбинацией, передается от родителя к потомству одним пакетом, а не независимо.

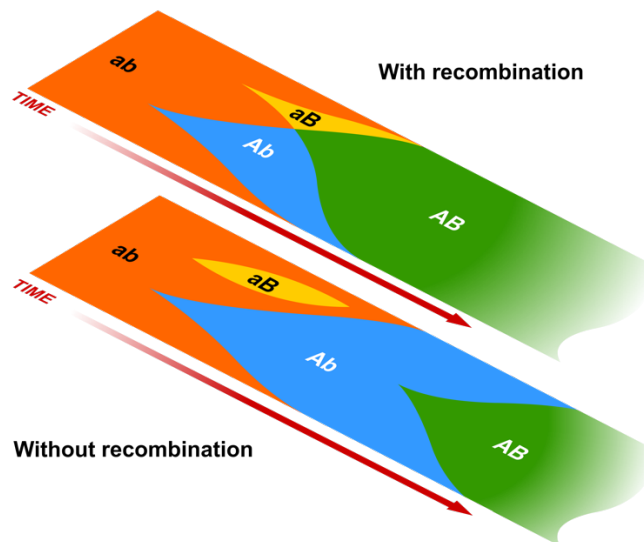


Рисунок 3.1. Клональной интерференции противостоит рекомбинация. Внизу: полезные мутации в двух разных локусах исходного варианта ab приводят к созданию более приспособленных гаплотипов, Ab и aB . Эти новые линии препятствуют росту друг друга. Более приспособленный клон захватывает население. В конце концов, вторая мутация генерирует наиболее приспособленный гаплотип AB . Вверху: при частой рекомбинации геномов (половое размножение) два клона быстро объединяются, образуя наиболее приспособленный гаплотип AB . По работе [91].

«Клональная интерференция», также известная как эффект Фишера-Мюллера [59, 70], является одним из наиболее важных последствий генетического сцепления. По сути, этот эффект можно рассматривать, как конкурентное вытеснение. Особи с разными полезными аллелями

конкурируют друг с другом за место в популяции (рисунок 3.1). Победителем становится генетическая линия с наибольшей приспособленностью, которая вытесняет из популяции все другие линии (выживание сильнейшего).

Этот эффект клональной интерференции исчезает, только если популяция астрономически велика (экспоненциально велика по количеству локусов) [7]. В этом пределе, частые мутации могут почти мгновенно создать любую последовательность. Вместо двух конкурирующих аллелей, мгновенно возникает система глубоко вложенных клонов со всевозможными последовательностями. Следовательно, эффект клональной интерференции возникает из-за комбинации трех факторов: естественного отбора, отсутствия рекомбинации или близости локусов и ограниченного размера популяции.

Задачей большого практического значения является предсказание средней скорости эволюции популяции через заданные постоянные параметры системы, включая количество локусов L , размер популяции N , коэффициенты отбора, вероятность мутаций на локус μ и среднюю частоту менее приспособленных аллелей, обозначенная α . Двумя стандартными мерами скорости эволюции являются эффективная частота замен равная, по определению, разнице между частотой накопления выгодных аллелей и частотой накопления невыгодных аллелей, и скорость адаптации (скорость изменения средней приспособленности), которые могут быть как положительными, так и отрицательными. В настоящей главе обсуждаются две работы, где скорость эволюции выражена через эти параметры системы для двух случаев: однородного коэффициента отбора [7, 11] (§ 3.1) и распределенного коэффициента отбора, который изменяется между локусами [16] (§ 3.2).

3.1 Многолокусная эволюция с фиксированным коэффициентом отбора

Главная математическая трудность заключается в том, что количество возможных последовательностей экспоненциально увеличивается с номером локуса, 2^L для двоичных последовательностей и 4^L , если все 4 основания возможны. Строгое описание этой системы требует решения 2^L сцепленных уравнений Колмогорова [33, 64], что невозможно практически. Для преодоления этой трудности, было предложено классифицировать геномы по их приспособленности и применить детерминистическую динамику к этим классам приспособленности [380, 381]. Эти авторы показали существование распределения, перемещающегося вдоль координаты приспособленности со скоростью, определяемой обрезанием распределения на краю на один геном.

Более точная и биологически правильная реализация этой идеи, описанная ниже в этом разделе, позволила учесть дискретность мутаций, включить стохастические эффекты на краю распределения и, в итоге, впервые получить асимптотически правильное выражение для скорости многолокусной эволюции [7, 11]. Этот метод автоматически учитывает произвольное количество конкурирующих и вложенных клонов, на любом уровне вложенности, а также учитывает другие эффекты сцепления.

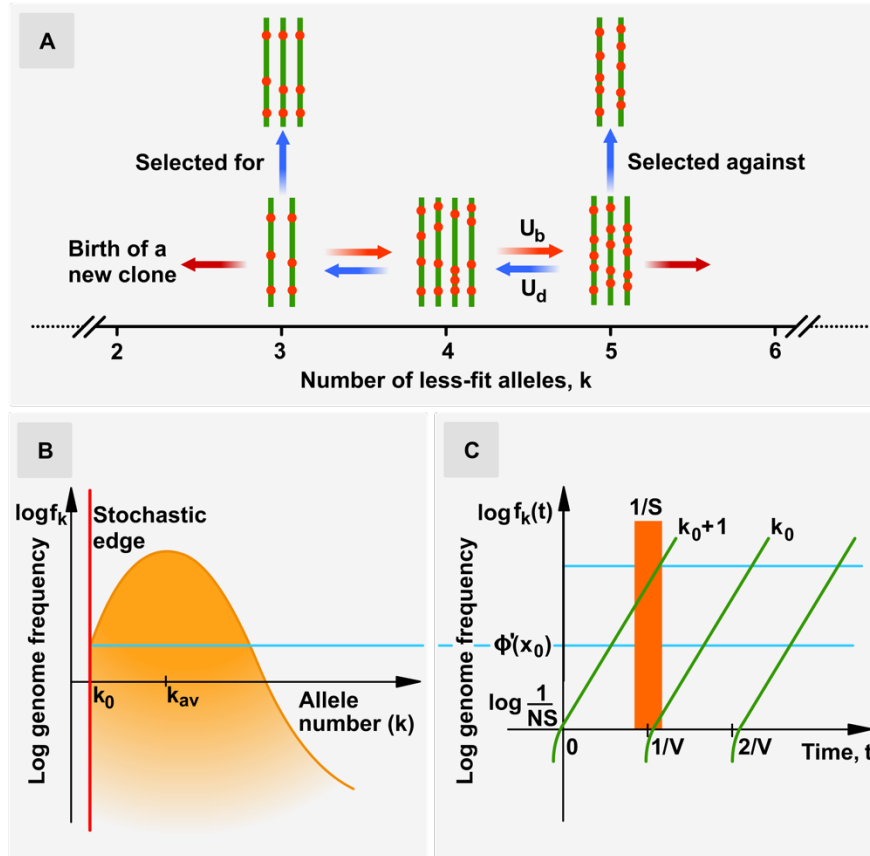


Рисунок 3.2. Адаптация большого числа сцепленных локусов в бесполой популяции представляет собой бегущую волну приспособленности, скорость которой определяется стохастической динамикой наиболее приспособленных индивидумов. (А) Последовательности (зеленые линии) классифицируются по дискретным классам приспособленности в соответствии с количеством невыгодных аллелей k (красные кружки). Эволюционными факторами, влияющими на изменение размера каждого класса во времени, являются полезные и вредные мутации с соответствующими показателями U_b и U_d на геном, случайный генетический дрейф и естественный отбор. (В) Полудетерминистический подход [7, 11]. Динамика доли класса f_k определяется из детерминистического уравнения (3.1). Решение представляет собой уединенную волну с отсечкой при малых числах вредных аллелей $k = k_0$ и бегущим максимумом $k_{av}(t)$. (С) Для определения скорости волны $V = -dk_{av}/dt$, рассматривается стохастическая динамика доли наиболее приспособленных в популяции $f_{k_0}(t)$; показан случай $V > 0$. Новый класс увеличивается детерминистически, если его доля превышает стохастический порог $\sim 1/NS$. По работе [7].

Чтобы учесть огромное количество возможных последовательностей, нужно, во-первых, отклассифицировать их на большие группы с одинаковой приспособленностью. Бинарные последовательности помещаются в дискретные ячейки с фиксированным числом невыгодных аллелей k (рисунок 3.2А). Конкретное распределение этих аллелей среди L локусов не имеет значения. Значение k определяет логарифм приспособленности последовательности $-sk$ по

отношению к наиболее приспособленной последовательности с $k = 0$. Предполагается, что каждое мутационное событие в основании имеет фиксированный эффект приспособленности s или $-s$, в зависимости от аллеля, существующего на локусе. При смене поколений, более приспособленный аллель может мутировать с эффектом $-s$ и с вероятностью $\mu \ll 1$, а менее приспособленный аллель может мутировать в более приспособленный аллель, с эффектом s и с той же вероятностью.

Основное приближение, подтвержденное в конце, состоит в том, что, при достаточно больших размерах популяции (насколько больших, станет известно в конце), почти любой непустой фитнес-класс достаточно велик по размеру, чтобы его можно было рассматривать детерминистически, т. е., без учета случайного генетического дрейфа и с заменой мутаций их средним эффектом (рисунок 3.2b). Единственным исключением из этого приближения будет наиболее приспособленный класс с наименьшим $k = k_0$, который мал и должен рассматриваться стохастически, как описано ниже. Предполагается, что уже следующий класс, $k = k_0 + 1$, достаточно велик, чтобы его динамика была почти детерминистической [7, 11].

Динамика классов приспособленности, рассчитанная численно в рамках такого «полудетерминистическим» подхода, была проверена с помощью прямого компьютерного моделирования Монте Карло описанного в § 2.3, в отсутствие рекомбинации ($r = 0$) (рисунок 3.3). Результат обоих численных методов — это волна, бегущая к более высокой или более низкой приспособленности (большее или меньшее k), в зависимости от начального состояния и параметров модели: N , s , L и частоты мутаций на геном μL . Разница между методами в том, что волна в методе Монте-Карло волна имеет флуктуации, а при полудетерминистическом подходе она гладкая (рисунок 3.3A и C). Оба метода предсказывают почти одинаковую траекторию волны, что подтверждает точность «полудетерминистического» подхода (рисунок 3.3B и D). При высокой начальной приспособленности, происходит накопление невыгодных аллелей, а при низкой начальной приспособленности, их число уменьшается. Рано или поздно, система приходит в равновесии (рисунок 3.3C, D), с гораздо более высоким числом невыгодных аллелей k чем предсказание однолокусной модели, $\mu L/s$ (§ 2.1).

В настоящем разделе, эти результаты выводятся в общей аналитической форме [7, 11]. Полудетерминистический подход разбивает сложную задачу на несколько простых шагов (§§ 3.3.1–3.3.5):

(i) Написать детерминированное уравнение баланса, которое контролирует динамику всех классов приспособленности, за исключением наилучшего класса.

(ii) Решить это уравнение и получить решение бегущей волны приспособленности, но с неопределенной скоростью (средняя частота замен).

- (iii) Продемонстрировать, что волна имеет край со стороны высокой приспособленности.
- (iv) Выразить разницу в высоте волны меж краем и пиком через скорость волны.
- (v) Найдите максимум волны из условия нормализации.
- (vi) Определить точку отсечки волны на краю высокой приспособленности.

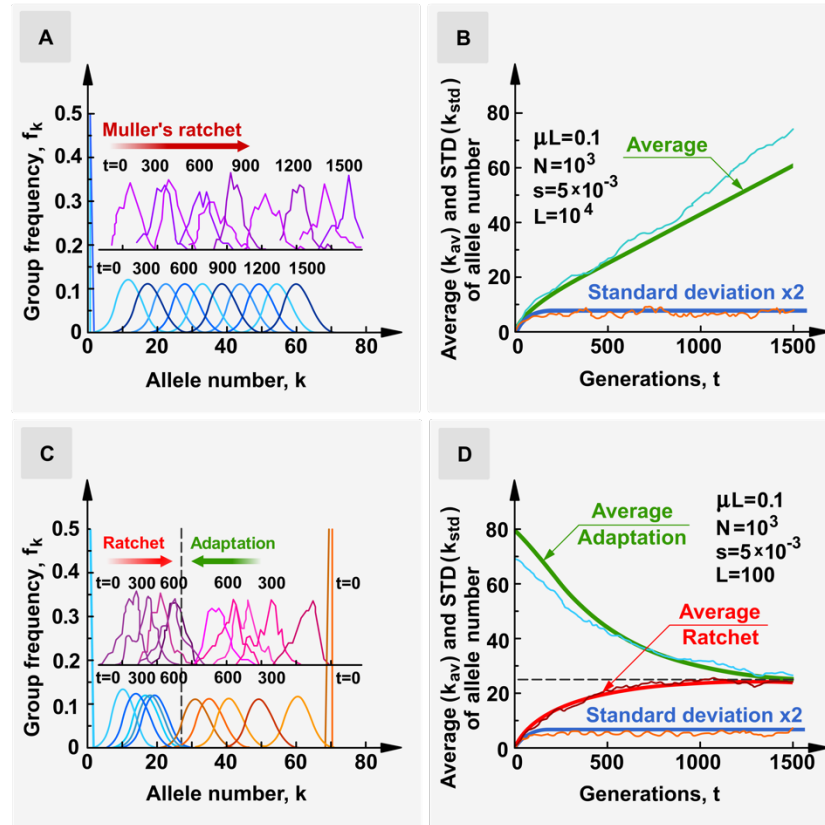


Рисунок 3.3. Два численно полученных примера эволюции популяции в длинном (A и B) и относительно коротком (C и D) геноме. Значения параметров показаны на B и D. (A и C) Частота последовательностей с k неприспособленными аллелями в разные моменты времени, показанных на кривых. Широкие стрелки показывают направление эволюции. Рваные кривые, полученные с помощью псевдослучайного моделирования (§ 2.3), соответствуют либо «храповику Мюллера» (красная стрелка), начальное значение $k(0) = 1$, либо адаптации (зеленая стрелка), начальному $k(0) = 70$. Гладкие кривые ниже (синий для храпового механизма, коричневый цвет для адаптации) получаются с помощью другого типа моделирования с использованием детерминистической аппроксимации для основной массы и стохастической обработки для наиболее приспособленного класса. (B и D) Соответствующая временная зависимость для среднего и стандартного отклонения k (ширина волны) для двух методов моделирования. Пунктирные линии показывают стационарное значение k_{av} . По работе [7].

- (vii) Оценить среднюю частоту наиболее приспособленного класса из отдельного стохастического рассмотрения.
- (viii) Сопоставить этот результат с высотой края (vi) и отсюда найти скорость волны.

3.1.1 Детерминистическое уравнение для фитнес-классов: бегущая волна

Отправная точка, это уравнение детерминистической динамики (2.5.2), полученное в главе 2 для одного локуса и двух аллелей. Вместо двух аллелей, имеется много классов приспособленности, состоящих из различных последовательностей с одинаковой приспособленностью, и похожее уравнение применяется к каждому классу. Доля класса с k вредных аллелями в популяции $f_k(t)$ описывается конечно-разностным уравнением вида

$$f_k(t + 1) - f_k(t) = U_d f_{k-1}(t) + U_b f_{k+1}(t) - \{U_d + U_b + s(k - k_{av})\} f_k(t) \quad (3.1)$$

где k изменяется от 0 до L и граничное условие $f_{-1}(t) \equiv f_{L+1}(t) \equiv 0$, введены частоты вредных и полезных мутаций на геном за поколение

$$U_d \equiv \mu(L - k_{av}), \quad U_b \equiv \mu k_{av}$$

соответственно, и

$$k_{av}(t) = \sum_k k f_k(t)$$

есть среднее число аллелей, что делает уравнение (3.1) нелинейным. По определению, $\sum_k f_k(t) \equiv 1$ в любой момент времени t . Здесь мы пренебрегаем множественными мутациями на геном и предполагаем $s \ll 1$. Интересно, что не используя эти два приближения, получается то же уравнение (3.1) [7, 11].

Поскольку решать дискретное уравнение (3.1) неудобно и не нужно (понадобилось бы преобразование Лапласа), оно успешно приближается дифференциальным уравнением в частных производных. Поскольку как частота мутаций так и коэффициент отбора предполагаются маленькими, $f_k(t)$ эволюционирует медленно, так что $f_k(t + 1) - f_k(t) \approx \frac{\partial f_k}{\partial t}$. Непрерывное приближение для $f_k(t)$ как функции k менее тривиально. На самом деле зависимость $f_k(t)$ от k может быть достаточно резкой на важном краю волны, вдали от пика. Однако его логарифм меняется с k медленно (см. ниже), так что можно воспользоваться приближением

$$\frac{f_{k+1}}{f_k} \approx \exp \left[\frac{\partial \log f(k, t)}{\partial k} \right]$$

После введения перенормированных величин

$$\sigma = \frac{s}{U_b + U_d}$$

$$\tau = (U_b + U_d)t$$

уравнение (3.1) можно привести к непрерывному как по k , так и по τ виду:

$$\frac{\partial \log f_k(t)}{\partial \tau} = (1 - \alpha) e^{-\frac{\partial \log f_k(t)}{\partial k}} + \alpha e^{\frac{\partial \log f_k(t)}{\partial k}} - \sigma(k - k_{av}) - 1 \quad (3.2)$$

где

$$\alpha \equiv \frac{U_b}{U_b + U_d} = \frac{k}{L}$$

доля вредных аллелей в геноме (в главе 2 обозначенная f). Нелинейное дифференциальное уравнение в частных производных (3.2) трансляционно инвариантно и имеет решение бегущей волны

$$\log f(k, t) \equiv \phi(x), x = k - k_{av}(t) \quad (3.3)$$

После подстановки (3.3) в (3.2), получается обыкновенное дифференциальное уравнение в система отсчета, связанной с движущейся волной [7]

$$\sigma x = (1 - \alpha)e^{-\phi'(x)} + \alpha e^{\phi'(x)} + v\phi'(x) - 1 \quad (3.4)$$

Здесь $\phi'(x) \equiv \frac{d\phi(x)}{dx}$, а новое обозначение v — перенормированная скорость движения волны в сторону ухудшения приспособленности

$$v \equiv \frac{(U_b + U_d)^{-1} dk_{av}(t)}{dt},$$

Это значение может быть как отрицательным (случай адаптации), так и положительным (накопление вредных мутаций), и соответствует в эксперименте суммарной частоте накопления геномных мутаций (вредные минус полезные).

Для заданного значения v , уравнение (3.4) и граничное условие $\phi(0)$ полностью определяют форму распределения приспособленности $\exp[\phi(x)]$. Хотя решить уравнение (3.4) в явном виде трудно, все важные размеры и скорость волны v получены в самом общем виде. Оказывается, форма волны приблизительно гауссова вблизи центра, но отклоняется от этой зависимости вдали от центра и имеет отсечку при высокой приспособленности ($x_0 < 0$). Длина хвоста волны с высокой приспособленностью («лидер») выражается через скорость волны v и параметры модели.

3.1.2 Центр волны и край высокой приспособленности

Следующий шаг — вывести максимум распределения $\phi(0)$. Основное приближение состоит в том, что $\phi(x)$, логарифм плотности распределения относительной x , является медленной функцией x , так как характерная ширина $\phi(x)$ по x много больше единицы. Следовательно, можно разложить экспоненты в правой части (3.4) линейно по $\phi'(x)$. После интегрирования по x и с учетом условия нормировки $\int dx \exp[\phi(x)] = 1$ получается

$$\phi(x) = -\frac{\sigma x^2}{2(1-2\alpha-v)} + \log \sqrt{\frac{\sigma}{2\pi(1-2\alpha-v)}} \quad (3.5)$$

Из (3.5) следует, что распределение геномов по приспособленности $\exp[\phi(x)]$ вблизи $k = k_{av}$ может быть аппроксимировано гауссианой с дисперсией

$$Var[k] = (1-2\alpha-v)/\sigma \gg 1 \quad (3.6)$$

Из-за условия, что дисперсия (3.6) положительна, скорость волны v должна быть меньше $1-2\alpha$. Другими словами, если вредные мутации в целом накапливаются, что соответствует случаю $v > 0$, то скорость их накопления не может превышать $1-2\alpha$. Формула (3.6), известная в популяционной генетике как «фундаментальная теорема Фишера», является куда более общей, чем данная модель. Она связывает ширину распределения геномов по приспособленности с частотой замен: более широкая волна движется быстрее к более высокой приспособленности (или медленнее к более низкой).

Край высокой приспособленности.

Далее рассматривается хвост волны с высокой приспособленностью. Легко показать, что волна там имеет край (рисунок 3.3). Чтобы найти его положение по x , уравнение (3.4), вместо дифференциального уравнения для $\phi'(x)$, рассматривается, как алгебраическое уравнение для функции $x(\phi')$. Заметим, что значение функции $x(\phi')$ в (3.4) положительно-бесконечно в пределах $\phi' \rightarrow \pm\infty$. В середине, функция $x(\phi')$ имеет один минимум

$$e^{\phi'(x_0)} = \frac{1}{2\alpha} \left[-v + \sqrt{v^2 + 4\alpha(1-\alpha)} \right] \equiv u \quad (3.7)$$

$$x_0 = -\frac{1}{\sigma} (1 - 2\alpha u - v \ln u - v) < 0 \quad (3.8)$$

Координата x_0 («лидер») и есть край волны. За краем нет геномов: все высокоприспособленные классы $x < x_0$ пусты.

Предполагается, что лидер длинный, $|x_0| \gg 1$, что подразумевает $s \ll U_b + U_d$. Например, для полиовируса и ВИЧ, $U_b + U_d \approx 0.3$ [22, 382], так что данный вывод имеет смысл для мутаций с коэффициентами отбора $s \sim 0.1$ и меньше. Редкие выгодные мутации с большими значениями s фиксируются в популяции быстро и по одиночке. Вредные же мутации с большими значениями s уходят из популяции в течение одного поколения и никакой роли не играют.

После нахождения края и лидера x_0 , нужно найти разность $\phi(0) - \phi(x_0)$. Можно тривиально написать

$$\phi(0) - \phi(x_0) = \int_{x_0}^0 \phi'(x) dx \quad (3.9)$$

Делая замену переменной интегрирования $x = x(\phi')$ и используя (3.4) и (3.7), получается перепад высоты волны между пиком и краем

$$\phi(0) - \phi(x_0) = \frac{1}{\sigma} \left\{ 1 - 2\alpha - \frac{\nu}{2} [\ln^2(eu) + 1] - 2\alpha u \log u \right\} \quad (3.10)$$

где e — постоянная Эйлера, $\log e = 1$, а $\phi(0)$ определяется (3.5). Величина $\phi(x_0)$ представляет собой средний логарифм частоты наиболее приспособленного класса на краю. Поскольку в его динамике преобладают случайные мутации и стохастический дрейф, оценить само $\phi(x_0)$ из детерминистического рассмотрения нельзя, и нужно отдельное рассмотрение.

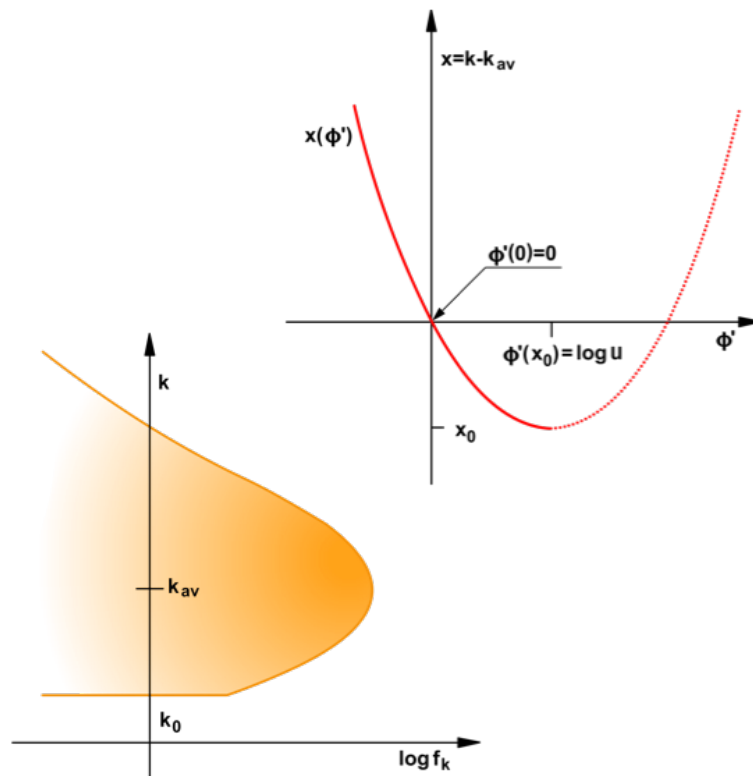


Рисунок 3.4. Минимум функции $x(\phi')$ определяет положение края волны. По работе [11].

3.1.3 Вероятностный подход к наиболее приспособленным

Детерминистический подход в § 3.1.2 дает непрерывный набор решений в виде уединенной волны с неопределенной скоростью. Чтобы выбрать правильное решение, нужно рассмотреть динамику наиболее приспособленного класса, $k = k_0$. Лучший класс либо создается выгодными мутациями в режиме адаптации, либо вымирает в режиме храповика Мюллера. Среднее время требуемое для приобретения нового лучшего класса или его потери, и является ограничивающим фактором скорости прогресса (или регресса) волны. Случайный и дискретный характер мутационных событий, а также случайный генетический дрейф, критически важны для динамики

лучшего класса.

Стохастическую динамику лучшего класса можно рассматривать в рамках классической однолокусной двухаллельной модели, описываемой уравнением Колмогорова; его различные частные случаи анализируются в обзоре [33], который включен в диссертацию. Лучший класс рассматривается, как маолчисленная выгодная аллель с эффективным коэффициентом отбора S , в то время как остальная часть населения объединяется в невыгодный аллель. Хорошо известно, что естественный отбор доминирует над случайным генетическим дрейфом, когда число аллелей $Nf_{k_0}(t)$ выше стохастического порога $1/|S|$, в каком случае полезная аллель фиксируется в популяции с вероятностью, близкой к 1. Ниже этого порога, правит случайный генетический дрейф, и полезная аллель, скорее всего, вымрет. Если $f_{k_0}(t)$ гораздо выше порога $1/|S|$, детерминистическое уравнение (3.1.1) применимо, откуда

$$\begin{aligned}\frac{\partial f_{k_0}}{\partial t} &= M(t) + Sf_{(k_0)(t)}, \\ M(t) &\equiv U\alpha f_{k_0+1}(t), \\ S &\equiv U(\sigma|x_0| - 1)\end{aligned}\tag{3.11}$$

В эффективном коэффициенте отбора S лучшего класса, первый (положительный) член обусловлен приспособленностью лучше среднего, а второй, отрицательный член обусловлен мутациями из этого класса. $M(t)$ имеет смысл эффективной частоты мутаций в двухаллельной модели. Как мы покажем ниже, знак S определяется знаком скорости v , положительной в случае адаптации, $v < 0$, и отрицательной в случае накопления вредных аллелей, $v > 0$ (храповик Мюллера).

Размер наиболее приспособленного класса зависит от времени пилообразным образом, потому что число вредных аллелей в лучшем классе k_0 меняется на единицу, когда новый класс приобретает или теряется. В режиме адаптации, наиболее приспособленный класс растет до тех пор, пока полезная мутация внутри него не приведет к рождению нового наиболее приспособленного класса (рисунок 3.2С).

Для согласования динамики лучшего класса с основной частью волны, $\log f_{k_0}(t)$, усредненный за один период «пилы» на рисунок 3.2С, должен был равен значению волны на краю, $\phi(x_0)$, полученному выше в § 3.1.2, т.е.

$$\phi(x_0) = \frac{1}{2} \left[\log \frac{1}{N|S|} + \log f_{max} \right]\tag{3.12}$$

где f_{max} — максимальная частота лучшего класса перед рождением следующего (рисунок 3.2С).

Выражение для f_{max} в (3.12) находится из условия, что лучший класс приобретает или теряется за период времени $1/|V|$, где

$$V \equiv -\frac{dk_{av}}{dt} = -(U_b + U_d)v \quad (3.13)$$

— скорость волны (частота замен). Конкретное выражение для f_{max} различается между случаями адаптации ($V > 0$) и храповика Мюллера ($V < 0$) следующим образом.

Адаптация, $V > 0$

В процессе адаптации, общая приспособленность популяции повышается. Новый наиболее приспособленный класс возникает в результате мутации в одном из геномов текущего наиболее приспособленного класса. Если он достигает размера порядка стохастического порога, что случается редко, новый класс устанавливается в популяции. Таким образом, волна перемещается на единицу по k за интервал времени $1/V$ (рисунок 3.2С).

Максимальная высота зубьев пилы $f_{max} = f_{k_0}(t = 1/V)$ в момент появления нового установленного аллеля оценивается следующим образом. Линия переживает генетический дрейф и достигает стохастического порога с небольшой вероятностью S [383, 384], так что среднее число новых аллелей, устанавливаемых в течение характерного времени роста лучшего класса $1/S$, равно

$$\frac{SU_b N f_{max}}{S} = U_b N f_{max}$$

По определению, за период устанавливается ровно один новый фитнес-класс, так что эта величина равна 1, откуда

$$f_{max} \approx \frac{1}{U_b N} \quad (3.14)$$

Храповик Мюллера, $V < 0$

В противоположном случае, когда популяция начинается с высокой приспособленности, любая мутация может только снизить приспособленность. В этом случае новые лучшие классы не могут родиться, а существующий лучший класс $k = k_0$ в конечном итоге теряется из-за случайного генетического дрейфа. В отсутствие выгодных мутаций и рекомбинации, эта потеря необратима, отсюда и термин «храповик» [69, 70]. После этого, все распределение приспособленности смещается в сторону большего k на единицу («щелчок храповика»). На стохастическом краю, зависимость от времени имеет пилообразный вид, аналогичный случаю адаптации на рисунок 3.2С, но с отрицательным наклоном: уменьшение лучшего класса, его потеря, переход к следующему и т.д. Эффективный коэффициент отбора S отрицателен при всех $0 < v$, что означает, что в этом режиме выбирается наиболее приспособленный класс. Даже небольшое

количество полезных мутаций может остановить храповик Мюллера [7, 385]. Однако, когда популяция все еще далека от равновесия и очень хорошо приспособлена, полезных мутаций слишком мало, и можно заменить $M(t) \rightarrow 0$. После интегрирования уравнения (3.11), получается затухающая динамика

$$f_{k_0}(t) = f_{max} e^{-|s|t} \gg 1/N|S| \quad (3.15)$$

где $t = 0$ обозначает момент потери предыдущего лучшего класса с $k = k_0 - 1$ аллелями. Ниже порога $f_{k_0}(t) < 1/(N|S|)$, дрейф становится доминирующим фактором, и класс $k = k_0$ вымирает с вероятностью, близкой к единице. Время, когда наиболее приспособленный класс проходит стохастический порог, t_{loss} , можно оценить из формулы $f_{k_0}(t_{loss}) = 1/|S|N$. Скорость волны определяется промежутком времени между потерей двух лучших классов, $t_{loss} = 1/(U_d v)$, откуда, с учетом (3.15), получается

$$f_{max} = \frac{1}{|S|N} e^{|s|/U_d v} \quad (3.16)$$

3.1.4 Результаты для средней частоты замен

Быстрая адаптация

Адаптация имеет место после того, как происходит резкое изменение условий среды, миграции или (для вирусов) иммунного ответа (§§ 2.1, 7.1, 7.2). Полезные мутации накапливаются до тех пор, пока популяция не достигнет стационарного состояния вблизи наилучшей последовательности, $k = 0$. Иными словами, адаптация имеет место, когда начальная приспособленность ниже, чем ее значение в стационарном состоянии. Ниже рассматривается случай вдали от стационарного состояния, когда частота замен много больше частоты мутаций

$$V = -(U_b + U_d)v \gg U_b + U_d$$

где вредными мутациями можно пренебречь. Я также предполагаю, что s больше, чем U_b , которое обычно довольно мало, $10^{-5} - 10^{-2}$ и что $V \gg s$, что выполняется для достаточно больших популяций. В этом пределе, все выражения для детерминистической части волны (3.5), (3.7) — (3.11) значительно упрощаются. Объединяя их с формулами (3.12) и (3.14) для лучшего класса, значение волны на краю принимает вид [11]

$$\phi(x_0) = -\log[N\sqrt{VU_b \log(V/U_b)}] \quad (3.17)$$

где числовая константа внутри большого логарифма приближена единицей. Этот результат, как уже упоминалось, основан на аппроксимации распределения приспособленности непрерывной функцией от k . Поправка к данному приближению из-за дискретности k получена в работе [11]. Добавляя этот поправочный член к $\phi(x_0) - \phi(0)$ и используя (3.17), окончательный результат

для средней частоты замен V имеет вид [11]

$$V \approx \frac{2s \log(N\sqrt{sU_b})}{\log^2\left[\frac{s}{U_b} \log(N\sqrt{sU_b})\right]} \gg U_b \quad (3.18)$$

Ускоренное накопление вредных аллелей

Теперь рассматривается противоположный предел, когда популяция начинается с максимальной приспособленности, $k = 0$, так что $M(t) \equiv 0$. Средняя средняя приспособленность будет снижаться до тех пор, пока отбор и полезная мутация не остановят этот процесс (устойчивое состояние). В хорошо адаптированных популяциях большинство потенциальных мутаций убивают приспособленность штамма и, следовательно, быстро исчезают из популяции. Однако многие вредные мутации имеют достаточно небольшой эффект приспособленности, $|s| \ll U_d$, и могут накапливаться в геноме до высоких уровней.

Полагая $\alpha = 0$ в выражении для u , выражения для детерминистической части волны (3.5), (3.7) — (3.9) упрощаются. Комбинируя их с формулами (3.12) и (3.16) для стохастического класса, получается [11]

$$\phi(x_0) = -\log[Nv^{3/2}U_d \log(e/v)] \quad (3.19)$$

где числовой коэффициент ~ 1 в аргументе логарифма опущен. Следующая поправка к $\phi(x_0) - \phi(0)$ из-за дискретности k получена в работе [11]. Используя ее и (3.19), окончательный результат имеет вид

$$\begin{aligned} \sigma \log\left(NU_d\sigma^{\frac{3}{2}}\right) &\approx \left[1 - \frac{v}{2}\left(\log^2\frac{e}{v} + 1\right)\right] \\ &- \sigma \log\left[\sqrt{\frac{v^3}{1-v}} \frac{\log(e/v)}{1-v \log(e/v) + 5\sigma/6}\right] \end{aligned} \quad (3.20)$$

Второй член в (3.20) — формально малая, но не пренебрежимо малая на практике, поправка к первому члену в пределе $\sigma \ll 1$. Аналитический результат подтверждается моделированием методом Монте-Карло (рисунок 3.5а).

Общий случай

Выше предполагалось, что популяция по приспособленности либо значительно ниже, либо выше стационарного состояния, так что либо вредными, либо полезными мутациями можно пренебречь. Когда популяция не очень далека от стационарного состояния, важны как те так и другие мутации. В общем случае, численность популяции N можно выразить через v и α следующим образом [7]

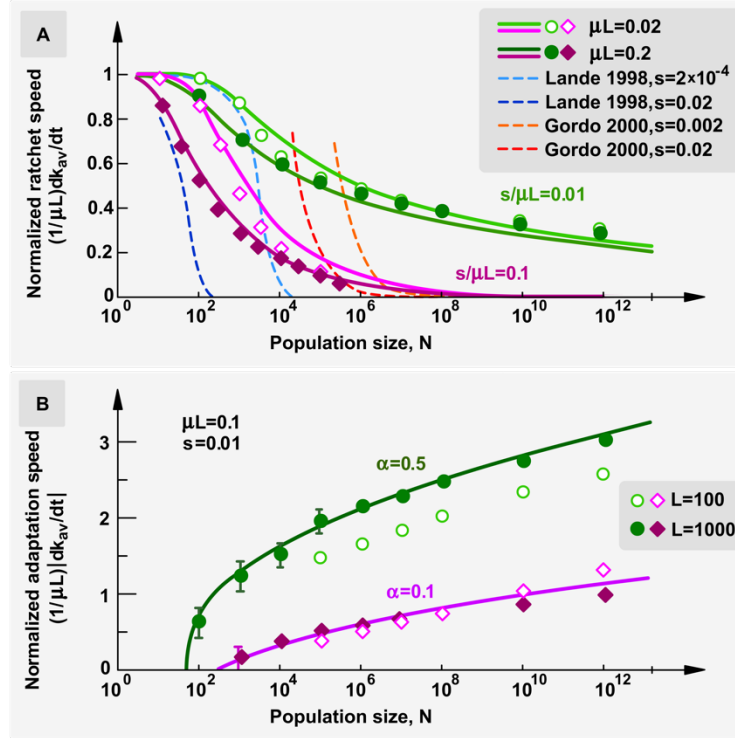


Рисунок 3.5. Сравнение аналитических результатов для частоты замен с результатами Монте Карло. (А) Нормализованная скорость храповика как функция размера популяции N : результаты анализа (сплошная линия) в сравнении с моделированием (символы). Полезные мутации отсутствуют, $\alpha = 0$. Фиолетовый: результаты для $\sigma = s/U_d = 0.1$, формула (3.20). Штриховые линии — асимптотика больших N из работы [[86], формулы (3а) и (3б)] и асимптотика малых N из работы [[83], выражение (2с), умноженное на NU_d]. Зеленый: результаты для $\sigma = s/U_d = 0.01$. Показаны параметры. (В) Нормализованная скорость адаптации как функция N из данных аналитических результатов (сплошная линия) и моделирования (символы). Зеленые и фиолетовые кривые соответствуют двум различным значениям менее подходящей аллельной фракции α (показано). Параметры, включая полную частоту мутаций μL , коэффициент отбора s и общий номер локуса L , показаны. По работе [7].

$$\sigma \log \left(\frac{N}{N^*} \right) \approx 1 - 2\alpha - v - \frac{v}{2} \log^2 u - v \log u - 2\alpha u \log u, \quad v < 1 - 2\alpha \quad (3.21)$$

где префактор N^* является степенной функцией $Var[k]$, U , α и v [[7], приложение, формулы (19)-(21)]. Лидер фитнес-волны x_0 определяется формулой (3.8). Аналитические результаты при различных значениях α и N подтверждаются моделированием Монте-Карло (рисунок 3.5В).

Устойчивое состояние

В конце концов, популяция достигает устойчивого состояния при значении α , где процессы адаптации и храповика Мюллера (§ 3.1.4) уравновешивают друг друга (рисунок 3.3С и D). Полагая $v = 0$ в формуле (3.21), получается уравнение для доли вредных аллелей α в равновесии [7]

$$\sigma \log(N/N^*) = 1 - 2\alpha - \sqrt{\alpha(1-\alpha)} \log \frac{1-\alpha}{\alpha}, \quad \sigma \log(N/N^*) < 1 \quad (3.22)$$

где $N^* = \alpha \sqrt{sU_d}$. За исключением крохотных популяций, эти значения α очень малы по сравнению с единицей, $\alpha \ll 1$. Другими словами, достаточно весьма небольшого количества

полезных мутаций, чтобы остановить храповик Мюллера [7, 385]. В больших популяциях, таких что $\sigma \log(N/N^*) > 1$, храповик Мюллера невозможен, и $\alpha = \mu/s$ как в одно-локусной модели (§ 2.1).

3.1.5 Заключение

Результат для средней частоты замен (3.18) подтверждает предсказания упрощенных моделей о том, что адаптация бесполой популяции замедляется эффектами сцепления [79, 380, 381] и, самое главное, показывает, до какой степени она замедляется. Скорость адаптации V не пропорциональна частоте мутаций или размеру популяции, как это было бы для независимых локусов. В широком интервале размеров популяции, $\log(1/s) < \log N \ll k \log(\sigma L)$, скорость адаптации гораздо меньше предсказания одно-локусной модели (§ 2.1) и очень медленно растет с численностью популяции N (рисунок 3.7). Только в пределе очень больших N , скорость адаптации (3.13) сшивается с однолокусным детерминистическим результатом, $V = s\bar{k}$, где $k/L \ll 1$. Переход к независимо-локусному режиму происходит при размерах популяции [7]

$$N_{one-locus} \sim (\sigma L)^{Ck}, \quad C \sim 1 \quad (3.23)$$

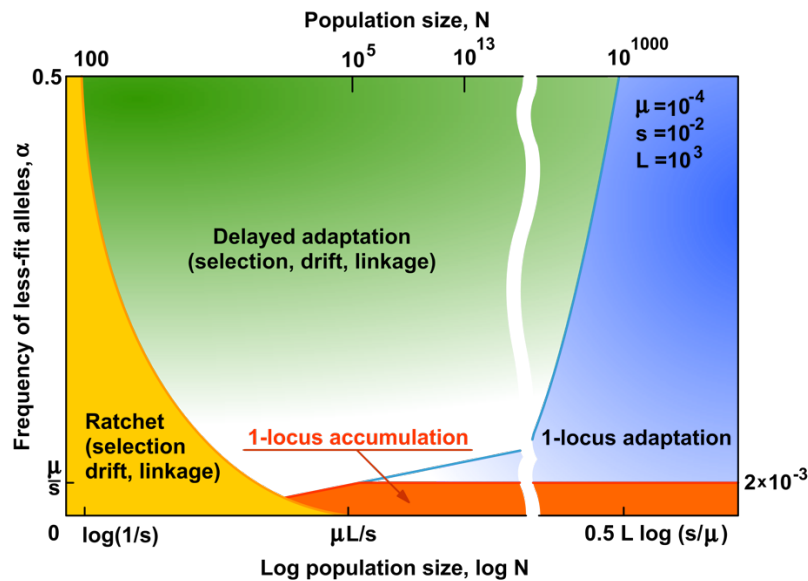


Рисунок 3.6. Схематическая фазовая диаграмма, показывающая границы однолокусного и многолокусных режимов, а также общее направление и доминирующие факторы эволюции эволюции, в зависимости от доли невыгодных аллелей и размера популяции. Верхняя ось и значения параметров (показаны) типичны для РНК-вирусов. Для организмов, частота мутаций меньше, $\mu \sim 10^{-9}$, а количество локусов больше, $L > 10^6 - 10^9$. По работе [7].

которые, для большого числа локусов, нереально велики. Количество полиморфных (сегрегирующих) локусов в большинстве организмов исчисляется миллионами, а в некоторых РНК-содержащих вирусах, таких как ВИЧ и SARS-CoV-2 — сотнями. Поэтому последняя оценка

$N_{one-locus}$, даже для вирусов с их короткими геномами, может легко превысить количество звезд во Вселенной. Рекомбинация, как мы покажем в следующей главе, чрезвычайно выгодна для адаптации, именно из-за устранения интерференционных эффектов сцепления. Границы монолокусного режима, общее направление и доминирующие факторы эволюции, в зависимости от доли вредных аллелей и численности популяции, показаны на рисунке 3.6.

Последующие исследования подтверждают точность результатов [7], включая улучшенную версию того же метода [11] и работы других исследователей [12, 72]. Результаты различных подходов [7, 11, 12, 72] сравниваются друг с другом (кривые) и с компьютерным моделированием (символы) на рисунке 3.7. Аналитический результат для стационарного состояния (3.21) был позже переполучен другой группой [385].

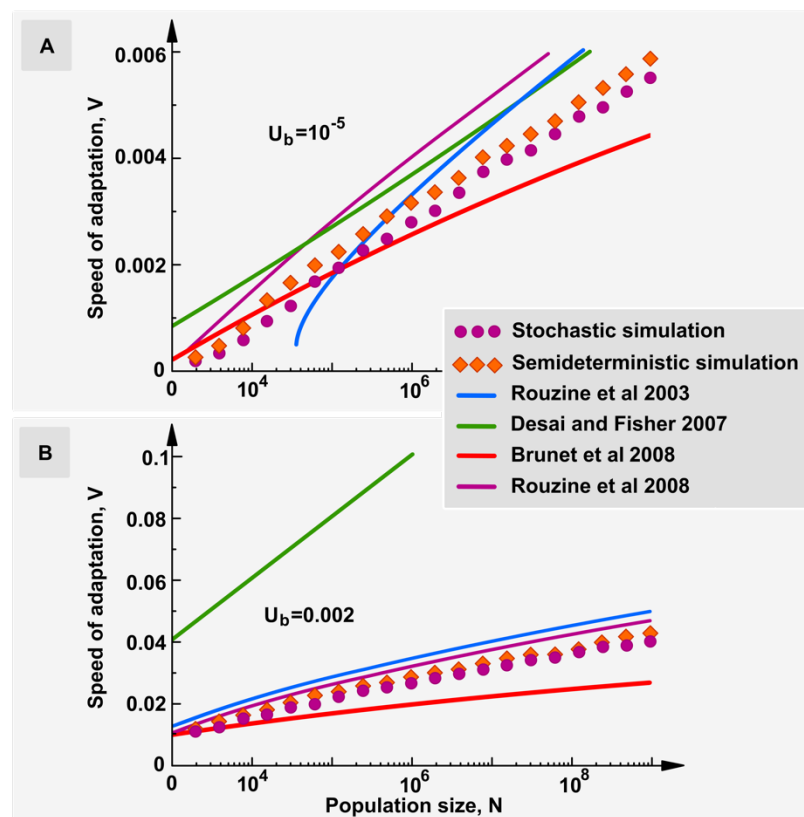


Рисунок 3.7. Сравнение четырех аналитических приближений для расчета скорости адаптации бесполой популяции и двух типов стохастического моделирования. Скорость адаптации увеличивается логарифмически с размером популяции N . (А, В): Кривые: аналитическая скорость адаптации, предсказанная двумя различными версиями теории бегущей волны и их улучшенными версиями: работа [7] (синий) и улучшенная версия [11] (фиолетовый); работа [72] (зеленый) и улучшенная версия [12] (красный). Символы: полная стохастическая симуляция (фиолетовые кружки) и полудетерминистическая симуляция, в которой стохастически рассматривается только наиболее приспособленный класс (оранжевые ромбы). Параметры: коэффициент отбора $s = 0.01$, частота полезных мутаций на геном U_b составляет (А) 10^{-5} и (В) 2×10^{-3} , вредных мутаций нет. По работе [12].

3.2 Многолокусная эволюция с распределенным коэффициентом отбора

В предыдущем параграфе, предполагалось, что все мутации имеют одинаковый эффект на приспособленность вируса, коэффициент отбора s . В реальных организмах, значение s меняется между локусами. Заранее не очевидно, усредняются ли локусы до некоего эффективного значения s [96], и какие мутации являются наиболее важными, с большими или малыми значениями s (рисунок 3.8). Исследования методом Монте-Карло [97] показали, что аппроксимация эффективного s может быть справедливой, если распределение s затухает при больших s экспоненциально или быстрее. Если хвост распределения затухает медленнее чем экспоненциальное, приближение многих мутаций с эффективным s не применимо, и лучшим приближением является модель, где зараз конкурируют только два клона, образованные выгодными мутациями на двух локусах [79]. Эта модель двух клонов была позже обобщена с учетом третьего, вложенного клона [81].

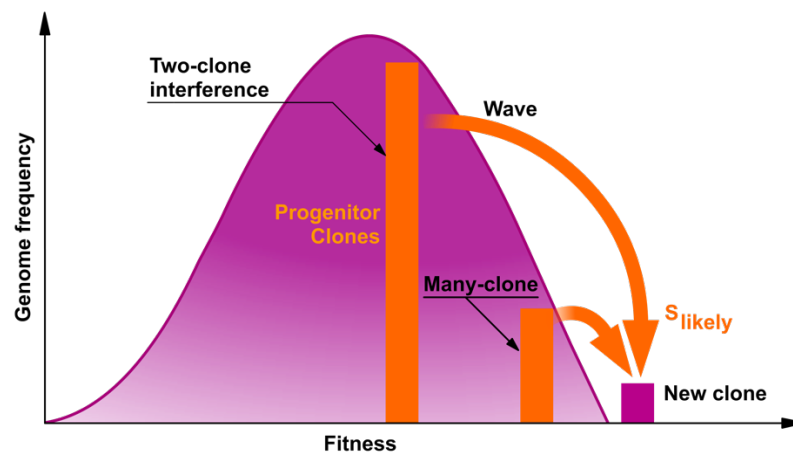


Рисунок 3.8. Механизм бесполой адаптации, когда коэффициент отбора s зависит от локуса. Механизм напоминает либо модель двухклоновой интерференции (§ 3.2) либо модель бегущей волны с фиксированным s (§§ 3.3–3.6) в зависимости от того, насколько крут хвост распределения s и насколько велик размер популяции [16]. Распределение в тексте рассматривается коэффициент отбора вида $\exp[-s/\sigma]$, где σ — постоянный параметр. Фиолетовая кривая: плотность распределения геномов по приспособленности. Оранжевые полосы: родители. Фиолетовая полоса: новый наиболее приспособленный клон. Из обзора [34].

Работа [16], представленная ниже в этом разделе, выполнена в сотрудничестве с тремя теоретическими группами. Рассматривается большое количество локусов с помощью трех параллельных методов: оригинального метода оптимальной мутации, метода марковской цепочки [76] и метода «настраиваемых моделей» [88]. Все три метода дают схожие результаты подтвержденные численно методом Монте Карло.

3.2.1 Модель

Как и в § 3.1, рассматривается популяция из N отдельных геномов, со средней частотой полезных мутаций U_b на геном. Предполагается, что эти мутации происходят на большом количестве локусов, аддитивно влияют на логарифм приспособленности генома, и выбраны из распределения коэффициентов отбора $\rho(s)$. Эпистаза и рекомбинации нет. Вредными мутациями с небольшим эффектом можно пренебречь, если система далека от стационарного состояния (§ 3.1). Вредные мутации с большим эффектом не важны для эволюции.

В оригинальной работе, рассмотрен случай общего распределения $\rho(s)$ [16]. Ниже рассматривается случай экспоненциального распределения, часто наблюдаемого в эксперименте [91-95], причины чего обсуждаются в § 5.4:

$$\rho(s) = \frac{1}{\sigma} e^{-\frac{s}{\sigma}}, \quad \sigma \gg U_b \quad (3.24)$$

Как показано в § 3.1, популяция развивается в уединенную волну, которая движется в координате приспособленности X с постоянной средней скоростью $v = d\bar{X}(t)/dt$ и формой $f(x)$, где $x = X - \bar{X}(t)$ — относительная приспособленность индивидуума по отношению к среднему приспособлению $\bar{X}(t)$ (рисунок 3.8). В § 3.1, $X = -sk$, где k — целое число вредных аллелей, существующих на фоне наиболее приспособленного к данным условиям генома. Здесь X может быть любым действительным числом.

Форма бегущей волны определяется, при постоянном s , полудетерминистическим методом использованным в § 3.1 [7, 8, 10, 11, 14]. В качестве альтернативы, его можно получить, используя другие варианты теории бегущей волны [12, 72, 88]. Однако все эти подходы менее удобны для случая распределенных s . Для данной цели, достаточно использовать гауссову аппроксимацию истинной формы профиля приспособленности. Как показано в § 3.1, волна близка к гауссовой с дисперсией, определяемой скоростью адаптации, уравнение (3.5):

$$f(x) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{x^2}{2v}} \quad (3.25)$$

где перемасштабированы обозначения $-\sigma v \rightarrow v$, $\sigma x \rightarrow x$. Скорость адаптации v , представляет собой скорость изменения приспособленности, в отличие от скорости замен менее приспособленных аллелей, используемой в § 3.1.

3.2.2 Вероятность установления генетической линии

Предположим, некий геном содержит полезный аллель, и мы хотим проследить его судьбу в популяции. Он может либо вымереть, либо выжить. Конкуренция аллеля в геноме с

приспособленностью X против остальной популяции зависит от временной зависимости средней приспособленности $\bar{X}(t)$, которая увеличивается с постоянной скоростью v в стационарном процессе. Пока вся линия потомков аллеля остается меньше размера популяции, мы можем анализировать ее поведение с точки зрения Марковского процесса с коэффициентом рождаемости $B(X, t) = 1 + X - \bar{X}(t)$ и смертностью 1, в выбранных единицах времени. Мутация увеличивает приспособленность от X до $X + s$ в интервале $[s, s + ds]$, с вероятностью $U_b \rho(s) ds$ на геном за поколение.

Скорость эволюции определяется вероятностью фиксации (невымирания) $w(X, t)$ линии генома с приспособленностью X в момент времени t . В простейшем случае, когда популяция невелика, а события фиксации аллеля редки и, следовательно, независимы в разных локусах, она равна $w = x$ [383]. Когда несколько событий фиксации происходят одновременно, они интерферируют друг с другом (§ 3.1). Когда генетическая линия, которая не прервалась, в конце концов захватывает большую часть популяции, она взаимодействует с другими линиями. Ниже предполагается, что все генетические линии, достигшие такого размера, уже гарантированно закрепляются в популяции, так что мы можем приравнять вероятность невымирания на низком уровне, $w(X, t)$, к вероятности закрепления линии.

Чтобы не прерваться, генетическая линия должна претерпеть множественные мутации для увеличения приспособленности. Это необходимо, чтобы избежать конкуренции с постоянно увеличивающейся средней приспособленностью населения. Такой процесс был описан в случае рекомбинации [76] и в случае распространения генов в популяциях, расширяющихся географически [386, 387]. Вероятность фиксации $w(x)$ удовлетворяет уравнению полученному в оригинальной работе [16]

$$v \frac{dw}{dx} = xw(x) - w(x)^2 + U_b \int_0^{\infty} ds \rho(s) [w(x+s) - w(x)] \quad (3.26)$$

3.2.3 Условие самосогласования скорости эволюции

Выживание аллельных линий связано с бегущей волной тем, что популяция все время адаптируется, генерируя новые мутации, которым удается закрепиться. Вероятность фиксации аллели с коэффициентом отбора s , обозначаемая $\pi(s)$, может быть получена из вероятности фиксации линии индивидуума x , $w(x)$, путем усреднения по распределению приспособленности генома $f(x)$, формула (3.25), в котором этот аллель появился

$$\pi(s) = \int_{-\infty}^{\infty} dx w(x) f(x-s) \quad (3.27)$$

где $w(x)$ находится из уравнения (3.26). Затем согласованность требует, чтобы скорость

адаптации v определялась средней скоростью фиксации новых мутаций, взвешенной по их эффекту приспособленности

$$v = NU_b \int_{-\infty}^{\infty} ds s \pi(s) \rho(s) \quad (3.28)$$

где под интегралом находится распределение фиксированных мутаций по их эффекту

$$\rho_f(s) \propto \pi(s) \rho(s) \quad (3.29)$$

Формулы (3.24)-(3.28) определяют распределение фиксированных мутаций по s и всю адаптационную динамику.

3.2.4 Вероятность фиксации и скорость адаптации

Когда размеры популяции велики, а мутации редки, решение (3.26) претерпевает резкое изменение при пороговой приспособленности $x = x_c$, выше которого оно близко к однолокусному результату, $w = x$ (рисунок 3.10), а ниже которого оно экспоненциально затухает. Можно кусочно аппроксимировать вероятность фиксации $w(x)$ как [16]

$$w(x) \approx \begin{cases} 0 & x < 0 \\ x_c e^{\frac{x^2 - x_c^2}{2v}} & 0 < x < x_c \\ x & x > x_c \end{cases} \quad (3.30)$$

где x_c определяется условием

$$2 = U_b \int_0^{\infty} ds \rho(s) \frac{e^{\frac{sx_c}{v}} - 1}{s} e^{-\frac{s^2}{2v}} + \frac{U_b}{vx_c} \int_{x_c}^{\infty} dx x e^{\frac{x_c^2 - x^2}{2v}} \int_0^{\infty} ds \rho(s) e^{-\frac{s^2}{2v} + \frac{xs}{v}} \quad (3.31)$$

Точка x_c аналогичная лидеру волны x_0 в § 3.1, интуитивно интерпретируется как граница приспособленности, выше которой клональная интерференция не снижает значительно вероятность фиксации (рисунок 3.9).

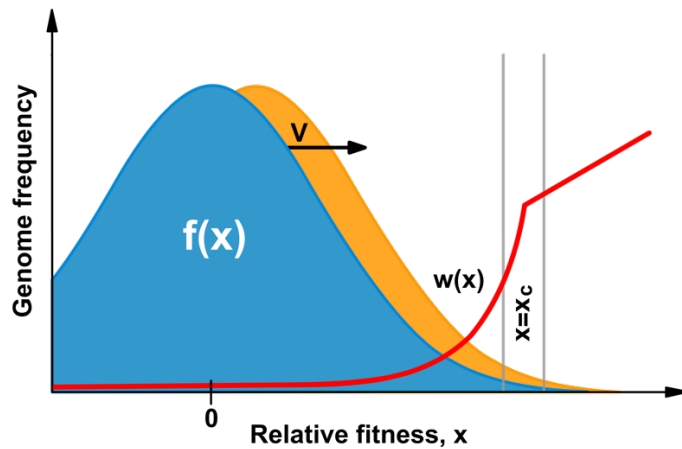


Рисунок 3.9. Процесс адаптации (схема). Плотность распределение геномов по приспособленности $f(x)$ движется с постоянной скоростью v . Вероятность фиксации $w(x)$, формула (3.30), резко возрастает с увеличением x , пока не достигает тонкого пограничного слоя $x = x_c$, после чего переходит к стандартному результату теории одного локуса (глава 2), $w(x) = x$. По работе [16].

Чтобы рассчитать среднюю вероятность фиксации $\pi(s)$, мы подставляем $w(x)$ из уравнения (3.30) в уравнение (3.27) и интегрируем, что дает

$$\pi(s) \propto \frac{e^{\frac{sx_c}{v}} - 1}{s} e^{-\frac{s^2}{2v}} + \frac{x_c^2}{vx_c} \int_{x_c}^{\infty} dx x e^{-\frac{(x-s)^2}{2v}} \quad (3.32)$$

Распределение фиксированных мутаций представляет собой произведение $\rho_f(s) \propto \pi(s)\rho(s)$ (рисунок 3.10). Интересно, что $\pi(s)$ имеет интервал эффективной нейтральности при $s < v/x_c$, где аллели фиксируются с вероятностью, примерно равной $\pi(s) = 1/N$ [81]. Выше этого характеристического значения, вероятность фиксации $\pi(s)$ экспоненциально возрастает с s до достижения предела одного локуса $\pi(s) = s$ при $s > x_c$.

Приближенные выражения для $w(x)$ и $\pi(s)$ содержат скорость адаптации v , которая самосогласованно получается путем подстановки (3.32) в (3.28). На рисунках 3.11-3.13, окончательные аналитические результаты данные ниже сопоставлены с результатами численного моделирования.

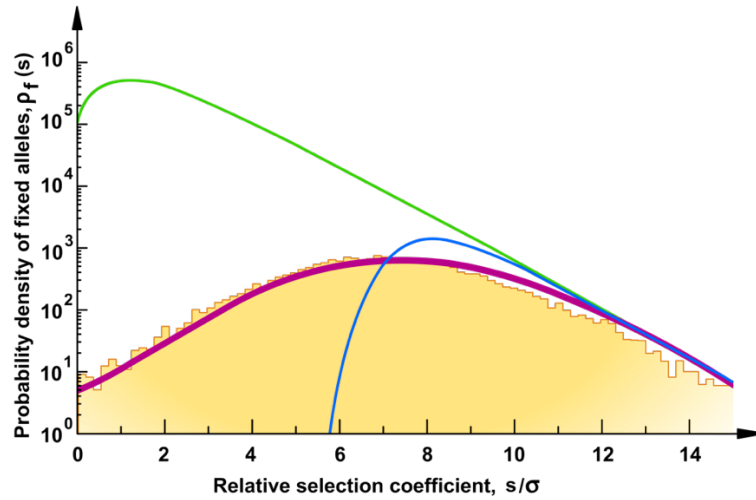


Рисунок 3.10. Распределение эффектов приспособленности фиксированных мутаций, $\rho_f(s)$. Сплошная фиолетовая линия: аналитический результат (3.29). Синяя линия: предсказание модели клональной интерференции [79]. Зеленая линия: предсказание однолокусной модели (глава 2). Параметры: $N = 10^7$, $U_b = 10^{-5}$, $\sigma = 0.01$. По работе [16].

Предполагая экспоненциальное распределение эффектов приспособленности (3.24), интегралы по коэффициенту отбора s в (3.31) и (3.28) имеют резкий пик при

$$s = s^* = x_c - \frac{v}{\sigma} \quad (3.33)$$

который представляет эффективный коэффициент отбора. Этот факт приводит к двум связанным уравнениям для v и x_c [16]

$$2 = \frac{U_b}{\sigma} \sqrt{\frac{2\pi\sigma^2}{v}} \left[1 + \frac{\sigma}{x_c} + \frac{v}{\sigma x_c - v} \right] e^{\frac{(x_c - v/\sigma)^2}{2v}} \quad (3.34)$$

$$1 = NU_b \left[\frac{x_c^2}{v} - 1 + \frac{2x_c\sigma}{v} + \frac{2\sigma^2}{v} \right] e^{-\frac{x_c - v/2\sigma}{\sigma}} \quad (3.35)$$

Численное решение уравнений (3.34) и (3.35) дает результаты, показанные на рисунке 3.11. Асимптотические аналитические выражения для этих величин допустимы в следующих двух важных пределах.

Если доминирующий эффект мутации относительно велик по сравнению с лидером волны, $s^* \sim x_c$ в (3.33), что имеет место при промежуточных значениях NU_b (см. ниже), то доминантные мутации представляют собой большие скачки приспособленности по сравнению с опережением распределения, $|x - x_c| \sim x_c$ по порядку величины (рисунок 3.8, левая оранжевая полоса). В этом случае, решение уравнения (3.34) и (3.35) итерациями дает

$$v \approx \frac{\sigma^2 \log^2(NU_b)}{2 \log\left(\frac{\sigma}{U_b}\right)}, \quad \log(NU_b) \ll 2 \log\left(\frac{\sigma}{U_b}\right)$$

который выполняется на промежуточном уровне NU_b . Здесь логарифмические коэффициенты внутри больших логарифмов опущены.

Напротив, при достаточно большом NU_b , наиболее важные мутации имеют относительно небольшой эффект по сравнению с лидером, $s^* \ll x_c$ в (3.33), и возникают в геномах, близких к точке перехода x_c (рисунок 3.8, правая оранжевая полоса). В этом случае, приближенное выражение для скорости адаптации полученное из уравнений (3.34) и (3.35) имеет вид

$$v = 2\sigma^2 \log[NU_b \log(NU_b)], \quad \log(NU_b) \gg 2 \log\left(\frac{\sigma}{U_b}\right)$$

Следовательно, для экспоненциального распределения эффектов приспособленности существуют два различных режима адаптации в двух интервалах размера популяции, как показано ниже. Один тип адаптации происходит за счет больших скачков приспособленности, сравнимых с опережением распределения, и он соответствует модели интерференции двух клонов [79]. Другой тип адаптации происходит посредством относительно небольших скачков и эквивалентен многолокусной модели с фиксированным s рассмотренной в § 3.1, где s и U_b заменяются эффективными значениями

$$s_{\text{eff}} = s^*, \quad U_{\text{eff}} \sim U_b \sqrt{2\pi v \rho(s^*)}$$

Можно показать, что второй тип существует в более общем случае, чем предполагаемый здесь экспоненциальный спектр мутаций [16]. Сравнение аналитических результатов с результатами моделирования Монте-Карло показано на рисунке 3.13.

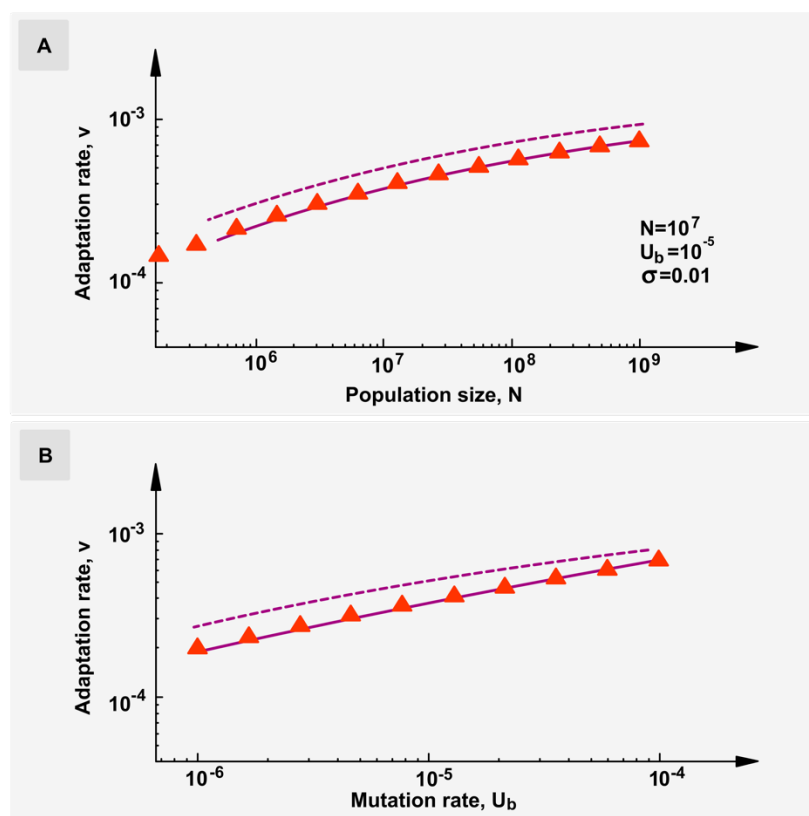


Рисунок 3.11. Скорость адаптации v как функция параметров модели. (A) v как функция размера популяции N . (B) v как функция вероятности полезной мутации U_b в случае относительно небольшого NU_b . Параметры: $\sigma = 0.01$ – средний коэффициент отбора (3.24). Символы и сплошная линия обозначают результаты численного моделирования и анализа соответственно (см. текст). Пунктирная линия: предсказание теории интерференции двух клонов (§ 3.2). По работе [16].

3.2.5 Заключение

Механизм адаптации и зависимость скорости адаптации от параметров модели существенно зависят от формы хвоста распределения s , а также от диапазона численности популяции (рисунок 3.9). Если спад хвоста происходит медленнее экспоненциального, важными мутациями являются те немногие, которые имеют наибольший возможный эффект приспособленности, подобно моделям интерференции с двумя или тремя клонами [79, 81]. Для более быстрого, чем экспоненциальное, затухания распределения s важны множественные основания, и результат сводится к результату адаптации при фиксированном s в § 3.1, но только с некоторым эффективным значением s , которое, в простейшем случае, выражается через три параметра модели: среднее значение s , размер популяции, и среднюю частоту выгодных мутаций. Наиболее интересным случаем является экспоненциальное убывание распределения s . В этом случае может иметь место любой режим, в зависимости от размера популяции: при умеренно малых размерах популяции, применяется двухлокусная модель [79], а при больших размерах популяции, актуальна многолокусная модель [16] (§ 3.1). В самых малых популяциях, эффекты

сцепления не важны, и выгодные мутации появляются и фиксируются без перекрытия во времени.

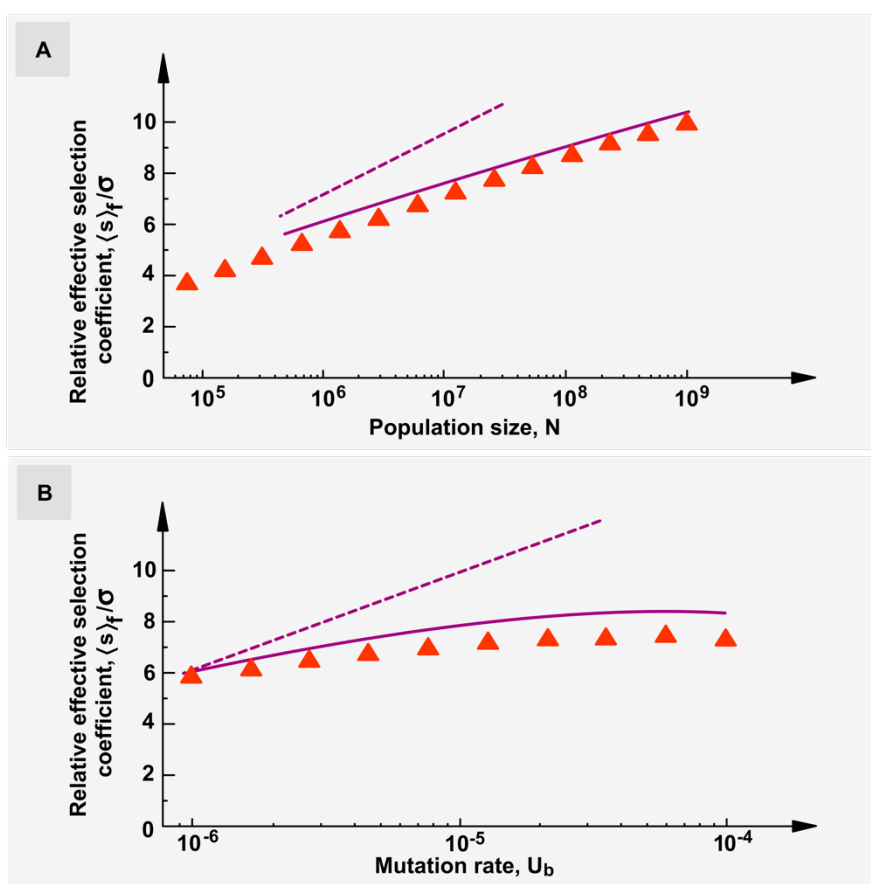


Рисунок 3.12. Средний эффект приспособленности фиксированной мутации s^* , в зависимости от параметров модели. Использована формула (3.33). (A) s^* как функция размера популяции N и (B) s^* как частоты полезных мутаций U_b . Оранжевые треугольники и фиолетовая сплошная кривая показывают результаты моделирования и анализа, соответственно. Фиолетовая пунктирная линия: результаты интерференционной модели с двумя клонами [79]. Параметры как на рисунке 3.11. По работе [16].

Глава 4. Многолокусная эволюция частично половой популяции

Отличием половых организмов от бесполовых является диплоидность (наличие двух наборов хромосом) и наличие рекомбинации. Во время спаривания, генетическая информация объединяется из двух родительских ДНК, по случайной половине от каждого, и передается потомству. Эволюционная функция рекомбинации заключается в создании новых, более приспособленных геномов, число которых увеличивается естественным отбором. Как показано в настоящей главе, рекомбинация может эффективно уменьшать клональную интерференцию и ускорять адаптацию на порядки величины, даже если рекомбинация происходит относительно редко.

Заметим, что диплоидность важна только для локусов с доминированием гомологичного аллеля. Если аллельное доминирование не в центре внимания, диплоидную популяцию из $N/2$ животных можно заменить гаплоидной из N геномов. Поэтому ниже рассматривается гаплоидная популяция с рекомбинацией.

Новым ключевым параметром является вероятность рекомбинации r , определяемая как вероятность на поколение, что геном имеет рекомбинацию с другим, случайно выбранным геномом (§ 2.3). Еще одним новым параметром является среднее количество кроссоверов на геном, M . В геноме человека $r = 1$, что соответствует исключительно половому размножению, $M = 2.5$ на хромосому, и 23 пары хромосом испытывают случайную перетасовку.

Роль рекомбинации заключается в соединении полезных аллелей в один геном и, наоборот, в удалении вредных аллелей. В работах вошедших в данную диссертацию [8-10, 14, 103], предполагается, что начальная популяция содержит небольшое количество полезных аллелей на рассматриваемых локусах. Так бывает, если окружающая среда претерпела недавние изменения, или вирус попал в нового хозяина, или появился иммунный ответ, или в организм введено лекарство. Предполагается, что выгодные аллели случайно разбросаны между локусами и геномами. Новыми мутациями пренебрегается, так рекомбинация создает новые последовательности гораздо быстрее чем мутация (см. ниже). Эти отдельные аллели объединяются, путем рекомбинации, в пары, тройки и т. д., что в конечном итоге приводит к появлению бегущей волны по координате приспособленности, в сторону более высоких значений приспособленности. Как и при эволюции бесполовых организмов (глава 3), скорость волны ограничена процессами на ее стохастическом фронте, где возникают новые геномы самой высокой приспособленности. В данной модели, они появляются благодаря рекомбинации, а не

мутации (рисунок 4.1).

Как показано ниже, филогенетическое родство последовательностей приводит к снижению эффективности рекомбинации [9, 10, 14]. Частоты полезных аллелей в одних локусах постепенно увеличиваются, а в других локусах полезные аллели со временем теряются из-за клональной интерференции (и случайного дрейфа). Когда все полезные аллели закрепятся в популяции или исчезнут, эволюция, в данной модели, закончивается. При этом оказывается, что, в достаточно длинном геноме, даже предельно частая рекомбинация не может компенсировать дальнедействующие эффекты сцепления между локусами и массовую потерю большинства полезных аллелей [103].

В долгосрочной перспективе, после того как все аллели на каждом локусе будут либо потеряны, либо закреплены, возникает новый режим, при котором редкая мутация необходима для внедрения новых аллелей [74, 76]. Рекомбинация способствует выживанию новых полезных аллелей, генерируемых мутациями, обеспечивая регулярные скачки аллелей из одного генома на другой, более приспособленный [10, 76].

4.1 Эволюция с рекомбинацией и начальным генетическим разнообразием

4.1.1 Приближение некоррелированных геномов

Первая модель [8] рассматривает популяцию из N гаплоидных геномов, имеющих большое количество локусов L (или диплоидную популяцию из $N/2$ особей без аллельного доминирования) (рисунок 4.1). Все локусы имеют два аллеля с небольшой разницей в приспособленности, $s \ll 1$. Поколения не перекрываются: все геномы умирают и заменяются их потомками. Приспособленность генома по отношению к наилучшей возможной приспособленности (число потомков) определяется выражением $\exp(-sk)$, где k — количество невыгодных аллелей. По определению, наилучший возможный геном имеет $k = 0$. Эпистаз (биохимическое взаимодействие между локусами, глава 5) отсутствует, и все аллели имеют одинаковый эффект приспособленности. Механизм рекомбинации зависит от организма. Рассматривается механизм кроссовера, распространённый у вирусов и животных, когда рекомбинация между двумя геномами происходит благодаря случайным переходам белка полимеразы между двумя матрицами РНК (Levy et al. 2004) (рисунок 4.1).

Размер вирусной популяции определяется общим числом вирусных геномов внутри клеток, N , которые производят новые инфекционные вирусные частицы, заражающие новые

клетки. Клетка, инфицированная двумя вирусными штаммами, может продуцировать рекомбинантные вирусные геномы (рисунок 4.1, внизу). Следовательно, фракция r всех геномов подвергнется рекомбинации с геномом другого вируса, а фракция $1 - r$ представляет собой копию единственного родительского генома. Ниже предполагается, что количество кроссоверов на геном M случайно меняется в соответствии с распределением Пуассона, и что рекомбинант состоит из примерно равной смеси каждого родительского генома. (Иногда частота рекомбинации определяется не для генома, а между двумя локусами, r_2 . Два определения связаны, как $r_2 = \frac{rM\Delta L}{L}$, где ΔL — количество пар оснований между двумя локусами.)

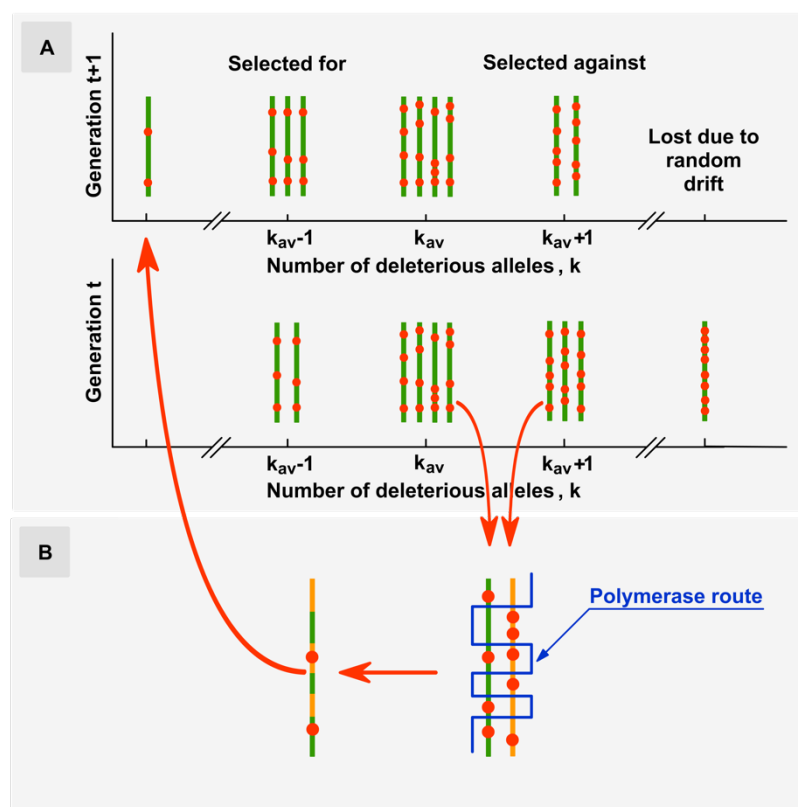


Рисунок 4.1. Модель эволюции при наличии естественного отбора, рекомбинации и случайного дрейфа. (А) Гаплоидная популяция в двух последовательных поколениях. Зеленые линии: геномы. Красные кружки: менее приспособленные аллели. (В) Механизм рекомбинации. Синяя пунктирная линия: маршрут РНК/ДНК-полимеразы между двумя родительскими геномами. По работе [8].

Важное предположение заключается в том, что, при заданном числе k невыгодных аллелей в геноме, они распределяются равномерно и случайным образом среди L локусов, и что их расположение между разными геномами никак не коррелирует. В § 4.2, это приближение будет снято, и корреляции между локусами будут учтены в численной модели. Там же будет показано, что геномы не коррелируют если $5 \ln(Nr f_0) \gg \sqrt{L}$. В пределах этого приближения, данный ниже анализ асимптотически точен в широком диапазоне параметров, задаваемом сильными неравенствами [8]:

$$1 \ll d \ll \frac{1}{s^2}$$

$$s \ll r \ll s\sqrt{d}$$

$$1 \ll \ln(Nr) \ll \min \left[d, \frac{1}{s^2 d}, d \left(\frac{s}{r} \right)^2 \log \left(\frac{s\sqrt{d}}{r} \right) \right]$$

Где $d = \bar{k} \left(1 - \frac{\bar{k}}{L} \right)$ — половина генетического расстояния (степень полиморфизма).

Эти условия не слишком ограничительны. На самом деле, при $d = 100 - 1000$ и $s = 0.01 - 0.05$, диапазон допустимых значений r, N весьма широк. Для быстро эволюционирующих вирусов, этот интервал для r является весьма широким, так их популяции обычно далеки от наилучшей последовательности, и количество вредных и полезных аллелей много больше 1.

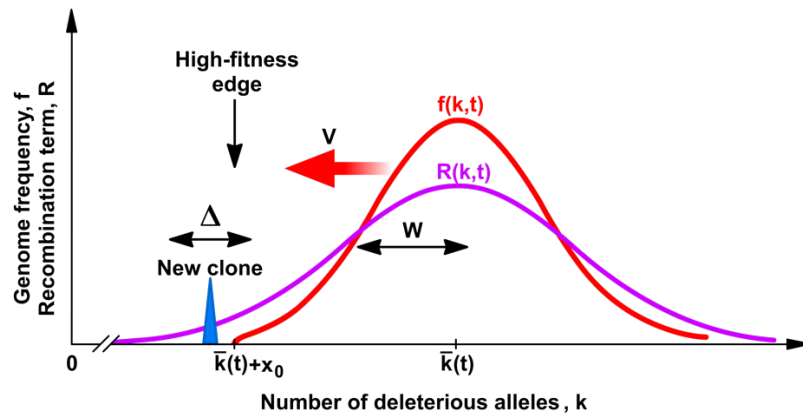


Рисунок 4.2. Схема движущейся уединенной волны. Красные и фиолетовые кривые: и тонкие линии, частота класса приспособленности, $f(k, t)$ и функция рекомбинационного генератора $R(k, t)$ соответственно. Синий пик: новый рекомбинантный клон, созданный за краем волны; Δ — интервал, в котором образуется наибольшее количество таких клонов; w и V — ширина и скорость (средняя частота замен) волны соответственно. По работе [8].

Обозначим через $f(k, t)$ среднюю частоту геномов с k вредными аллелями по отношению к наилучшей возможной последовательности. Для нашей модели рекомбинации, в интересующей нас области, динамическое уравнение для $f(k, t)$ имеет вид [8]

$$\begin{aligned} \partial f / \partial t &= -s[k - \bar{k}(t)]f(k, t) + r[R(k, t) - f(k, t)] \\ \bar{k}(t) &\equiv \int f(k, t)k dk \end{aligned} \quad (4.1)$$

где функция генерации рекомбинантов $R(k, t)$ имеет вид

$$R(k, t) = \frac{1}{\sqrt{\pi \bar{k}}} \int dk_1 \int dk_2 f(k_1, t) f(k_2, t) e^{-[k - \frac{k_1 + k_2}{2}]^2 / d} \quad (4.2)$$

Распределение приспособленности $f(k, t)$ и функция генерации рекомбинантов $R(k, t)$ показаны на рисунке 4.2.

4.1.2 Бегущая волна классов приспособленности

Уравнения (4.1) и (4.2) имеют решение вида бегущей волны

$$f(k, t) = \phi[k - \bar{k}(t)], R(k, t) = \rho[k - \bar{k}(t)] \quad (4.3)$$

Где среднее число вредных аллелей на геном $\bar{k}(t)$ уменьшается со временем, за счет рекомбинации и отбора. Подставляя уравнение (4.3) в (4.1) и (4.2), получается нелинейное интегро-дифференциальное уравнение для функции одной переменной

$$V \frac{\partial \phi}{\partial x} = -sx\phi(x) + r[\rho(x) - \phi(x)] \quad (4.4)$$

$$\rho(x) = \frac{1}{\sqrt{\pi \bar{k}}} \int dx_1 \int dx_2 \phi(x_1)\phi(x_2)e^{-[x-(x_1+x_2)/2]^2/d} \quad (4.5)$$

где относительное число аллелей $x \equiv k - \bar{k}(t)$ и средняя частота замен $V \equiv -d\bar{k}/dt$. Общее решение уравнения (4.4) имеет вид

$$\phi(x) = \frac{b}{w^2} e^{-(x+b)^2/2w^2} \int_{x_0}^x dx' \rho(x') e^{(x'+b)^2/(2w^2)} \quad (4.6)$$

где x_0 — константа, определенная ниже, и новые обозначения

$$\begin{aligned} b &\equiv \frac{r}{s} \\ w^2 &\equiv \frac{V}{s} \end{aligned} \quad (4.7)$$

При бесконечном размере популяции, $N = \infty$, уравнение (4.6) должно выполняться при любом значении x , даже очень далеко от центра волны. Отсюда имеем $x_0 = -\infty$, иначе функция $\phi(x)$ при $x < x_0$ была бы отрицательной. Единственное решение уравнений (4.5) и (4.6), для которых интеграл в (4.6) не расходится при $x' = -\infty$, имеет гауссовский вид с $w^2 = d$

$$\phi(x) = \rho(x) = \frac{1}{\sqrt{\pi \bar{k}}} e^{-\frac{x^2}{(2d)}}, \quad N = \infty \quad (4.8)$$

Для скорости волны V имеем

$$V = sw^2 = sd, \quad N = \infty \quad (4.9)$$

Формула (4.8) показывает, что разные аллели распределяются между геномами согласно распределению Пуассона, т. е. в пределе бесконечной популяции, локусы статистически независимы. Действительно, формула (4.9) — результат детерминистической однолокусной модели [§ 2.1, первый член в уравнении (2.4) при $f \ll 1$].

В конечной популяции, наиболее приспособленных последовательностей еще не

существует. Они добавляются постепенно на переднем фронте, $x = x_0$ (рисунок 4.2). При конечном x_0 , интеграл в формуле (4.6) не обязан сходиться при $x = -\infty$, и возможны значения ширины распределения w^2 , меньшие пуассоновского значения \bar{k} . Если общее число новых рекомбинантов в поколении велико, $Nr \gg 1$, то можно показать, что (4.6) принимает вид [8]

$$\phi(x) = \frac{1}{w\sqrt{2\pi}} e^{-\frac{(x+b)^2}{2w^2}}, \quad w^2 < d, \quad x_0 < x < |x_0| \quad (4.10)$$

$$\frac{b\sqrt{2\pi}}{w} \int_{x_0}^x dx' \rho(x') e^{-\frac{(x'+b)^2}{2w^2}} = 1 \quad (4.11)$$

где второе уравнение следует из (4.6) и условия нормировки $\int dx \phi(x) = 1$. Функция генерации рекомбинантов (4.5) тоже гауссовская, но с другой шириной

$$\rho(x) = \frac{1}{\sqrt{\pi(d+w^2)}} e^{-\frac{(x+b)^2}{d+w^2}}, \quad d > w^2 \quad (4.12)$$

Из (4.11) и (4.12), получается соотношение между лидером волны x_0 и ее скоростью (частотой замен) V

$$x_0^2 \approx d \frac{2p(1+p)}{1-p} \log \left[\frac{s\sqrt{d(1-p)}}{r} \right], \quad r \ll s\sqrt{d(1-p)} \quad (4.13)$$

где p это скорость волны нормализованная на результат однолокусной модели (предел $N = \infty$)

$$p \equiv \frac{w^2}{d} = \frac{V}{sd}$$

Как обычно, логарифмические и численные множители в аргументе больших логарифмов пренебрегаются.

4.1.3 Стохастический край высокой приспособленности и скорость адаптации

Пока что имеется одно соотношение для двух неизвестных величин, $|x_0|$ и V , формула (4.13). Чтобы получить второе уравнение, нужно изучить стохастическую динамику фронта волны. На рисунке 4.2 показано, как фронт волны надвигается со временем. Как и в § 3.1 для бесполой эволюции, рассматриваются два аллеля. Миноритарный аллель — это новый клон, формирующийся вблизи края волны, который имеет эффективный коэффициент отбора $S = s|x_0|$. Остальное население — это мажоритарный аллель. Новый рекомбинант появляется впереди фронта, при $x < x_0$. Частота его генерации равна $rN\rho(x)$. В не слишком маленьких популяциях, $Nr \gg 1$, лидер волны длиннее ширины, $|x_0| \gg w$. Большинство геномов за пределами края

рождаются в узком интервале x с шириной Δ

$$\Delta \sim \left| \frac{d \log \rho}{dx} \right|_{x=x_0}^{-1} \sim d/|x_0| \ll |x_0| \quad (4.14)$$

Частота образования этих рекомбинантов есть

$$G \sim rN\rho(x_0)\Delta \quad (4.15)$$

где $\rho(x_0)$ известно из (4.12). Следовательно, среднее время до установления нового клона равно

$$t_{seed} \sim 1/(GS) \quad (4.16)$$

где S — вероятность установления нового миноритарного аллеля несмотря на генетический дрейф (§ 2.2). Скорость края равно скорости волны

$$t_{seed} \sim \frac{\Delta}{V} \quad (4.17)$$

Формулы (4.7), (4.14) — (4.17) дают

$$x_0^2 \approx d(1+p) \log \frac{Nr}{p} \quad (4.18)$$

Наконец, из уравнений (4.13) и (4.18) для x_0^2 и p , окончательный результат для скорости эволюции (частоты замен) $V = s\bar{k}p$ имеет вид

$$V = sd \frac{\ln(Nr)}{\ln(Nr) + \ln\left(\frac{s^2 d}{r^2}\right)} < 1, \quad \frac{1}{N} \ll r \ll s\sqrt{d} \quad (4.19)$$

График этого результата как функции Nr показан на рисунке 4.3. Заметим, что в точке $Nr \sim 1$, адаптация полностью останавливается, а при $\ln(Nr) \gg \ln\left(\frac{s^2 d}{r^2}\right)$, скорость адаптации достигает предела независимых локусов.

4.1.4 Проверка результатов на компьютере методом Монте-Карло

Алгоритм Монте-Карло не был прямым моделированием популяции геномов, как в главе 2, а использовал то же приближение некоррелированных геномов, что и аналитическая часть. Программа запоминает число геномов $n(k, t)$ для каждого класса k в поколении t , $n(k, t) = Nf(k, t)$. При смене поколения, вычисляется среднее значение $\langle n(k, t+1) \rangle$ для всех $k = 1, \dots, L$ из формулы (4.1). Все классы со средним размером $\langle n(k, t+1) \rangle$ меньше заданного нижнего порога, $n_{emp} \ll 1$, опустошаются в следующем поколении. Новые размеры непустых классов, $n(k, t+1)$, генерируются случайным генератором согласно распределению Пуассона с рассчитанными выше $\langle n(k, t+1) \rangle$.

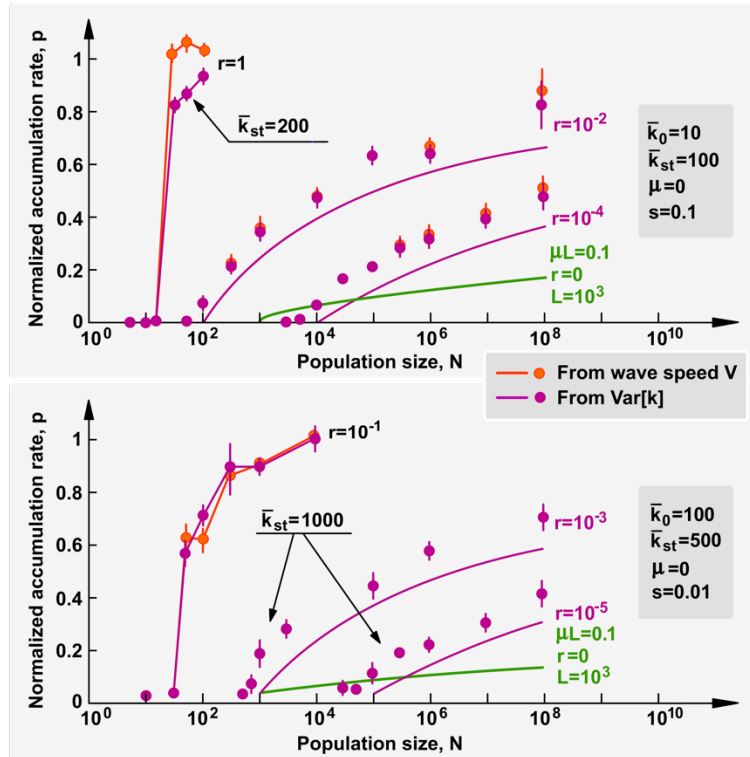


Рисунок 4.3. Средняя скорость V и квадрат ширины w^2 уединенной волны без мутаций ($\mu = 0$) в зависимости от размера популяции N . Обе величины нормализованы к их значениям в пределе бесконечного N . Оранжевые кружки: скорость волны $(dk/dt)/(s\bar{k})$; Фиолетовые кружки: квадрат ширины волны, w^2/\bar{k} . Коэффициент отбора $s = 0.1$ (вверху) и $s = 0.01$ (внизу). Вертикальные полосы: 67% статистический интервал для оценки среднего; фиолетовые линии, аналитический результат (4.19). Показаны средние числа аллелей в начале эволюции k_{st} и во время выборки k_0 . Значения r указаны на кривых. Результаты моделирования усредняются по прогонам: 40 (вверху) и 10 (внизу). Зеленые линии: результаты для бесполой популяции при $\mu = 10^{-4}$ и $L = 10^3$. По работе [8].

Логарифмически усредненная частота замен и квадрат ширины волны, поделенные на свои значения для независимых локусов, показаны на рисунке 4.3, в зависимости от N и r . В согласии с фундаментальной теоремой Фишера (4.7), нормированные значения $p = w_{av}^2/\bar{k}_0$ и $p = V_{av}/(s\bar{k}_0)$ очень похожи; они также близки к аналитическому предсказанию (4.19). Аналитическая теория несколько занижает результат моделирования, вероятно, из-за используемого в аналитической формуле непрерывного по k приближения (подобная ошибка наблюдалась в бесполом моделях в главе 2). При очень больших параметрах рекомбинации, $r \sim s\sqrt{k_0}$ и выше, приведенная выше аналитическая теория неприменима. Вместо этого мы наблюдаем крутой рост V от 0 до 1, когда размер популяции увеличивается на $N \approx 30/r$ (рисунок 4.3). На рисунке 4.4 показаны примеры смоделированных волн $f(k, t)$, динамика центра волны и средняя форма волны, соответствующая аналитическому прогнозу.

4.1.5 Заключение

Если начальная популяция имеет хотя бы небольшое количество полезных аллелей, адаптация, за счет редкой рекомбинации и естественного отбора, происходит гораздо быстрее, чем в отсутствие рекомбинации и за счет новых мутаций и естественного отбора. В бесполой эволюции, средняя частота замен достигает результата одно-локусной модели только в огромных популяциях, экспоненциально больших по параметру d , формула (2.23). Напротив, в присутствии рекомбинации, формула (4.19) достигает половины значения полученного в одно-локусной модели при весьма умеренном размере популяции, $N \sim (d/r)(s/r)^2$. Этот результат иллюстрирует фундаментальное эволюционное преимущество полового размножения.

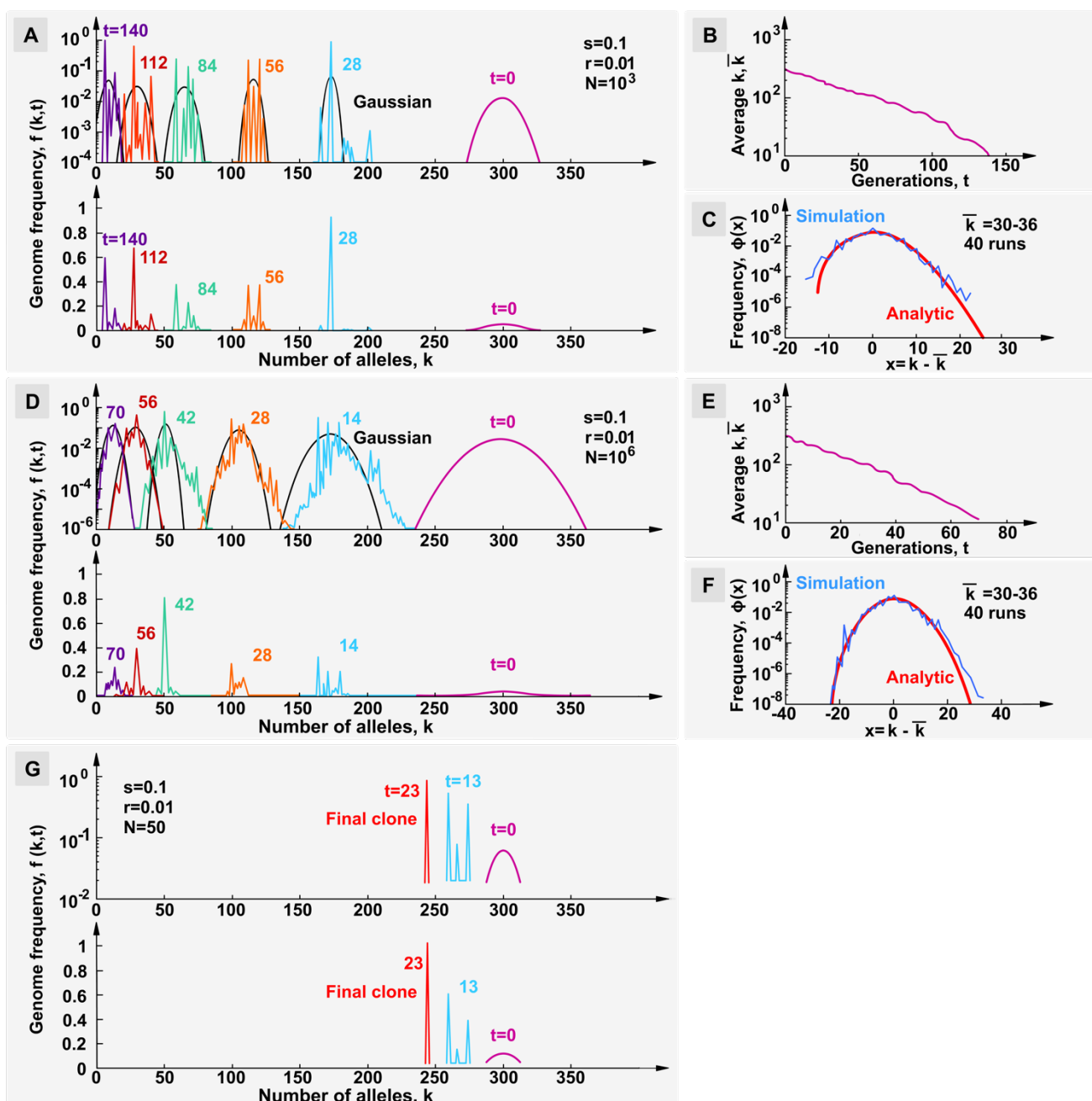


Рисунок 4.4. Примеры моделирования плотности распределения приспособленности $f(k, t)$ методом Монте-Карло с рекомбинацией и в отсутствие мутации. (А) Верх и низ: логарифмическая и линейная шкалы. Кривые в чередующихся цветах: $f(k, t)$ в разное время (показано). Черные линии: совпадение с функцией Гаусса. Показаны параметры модели. (В) Усредненное по населению число невыгодных аллелей \bar{k} , как функция времени. (С) Средняя центрированная форма волны $\phi(x)$. Синяя кривая: результат моделирования, усредненный по интервалу k (показан) и 40 прогонам. Красная линия: аналитический результат (4.10). (D–F) Аналогичные результаты для большего размера популяции, $N = 10^6$. (G) Моделирование ниже критического размера популяции, $Nr < 1$. По работе [8].

4.2 Эффекты общей родословной (филогенеза) на эволюцию с рекомбинацией

Важным ограничением вышерассмотренной модели является то, что она не учитывает корреляцию между геномами из-за общего происхождения части локусов. Предварительные работы [9, 10, 14], посвященные влиянию филогенеза на адаптацию с рекомбинацией, показали, что межгеномные корреляции, обусловленные общими предками, замедляют адаптацию и приводят к более резкой зависимости скорости адаптации от численности популяции и вероятности рекомбинации. Моделирование и анализ показал, что межгеномные корреляции между гомологичными локусами, обусловленные их семейным родством, нарастают во времени. Корреляции снижают эффективность рекомбинации, скорость адаптации и вызывают частичное исчезновение полезных аллелей. Причиной нарастающих корреляций является скрещивание родственных геномов. Одновременно, проявляются эффекты парного генетического сцепления. В работе обсуждаемой ниже [32], модель из § 4.1 изучена численно с учетом всех этих эффектов. Получено условие применимости приближения некоррелированных геномов использованного выше в § 4.1.

4.2.1 Модель Монте Карло

Как и в § 4.1, модель [32] включает эволюционными факторы: естественный отбор, случайный генетический дрейф, генетическое сцепление и рекомбинацию. Популяция, размножающаяся половым путем, состоит из N отдельных геномов (или $N/2$ диплоидных геномов без аллельного доминирования), где каждый геном имеет L локусов, и предполагается, что L намного больше единицы. Вначале каждый локус имеет долю f_0 полезных аллелей с фиксированным преимуществом приспособленности s и равномерно распределенных по всем сайтам и геномам. Предполагается, что значение f_0 находится в интервале $\frac{1}{Ns} \ll f_0 \ll 1$.

Эволюция моделируется в MATLAB™ с использованием процесса Райта-Фишера описанного в § 2.3. С некоторой вероятностью r , которая является входным параметром модели, геном

претерпевает случайные кроссинговеры с другим, случайно выбранным геномом. Среднее число кроссинговеров равно M . Один из двух родителей заменяется рекомбинантом.

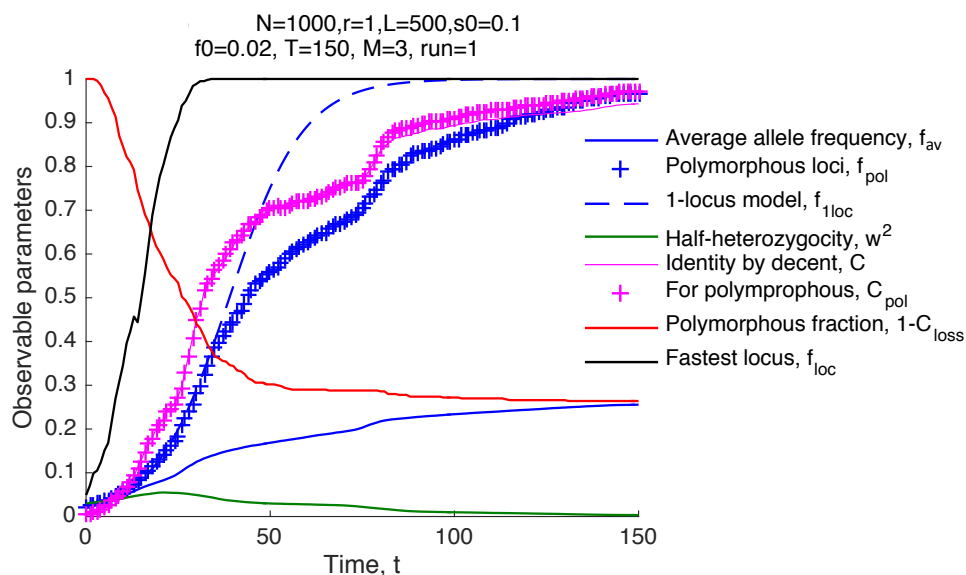


Рис. 4.5. Динамика наблюдаемых в модели со начальной генетической вариацией и рекомбинацией. Абсцисса показывает время в поколениях геномов. Ордината показывает наблюдаемые параметры: средняя частота полезных аллелей на locus на особь, f_{av} , то же значение, усредненное только по полиморфным locusам, f_{pol} , результат для f_{av} полученный в однолокусной модели, f_{1loc} , полу-генетическое расстояние $w^2 = \langle f(1 - f) \rangle$, доля гомологичных пар locusов с общим исходным предком C , одинаковое значение для полиморфных locusов C_{pol} , доля полиморфных locusов $1 - C_{loss}$ и наибольшая из частот аллелей среди locusов $\max(f_{loc})$. Значения параметров показаны. По работе [32].

Ниже предполагается $r = 1$, что соответствует обычному половому размножению, но все результаты легко обобщаются на случай любого r с помощью скейлинга во всех безразмерных выражениях: $(N, s) \rightarrow (Nr, s/r)$ [14, 76]. Параметры r и M могут быть связаны со средним количеством кроссинговеров между двумя сайтами r_2 , которые входят в модели с 2 locusами, как это задано $r_2 = \frac{rM}{L}$. Всего модель имеет 5 входных параметров (N, L, s, r, M) и начальное значение f_0 .

Новые события мутации отсутствуют. Эпистаз и аллельное доминирование не учитываются. Рассматривается гаплоидную популяцию. Как хорошо известно в популяционной генетике, диплоидная популяция с $N/2$ геномами и без доминирования фактически является гаплоидной, с двойным числом геномов N . Изученные диапазоны параметров: $s = [0.025, 0.4]$, $L = [10, 4000]$, $N = [10^2, 10^5]$, $M = [1, 300]$, $f_0 = [0.0001, 0.02]$.

4.2.2 Вымирание полезных аллелей и филогенетическая шкала времени зависят от одного составного параметра

В работе показано, что если число локусов L достаточно велико, полезные аллели в большинстве локусов в процессе эволюции вымирают. Доля оставшихся полиморфных локусов, обозначенная $1 - C_{loss}(t)$, уменьшается со временем от 1 до низкого плато (рисунок 4.5, красная кривая). Этот результат сильно отличается от предсказания однолокусной модели в изученном интервале $\frac{1}{Ns} \ll f_0 \ll 1$, что C_{loss} экспоненциально мало. Действительно, в однолокусной модели вероятность фиксации аллели равна s , а вероятность вымирания равна $1 - s$ [388]. Вероятность исчезновения Nf_0 выгодных аллелей, присутствующих вначале, есть $C_{loss}(\infty) = (1 - s)^{Nf_0} \approx e^{-Nf_0s}$. Таким образом, однолокусная модель предсказывает, что в каждом локусе будет зафиксировано множество аллелей, а потеря полиморфных локусов из-за генетического дрейфа [342] незначительна. Следовательно, наблюдаемая потеря не является эффектом генетического дрейфа в одно-локусной модели, а вызвана конкуренцией между клонами образованными полезными аллелями появившимися на разных локусах и геномах, которая имеется, несмотря на частую рекомбинацию [60, 65, 69, 389].

Варьируя параметры модели при моделировании, эмпирически установлено, что доля локусов с невымершими аллелями, $1 - C_{loss}(\infty)$, зависит в основном от одного составного параметра (рисунок 4.6А-С)

$$1 - C_{loss} = \begin{cases} 2.0 \frac{\log(Nf_0)}{\sqrt{L}} & 1 \ll \log(Nf_0) < 0.5\sqrt{L} \\ 1 & \log(Nf_0) > 0.5\sqrt{L} \end{cases} \quad (4.20)$$

Таким образом, имеется критическая точка $\log(Nf_0) = 0.5\sqrt{L}$, за пределами которой (размер популяции достаточно велик или количество локусов достаточно мало) заметной потери полиморфных локусов нет. Этот эффект генетического сцепления увеличивается с числом интерферирующих локусов, т. е. длины генома L , как и его эффект на скорость адаптации [7]. Величина $C_{loss}(t)$ примерно равна доля локусов, которые имеют общего предка для всех геномов (почти все из которых имеют в начале невыгодную аллель на локусе) [14].

Родственной величиной является идентичность по происхождению, $C(t)$, определяемая как вероятность того, что пара гомологичных локусов имеет одного и того же предка в момент времени t [14]. Эта величина не зависит от того, усреднена ли она по всем локусам или только по полиморфным локусам (рисунок 4.5). Результат симуляции показывает, что $C(t)$ больше, чем доля выгодных аллелей усредненная по полиморфным локусам $f_{pol}(t)$ (рисунок 4.5). В одно-локусной модели, общее происхождение пар геномов на локусе встречается куда реже, $C(t) < f_{pol}^2(t) < f_{pol}(t)$, потому что каждая пара локусов должна попасть в одну и ту же растущую генетическую линию, чтобы иметь одного и того же предка, а доля каждой линии меньше $f_{pol}(t)$.

В момент времени $t = T_2$ такой, что $C(T_2) = 1 - C_{loss}(T_2)$, обе величины близки к половине в широком диапазоне параметров (рисунок 4.5)

$$C(T_2) \approx C_{loss}(T_2) \approx 0.5$$

Этот результат является известным свойством многолокусной эволюции с естественным отбором, где среднее время до ближайшего предка для пары геномов и для всех геномов близки и определяются лидером бегущей волны (§§ 3.1 и 4.1) деленным на ее скорость [89, 390]. Для сравнения, в селективно-нейтральной эволюции это не так, и время до общего предка растет логарифмически с размером образца геномов [391, 392].

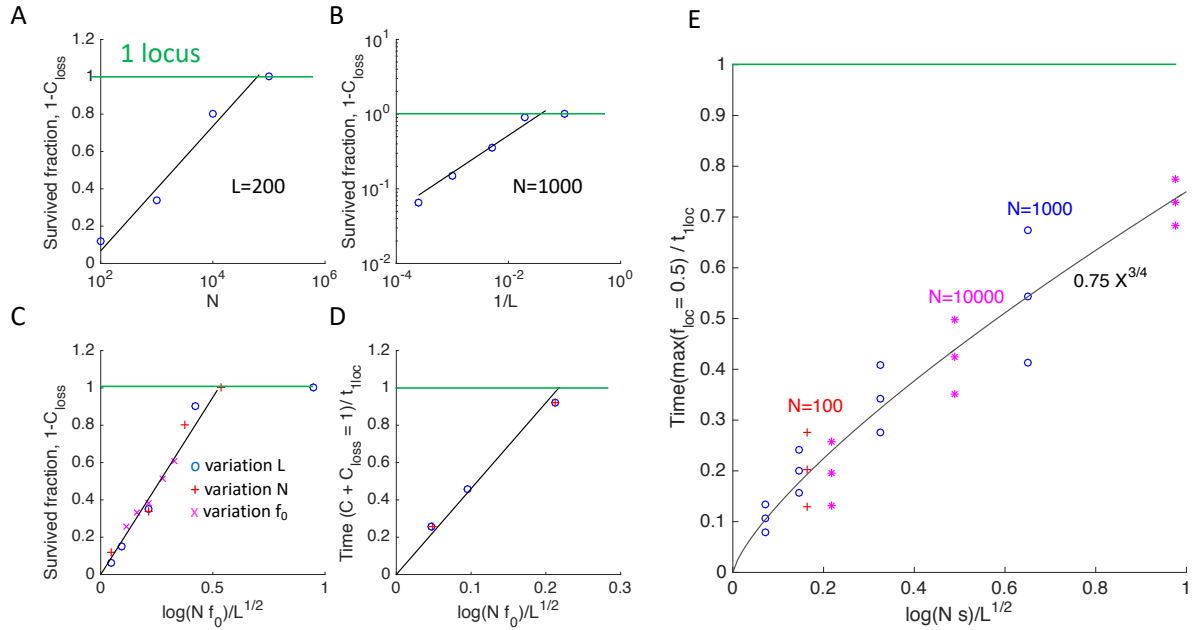


Рис. 4.6. Наблюдаемые в основном зависят от одного составного параметра.

A-C. Доля локуса, в которой выгодные аллели выжили и завершили адаптацию $1 - C_{loss}(\infty)$ пропорциональна составному параметру $\ln(Nf_0)/\sqrt{L}$ до критической точки когда он равен 0.5, после чего она перестает меняться. Цветные символы o, + и x соответствуют изменению параметров модели L , N и f_0 соответственно, где $f_0 > 1/Ns$. Зеленая горизонтальная линия показывает прогноз модели с одним локусом, $C_{loss} \approx 0$. **D.** Время t , когда доля выживших локусов $1 - C_{loss}(t)$ равна средней идентичности по происхождению $C(t)$ [пересечение красной и розовой кривых на рисунке 4.5] нормированное к времени полуадаптации в однолокусной модели. Эта величина также пропорциональна $\ln(Nf_0)/\sqrt{L}$, но не имеет критической точки, а продолжает увеличиваться выше 1 (не показано на рисунке). **E.** Время, когда частота аллелей в самом быстром локусе достигает 50%, как функция $\ln(Ns)/\sqrt{L}$. Тройки символов показывают среднее значение и 95% доверительный интервал. Цветные символы o, + и * показывают разные значения N . Чувствительность к изменению коэффициента отбора s и числа кроссинговеров M показана на рисунке 4.8. Значения параметров по умолчанию: $N = 1000$, $L = 200$, $f_0 = 0.02$, если не указано иное. Остальные параметры как на рисунке 4.5. По работе [32].

Время T_2 представляет приблизительную временную шкалу филогенетического дерева. Зависимость T_2 от параметров модели хорошо интерполируется формулой (рисунок 4.6D)

$$T_2 \approx t_{1loc} \frac{5.0 \log(Nf_0)}{\sqrt{L}} \quad (4.21)$$

$$t_{1loc} = \frac{1}{s} \log \frac{1}{f_0} \quad (4.22)$$

где t_{1loc} — время полуадаптации в однолокусной детерминистической модели (§ 2.1). Другими словами, время T_2 пропорционально тому же составному параметру, который контролирует долю успешных локусов, $1 - C_{loss}(\infty)$, определяемому формулой (4.20). Главная разница в том что T_2/t_{1loc} не имеет критической точки и может быть больше 1. Когда отношение времен T_2/t_{1loc} велико, что соответствует неравенству $5 \ln(Nf_0) \gg \sqrt{L}$, гомологичные локусы становятся родственниками достаточно поздно, когда фиксация выгодных аллелей почти завершена. В этом случае применим аналитический вывод в § 4.1, основанный на предположении о некоррелированных геномах. В противоположном пределе $5 \ln(Nf_0) < \sqrt{L}$, геномы в популяции имеют общих предков в большинстве гомологичных локусов уже посреди процесса адаптации, так предположение о некоррелированных геномах несправедливо, и нужно пользоваться численным моделированием. Здесь предположено $r = 1$ для полностью половой эволюции. В случае произвольного r , $0 < r < 1$, условие некоррелированности геномов имеет вид $5 \ln(Nrf_0) \gg \sqrt{L}$ (см. выше).

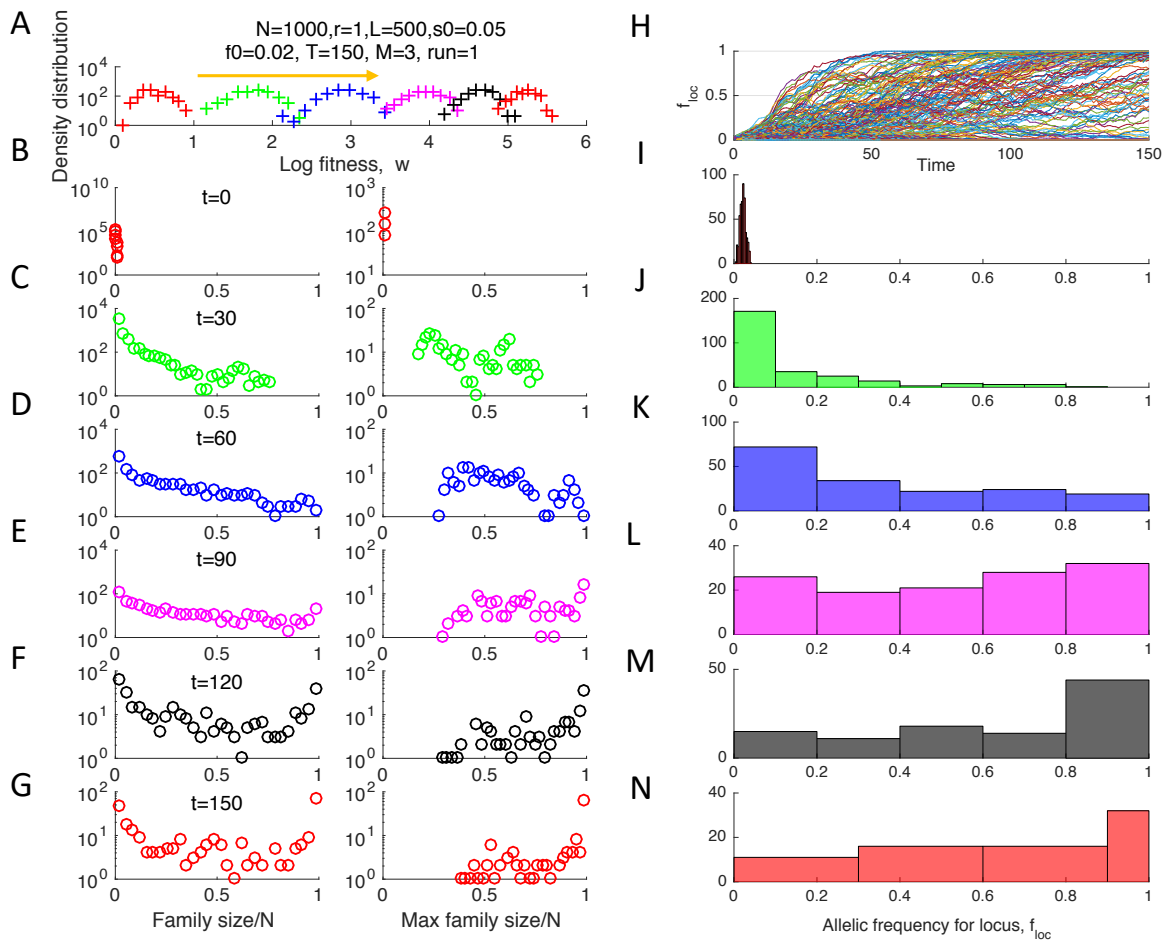


Рис. 4.7. Бегущая фитнес-волна и неравномерная динамика отдельных локусов. **А.** Плотность распределения геномов по приспособленности в разные моменты времени показана на (В-Г). **В-Г.** Первый столбец: гистограммы размера семейства, определенного как количество последовательностей с одним и тем же начальным предком в локусе. Второй столбец: Учитывается только самая крупная семья по локусу. **Н.** Средняя частота аллелей для каждого отдельного локуса f_{loc} как функция времени. **И-Н.** Гистограммы f_{loc} по локусам в разные моменты времени (показаны). Параметры как на рис. 4.5. По работе [32].

4.2.3 Сильная вариация скорости адаптации среди локусов

Поскольку при $2 \ln(Nf_0) \ll \sqrt{L}$ большинству локусов не удастся завершить адаптацию, средняя частота полезных аллелей на локус, $f_{av}(t)$, выходит на плато намного ниже 1 (рисунок 4.5). Частота аллелей, усредненная по полиморфным сайтам, $f_{pol}(t)$, увеличивается в том же общем временном диапазоне, что и результат модели с одним локусом, так что среднее время полуфиксации полиморфных сайтов $t_{50} \approx t_{1loc}$. В диапазоне параметров $s = 0.025 - 0.2, L = 200 - 2000, N = 1000 - 10000$ относительная разница между t_{50} и t_{1loc} мала и составляет от -0.11 до 0.14 . По сравнению с предсказанием однолокусной модели, зависимость $f_{pol}(t)$ испытывает задержку на поздних фазах адаптации и имеет заметную компоненту случайных колебаний (рисунок 4.5).

В некоторых локусах аллели накапливаются гораздо быстрее, чем предсказывает однолокусная модель (рисунок 4.5, черная кривая). Действительно, полупериод адаптации самого быстрого локуса $\max(t_{loc})$ намного короче, чем t_{1loc} (рисунок 4.6E). Сверхбыстрая эволюция некоторых локусов означает, что сегменты генома, содержащие эти локусы, имеют тенденцию иметь необычно большое количество полезных аллелей. Действительно, приспособленность генома пропорциональна этому числу.

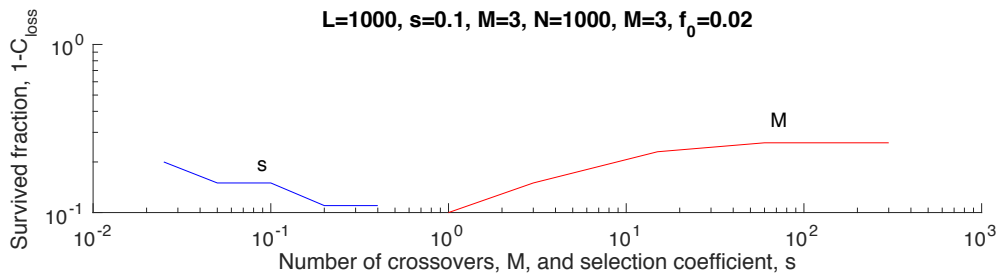
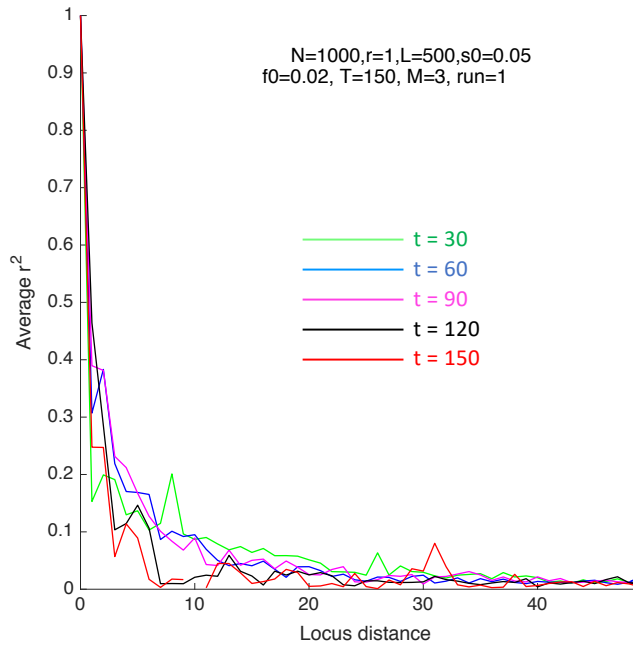


Рис. 4.8. Доля локусов, не потерявших и прошедших адаптацию, слабо чувствительна к среднему числу кроссинговеров и коэффициенту отбора. Значения параметров по умолчанию показаны вверху. По работе [32].

Другими словами, широкие различия в скорости эволюции между локусами с одинаковым коэффициентом отбора демонстрируют существование сильного эффекта “фонового отбора”, создаваемого событиями случайной рекомбинации. Рекомбинация объединяет разное количество благоприятных аллелей в разных сегментах, а естественный отбор отдает предпочтение наиболее приспособленным. Результирующее распределение геномов по приспособленности образует бегущую волну, хорошо известную как для бесполок, так и для популяций с половым размножением [388] (рисунок 4.7A). Новые геномы на краю волны высокой приспособленности рождаются путем рекомбинации. Их число растет быстрее, чем для среднего генома, заставляя волну двигаться вперед [8].

Популяция имеет сложную родословную структуру, которая варьируется в зависимости от локуса. Для данного локуса, генетическая линия определяется как совокупность особей, имеющих одного и того же начального предка. Все генетические линии изначально состоят из



одной особи основателя (рисунок 4.7В), но их размеры растут со временем, с разной скоростью для разных локусов, и распределены в очень широком диапазоне (рисунок 4.7С-Г). Распределение размеров генетических линий между локусами со временем смещается в сторону более крупных генетических линий, в конечном итоге занимающих почти всю популяцию. Если мы примем во внимание только крупнейшую генетическую линию для каждого локуса, их распределение по размерам выглядит схожим, но имеет

нижнюю отсечку растущую со временем (рисунок 4.7В-Г, столбец 2). Крупнейшие линии вырастают до половины популяции гораздо раньше, чем получается в однолокусной модели, t_{1loc} в формуле (4.22).

Рисунок 4.9. Корреляция между аллелями быстро убывает с увеличением расстояния между локусами. Ордината: коэффициент Пирсона r , формула (4.23). Моменты времени и параметры (показаны) такие же, как на рисунке 4.7. При $t = 0$ неравновесие по сцеплению тождественно равно нулю из-за первоначально случайного распределения аллелей. По работе [32].

Приведенные выше результаты на рисунках 4.5 to 4.7 слабо чувствительны к среднему числу кроссинговеров M . Во всем диапазоне от 1 до L доля локусов, не теряющих аллели, $1 - C_{loss}(\infty)$, варьируется примерно в 2 раза (рисунок 4.8). Изменение коэффициента отбора s меняет только масштаб времени, его эффект на $C_{loss}(\infty)$ также невелик.

4.2.4 Отсутствие аллельной корреляции в удаленных локусах

Коррелятор Пирсона между частотами аллелей в двух локусах вычислялся по формуле

$$r(l_{12}) = \frac{\langle (f_1 - \langle f \rangle)(f_2 - \langle f \rangle) \rangle}{\langle (f_1 - \langle f \rangle)^2 \rangle} \quad (4.23)$$

где $\langle \dots \rangle$ означает усреднение по парам достаточно гетерозиготных локусов, $2f_{loc}(1 - f_{loc}) > 0.1$. Значение r , определенное формулой (4.23), зависит от времени следующим образом. Первоначально $r \equiv 0$ из-за случайного начального распределения аллелей, заданного при

моделировании. По прошествии некоторого времени r становится положительным при любом расстоянии между локусами (рисунок 4.9, $t = 30$). После прохождения максимума r снова уменьшается во времени. В каждый следующий момент времени r все более резко зависит от расстояния между локусами l_{12} (рисунок 4.9). Другими словами, аллели в удаленных локусах разделяются независимо, как и ожидалось при наличии рекомбинации (закон Моргана).

4.2.5 Аллели фиксируются взаимозависимо

Вероятность фиксации выгодного аллеля определяется выражением

$$P_{fix} = \frac{\log [1/C_{loss}(\infty)]}{Nf_0} \approx \frac{1 - C_{loss}(\infty)}{Nf_0} \quad (4.24)$$

В заданном интервале $f_0 \gg \frac{1}{Ns}$, это значение намного ниже результата однолокусной модели, s . Значение P_{fix} достигает $P_{fix} = s$ только в пределе достаточно малого f_0 [76]. Точку перехода можно определить из условия $P_{fix}(f_0^{dilute}) = s$ и уравнений (4.20) и (4.24), откуда

$$f_0^{dilute} \sim \frac{1}{Ns\sqrt{L}} \log \frac{1}{s\sqrt{L}}, \quad s\sqrt{L} \ll 1 \quad (4.25)$$

Формулы (4.24) и (4.25) согласуется с результатами данного моделирования ([32], рисунок S1).

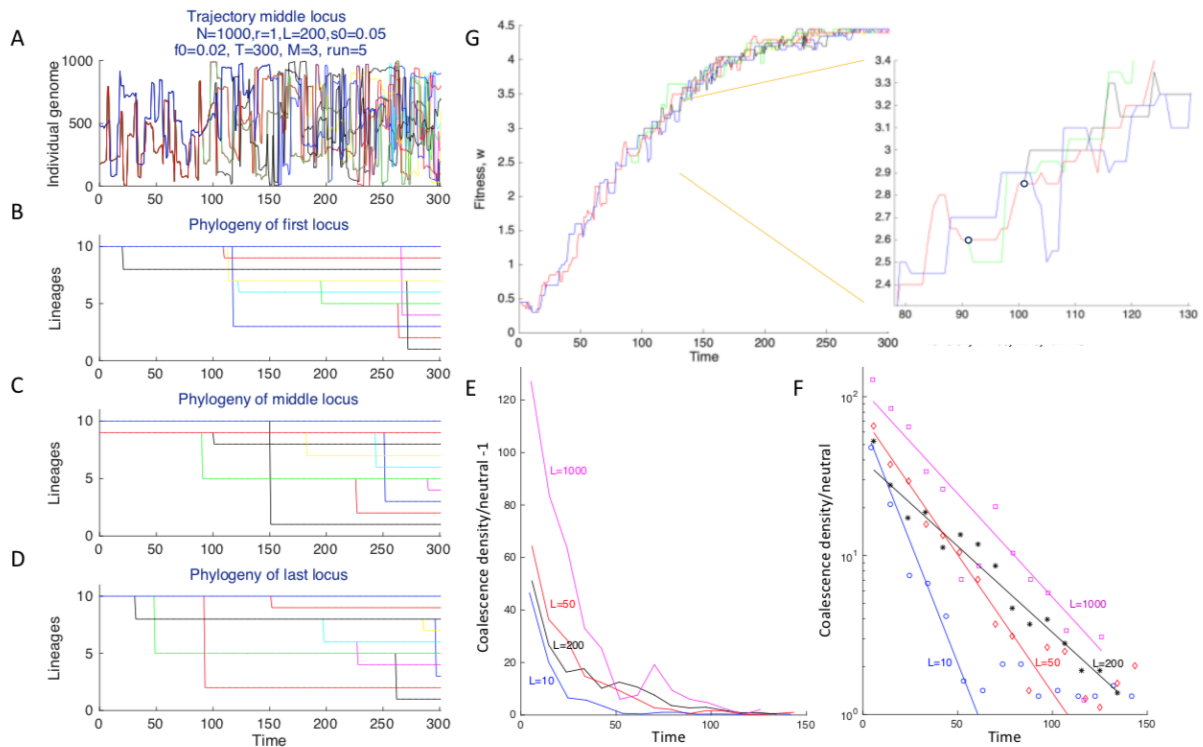


Рисунок 4.10. Филогенетическое древо и предковая история отдельных локусов. **А.** Траектория среднего локуса в обратном времени ($i = L/2$) у 10 особей с номерами 1, 101, 201, ... 901 во время $t = 300$, полученная путем отслеживания истории предков. **В-Д.** Филогенетические деревья для трех локусов (первого, среднего и последнего).

Е-Ф. Плотность времени слившихся событий усреднена по 10 запускам моделирования и нормализована к их значениям, предсказанным выборочно нейтральной моделью. Для оси Y используются линейная (Е) и логарифмическая (F) шкалы. Г. Траектории приспособленности для среднего локуса в (А, С). Вставка: небольшой сегмент показан оранжевым квадратом. Параметры указаны на (А), если не указано иное. По работе [32].

4.2.6 Генеалогия, филогения и траектория приспособленности

Помимо расчета шкалы времени филогении T_2 , построена генеалогия трех разных далеких локусов. Родитель каждого локуса записан в процессе моделирования. Траектория меняется между случайными особями из-за рекомбинации (рисунок 4.10А). Если выпрямить эти траектории и сохранить только топологию и время слияния линий, получатся филогенетические деревья для каждого локуса (рисунок 4.10В-Д). Как и ожидается, дерево сильно различается между локусами из-за рекомбинации, а ранние ветви намного короче поздних ветвей, что есть результат естественного отбора [393]. Средняя плотность событий слияния усредненная по 10 прогонам экспоненциально убывает со временем (рисунок 4.10Е, F). Это связано с тем, что плотность слияния событий пропорциональна размеру растущей субпопуляции вариантов [393]. Важно отметить, что плотность слияния значительно больше, чем в однолокусном пределе, и увеличивается с увеличением числа локусов L . Таким образом, в согласии с предыдущими исследованиями для бесполой эволюции, некомпенсированное генетическое сцепление при наличии отбора уплотняет филогенетические деревья и изменяет их форму, делая ранние стебли короче [89, 390, 394, 395] (рисунок 4.10Е, F).

В дополнение к генеалогии, т.е., номерам предков данного локуса как функции времени (рисунок 4.10А), построена типичная траектория приспособленности его предков (рисунок 4.10G). Она состоит из прямых горизонтальных сегментов за счет клональной экспансии и скачков, вызванными рекомбинацией. Скачки происходят в обе стороны, но чаще в сторону более высокой приспособленностью (рисунок 4.10G). Такое поведение «аллельного серфинга» было предсказано для половых популяций аналитически [10, 76]

4.2.7 Сравнение с другими моделями и применение к SARS-CoV-2

Если вначале генерируется незначительное количество полезных аллелей, например, из-за изменения внешней среды или из-за ранней мутации, эволюция может происходить некоторое время при отсутствии новых мутаций [8]. Такая эволюция, основанная на постоянной изменчивости и отборе, с рекомбинацией или без нее, наблюдалась для полиовируса [22] и VSV [13]. Если аллель сразу не потеряна из-за случайного генетического дрейфа, ее дальнейшее

накопление происходит благодаря совместной работе естественного отбора и рекомбинации. Процесс продолжается до тех пор, пока все аллели не зафиксированы или не исчезнут.

Хотя частота аллелей не коррелирует между удаленными локусами, в соответствии с законом Моргана, данное моделирование предсказывает существование сильных эффектов дальнего генетического сцепления, охватывающих весь геном. Эффекты включают исчезновение полезных аллелей в большинстве локусов из-за не случайного дрейфа, а клональной интерференции, слабую чувствительность результатов к числу кроссинговеров и сверхбыструю эволюцию в некоторых локусах, даже быстрее, чем в пределе независимых локусов. Последнее наблюдение предполагает существование геномных сегментов, обогащенных полезными аллелями в среднем выше среднего. В целом эти результаты означают, что удаленные области генома не развиваются независимо, а рекомбинация недостаточно сильна, чтобы разрушить эффекты сцепления, вызванные отбором, действующим, таким образом, на уровне целого генома, а также на уровне геномные сегменты.

При уменьшении числа локусов или увеличении размера популяции прогнозируется переход к пределу независимых локусов. Предсказанная зависимость всех эффектов сцепления от размера популяции N является логарифмической (рисунок 4.6). Для генома из 200 локусов и $f_0 = 0.02$, $s = 0.1$ переход к режиму независимых локусов можно наблюдать в популяции из 100 000 особей. В более длинном геноме из 1000 локусов, они будут эволюционировать независимо только для популяций численностью 10^{12} особей и более, что для большинства видов нереально. Популяция человека или животных имеет миллионы варибельных локусов, значительная часть которых находится под действием естественного отбора, так что модели независимых локусов, вероятно, никогда не работают в популяциях животных, за исключением очень редких мутаций, находящихся под очень сильным давлением отбора.

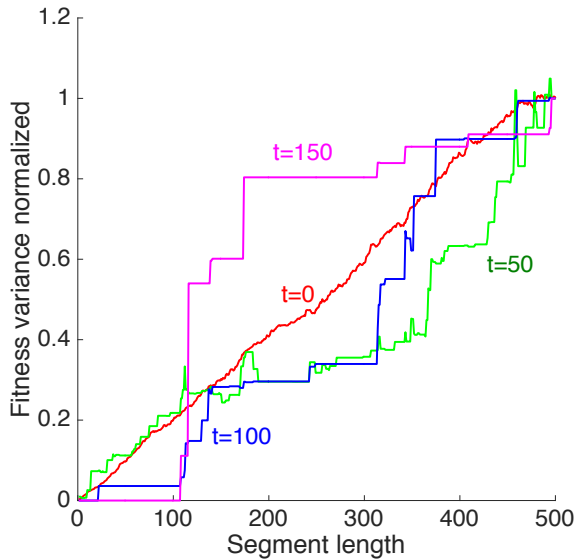
Результаты настоящего исследования, проведенного для среднесрочной эволюции с начальным разнообразием, когда мутации неважны, резко контрастируют с результатами полученными для долговременной эволюции, которая существенно зависит от синергии между новых мутациями и рекомбинации. После того как среднесрочная эволюция заканчивается потерей или фиксацией всех изначально присутствующих аллелей, дальнейшая эволюция требует постоянного притока новых мутаций. Для этой модели, было показано, что геном состоит из независимо эволюционирующих сегментов [74, 75, 396]. В обозначениях данного раздела, результат для среднего времени до последнего общего предка имеет вид [[74], уравнение 5]

$$T_{MRC A} \approx c \frac{M}{v} \log \left(\frac{Nv}{M} \right) \quad (4.26)$$

где v — скорость адаптации (прирост средней приспособленности в единицу времени), c — число порядка 1, а логарифм предполагается намного больше 1. В модели обсуждаемой здесь, T_{MRCA} по порядку величины равен T_2 из формулы (4.21), а скорость адаптации

$$v \approx \text{const} \frac{sL(1 - C_{loss})}{t_{50}} \quad (4.27)$$

Как уже упоминалось, среднее время полуфиксации для локусов, не теряющих аллели, t_{50} , всегда



близко к результату модели одного локуса t_{1loc} , формула (4.22). Подстановка уравнений (4.21), (4.22) и (4.27) в уравнение (4.26) дает

$$\frac{M \log(Nv/M)}{s \log^2(Nf_0)} = c$$

Последнее равенство явно неверно, поскольку M, N, f_0 и s — независимые параметры. Следовательно, формула (4.26) в данном случае не работает.

Рис. 4.11. Нелинейная зависимость дисперсии сегмента генома от длины сегмента. Абсцисса: длина сегмента генома, начиная с локуса 1. Ордината: вариация

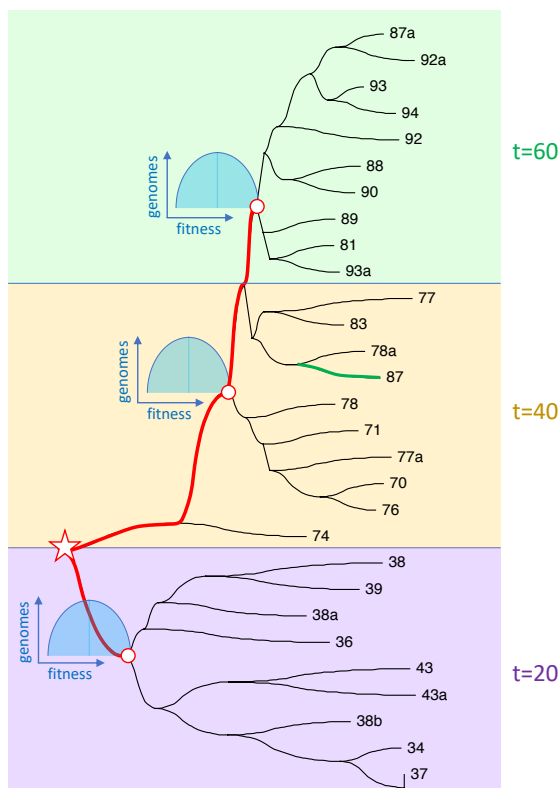
приспособленности между гомологичными геномными сегментами, деленная на вариацию приспособленности генома в один и тот же момент времени. Показан одиночный запуск Monte-Carlo. Параметры как на рисунке 4.5. По работе [32].

Вывод (4.26) в [74] был основан на двух утверждениях: (а) геном развивается как квазинезависимые бесполое сегменты без рекомбинации, (б) выражении времени до самого последнего общего предка для бесполой эволюции. Последнее подтверждено для различных многолокусных моделей. Поэтому вполне вероятно, что причиной расхождения является предположение о квазинезависимости. Другими словами, в среднесрочной эволюции, геном не развивается как набор квазинезависимых сегментов. Этот вывод подтверждается потерей аллелей из-за клональной интерференции на дальних локусах, рисунки 4.5 и 4.6А-С, а также результатами доказывающими, что полезные аллели могут образовывать хорошо приспособленные геномы, быстрый рост которых опережает смешение геномов за счет рекомбинации (рис. 4.7). Это подтверждается также тем что, в пределах одной реализации (один прогон) дисперсия приспособленности сегмента генома не пропорциональна его длине (рис. 4.11).

Результаты, полученные в настоящем исследовании, могут быть потенциально актуальны для вирусов с частой рекомбинацией, таких как ВИЧ, полиомиелит или SARS-CoV-2. Подобно сезонным человеческим коронавирусам или вирусу гриппа, SARS-CoV-2 постоянно приобретает

новые мутации в своем геноме и имеет сотни полиморфных локусов. Эволюция особенно быстра у рецептора белка Spike, 5 замен в год [201, 397-399]. Две основные причины объясняют высокую скорость эволюции этого вируса. Во-первых, у Спайка имеются рецептор-связывающие мотивы, влияющие на передачу, и их эволюция приводит к появлению вариантов с повышенной трансмиссивностью. Во-вторых, Spike содержит эпитопы — регионы, которые важны для иммунного ответа из-за их участия в связывании антител, способных нейтрализовать вирус. Мутации в эпитопах являются основным фактором, ограничивающим распознавание вируса иммунной системой и, следовательно, долговечность иммунной защиты [400-402].

Загадкой, важной для разработки будущих стратегий вакцинации, является происхождение вызывающих беспокойство вариантов (VOC), вызываемых большими группами



новых мутаций, которые возникают одновременно [403-407]. Популярные гипотезы появления VOC [408] включают обратный зооноз, эволюцию у пациентов с ослабленным иммунитетом [409, 410] и эволюцию в популяционных очагах, не охваченных генетическим надзором. Еще одной возможностью является эффект долины приспособленности, каскадное появление компенсирующих мутаций после первичной мутации, эффект, наблюдаемый или оцененный для разных вирусов, таких как ВИЧ и грипп [6, 28], и изученный теоретически [131].

Рисунок 4.12. Филогенетическое дерево в разные моменты времени имеет полифилетическую структуру с общим стволом. Филогения 10 смоделированных геномов, отобранных в три момента времени, $t = 20, 40, 60$, получена в программном обеспечении MEGA11. Жирной красной линией показан ствол, соединяющий поддеревья. Светлые кружки показывают корни трех поддеревьев, возникающих на границе распределения генома с высокой приспособленностью (показано). Количество полезных аллелей в геноме показано на листьях. Ниже представлена шкала генетической дистанции, измеряемая в количестве различий. Параметры модели такие же, как на рисунке 4.5. По работе [32].

На основании настоящего исследования можно добавить еще одно возможное объяснение. SARS-CoV-2 имеет наблюдаемую перекрестную рекомбинацию [411-413]. Следовательно, большие пакеты мутаций могут возникать благодаря комбинированным эффектам рекомбинации и естественного отбора и представляют собой последовательности, содержащие самые быстрые локусы (рисунок 4.7). Данная модель приводит к оценке

филогенетического дерева с помощью пакета MEGA11 из объединения случайных выборок из 10 смоделированных последовательностей, полученных в 3 момента времени (рисунок 4.12). Образцы генома, полученные в разные моменты времени, образуют отдельные поддеревья (клады), растущие на одном стебле. Каждое дерево имеет превышение количества полезных аллелей по сравнению с предыдущим деревом. Каждое поддерево в один момент времени также состоит из независимых поддеревьев. Эти особенности имеют сходство с наблюдаемой филогенией VOC.

Чтобы понять важность рекомбинации для SARS-CoV-2, нам необходимо знать частоту коинфицированных лиц среди всех инфицированных, что определяет вероятность рекомбинации r , важный входной параметр, входящий в модели половых популяций [10, 15, 74]. Значение r для SARS-CoV-2 в настоящее время неизвестно, так как на него влияет неоднородность передачи вируса включая события суперраспространения [414, 415]. Методы, разработанные ранее для количественной оценки рекомбинации на основе данных о последовательностях РНК ВИЧ, могут быть повторно применены к SARS-CoV-2 [15, 77]. Потенциальная значимость настоящей модели для SARS-CoV-2, а также подробный анализ SARS-CoV-2 в свете этой гипотезы подлежит дальнейшим исследованиям.

Предположения и ограничения. (i) Модель в данном разделе учитывает только очищающий (направленный) отбор. Хотя реальная популяция имеет множество консервативных сайтов, а также некоторые селективно нейтральные сайты, существование этих сайтов не влияет на эволюционную динамику сайтов, находящихся под очищающим отбором. Следовательно, эти сайты не включены в рассмотрение. (ii) Исследован случай постоянного коэффициента отбора, но ожидается, что результаты применимы также и для достаточно быстро убывающего распределения коэффициентов отбора, такого как экспоненциальное распределение. Распределения с более медленно затухающими хвостами могут иметь разные свойства, где бегущая волна заменяется парной клональной интерференцией [16]. Особым случаем является экспоненциальное распределение, часто наблюдаемое в экспериментах на патогенах, и этот факт получил объяснение [29]. (iii) Модель предполагает гомологичную рекомбинацию по следующим причинам. Во-первых, подавляющее большинство событий рекомбинации у вирусов и организмов происходит между гомологичными матрицами и не включает в себя вставки или делеции. Во-вторых, когда происходит негомологичная рекомбинация, эффект которой чаще встречается у вирусов, полученное потомство часто оказывается дефектным и, следовательно, не имеет значения для эволюционной динамики популяции.

4.2.8 Conclusion

При достаточно длинном геноме и в экспоненциально широком диапазоне размеров популяции, рекомбинация не может подавить эффекты дальнего сцепления из-за общей филогенеза разных локусов. Они включают чрезмерную потерю полезных аллелей из-за клональной интерференции, независимость наблюдаемых показателей от среднего числа кроссинговеров и сверхбыструю эволюцию в некоторых локусах из-за эффекта генетического фона. Эти эффекты сопровождаются сильными межгеномными корреляциями из-за родства геномов.

Глава 5. Определение адаптивного ландшафта из ГЕНОМНЫХ ДАННЫХ

Адаптивный ландшафт - это зависимость репродуктивного успеха (приспособленности), определяемого как среднее число потомков способных оставить потомков, от геномной последовательности. Он включает как изменения приспособленности, вызванные отдельными мутациями (коэффициенты отбора), так и их взаимодействие («эпистаз»). Адаптивный ландшафт можно рассматривать как частный случай отображения генотипа на фенотип, где приспособленность является интересующим нас признаком фенотипа. В то время как вирусы или бактерии позволяют, в принципе, измерить адаптивный ландшафт с помощью локус-направленного мутагенеза, в большинстве случаев, приспособленность трудно измерить напрямую. Цель этой главы — описать новые методы оценки ландшафта приспособленности из большого набора геномных последовательностей.

5.1 Универсальный эволюционный след эпистаза

Согласно известной теореме популяционной генетики, если вклады отдельных мутаций в логарифм приспособленности складываются, что имеет место в отсутствие биохимического взаимодействия, разные локусы, в бесконечной популяции, эволюционируют независимо. В реальном организме или вирусе, многие локусы генома, белков, мРНК биохимически взаимодействуют с другими локусами. Взаимодействия между локусами на уровне регуляции генов, транскрипции, мРНК или белков создают неаддитивные эффекты и запутывание эволюции локусов — даже в бесконечно большой популяции, где эффектов филогенетического сцепления нет. Это взаимодействие, эпистаз, является вездесущим свойством биологических систем.

Несмотря на то, что десятки статистических методов разработаны для определения взаимодействия локусов, обнаружение эпистатических взаимодействий в геноме остается серьезной проблемой. Одной из причин является небольшое количество эпистатических пар локусов среди всех возможных пар локусов. Другой причиной является огромный статистический шум, чья природа обсуждается в § 5.2. В дополнение к этим проблемам, до недавнего времени, метод оценки силы эпистаза без запутывания других параметров популяции отсутствовал.

В работе [24] обсуждаемой ниже, разработан и протестирован метод, который позволяет определить относительную величину эпистаза из геномных данных, напрямую, т.е., без

запутывания других параметров. Главный результат — это существование квазиравновесия между эпистазом и беспорядком, определяющего распределение аллелей по геномам и локусам в режиме медленной адаптации, характерном для режима бегущей волны [3]. Для пары взаимно компенсирующих локусов, это квазиравновесие приводит к универсальному соотношению между частотами гаплотипов пары локусов, названному «универсальной след эпистаза», которое зависит только от степени взаимной компенсации двух аллелей, но не других параметров модели или переменных состояния. Это соотношение легко обобщается для любой топологии эпистатической сети, несколько примеров которой рассмотрены в явном виде. (Этот тип квазиравновесия, которое существует в небольших популяциях даже при отсутствии рекомбинации, не следует путать с «квазиравновесием генетического сцепления» QLE [148, 150] которое предказано в очень больших популяциях при наличии не рассматриваемой здесь рекомбинации.)

5.1.1 Модель стохастической эволюции с эпистазом

Модель представляет собой процесс Райта Фишера [15], описанный в § 2.3.1. Эволюционные факторы — естественный отбор, включая эпистаз, мутации, случайный генетический дрейф и генетическое сцепление. Популяция состоит из N последовательностей $[K_i]$, где каждый эволюционирующий локус с номером $i = 1, 2, \dots, L$ является положением нуклеотида или аминокислоты, и $K_i = 0$ или $K_i = 1$ обозначает бинарную аллель. Рекомбинация отсутствует. Среднее число потомков (приспособленность) каждого генома e^W определяется выражением

$$W = \sum_{i=1}^L s_i K_i + \sum_{i<j}^L s_{ij} K_i K_j \quad (5.1)$$

$$s_{ij} \equiv E_{ij} (|s_i| + |s_j|) T_{ij} \quad (5.2)$$

Первый член в правой части уравнения (5.1) соответствует невзаимодействующим мутациям с коэффициентом отбора s_i для каждого локуса i . Второй член в формуле (5.1) описывает парные взаимодействия локусов с величинами s_{ij} вида (5.2); взаимодействия более высокого порядка пренебрегаются. Альтернативная модель глобального эпистаза «среднего поля», в которой эпистатические эффекты усредняются по большому количеству пар, исследована в [416, 417].

Фактор E_{ij} представляет собой относительную величину взаимодействия между локусами i и j , в то время как матрица топологии T_{ij} указывает взаимодействующую пару посредством $T_{ij} = 1$ и невзаимодействующую пару посредством $T_{ij} = 0$. Для простейшего случая пары оснований с невыгодными аллелями $s_i < 0, s_j < 0$, не взаимодействующей ни с каким другим локусом, значение $E_{ij} = 1$ соответствует полной взаимной компенсации, $W = 0$. По

определению, знак E_{ij} такой же, как знак члена взаимодействия в приспособленности s_{ij} , независимо от знаков коэффициентов отбора s_i, s_j . Если член взаимодействия увеличивает приспособленность, $E > 0$ (положительный эпистаз), а если он снижает приспособленность, $E < 0$ (отрицательный эпистаз).

5.1.2 След эпистаза для взаимодействующей пары локусов в длинном геноме

Пусть взаимодействующая пара оснований с вредными аллелями $-s_1, -s_2$ и величиной эпистаза E является частью длинного генома с логарифмом приспособленности W (рисунок 5.1). Эпистатическое взаимодействие пары с другими локусами генома отсутствует. Более сложные эпистатические сети рассмотрены в § 5.1.5 ниже. Предполагается, что популяция находится в процессе многолокусной адаптации. В этом случае, как демонстрируют различные многолокусные модели, плотность распределения приспособленности W по отдельным геномам представляет собой уединённую бегущую волну, с шириной по W гораздо меньше среднего [3] (главы 2 и 3). Анализ [24] основывался на том факте, что волна относительно узкая и относительно медленная, поскольку скорость движения ограничена редкими мутационными событиями, добавляющими новые геномы на фронт высокой приспособленности [7].

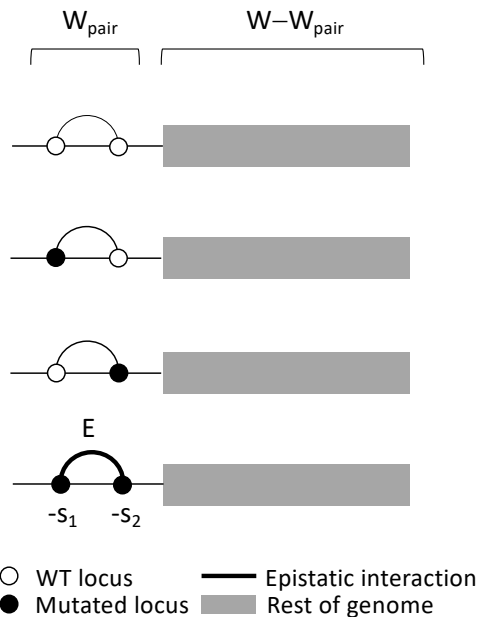


Рисунок 5.1. Пара взаимодействующих локусов в длинном геноме. Белые и черные кружки: дикий тип 0 и мутантный аллель 1. Толстая дуга: существующее взаимодействие. Тонкая дуга: потенциальное взаимодействие. Серый прямоугольник: остальная часть генома. Относительная сила эпистаза E определена в (5.2). По работе [24].

Цель — предсказать аллельную ассоциацию при такой медленной адаптации. Для этого, все индивидуальные геномы в популяции классифицируются в соответствии с гаплотипом данной пары локусов, 00, 01, 10 и 11 (рисунок 5.1). Каждый класс полиморфен во многих других локусах.

Вклад данной пары W_{ij} в логарифм приспособленности генома W зависит от гаплотипа ij следующим образом:

$$\begin{aligned} W_{00} &= 0 \\ W_{01} &= -s_1 \\ W_{10} &= -s_2 \\ W_{11} &= -(s_1 + s_2)(1 - E) \end{aligned} \quad (5.3)$$

Как обсуждалось много раз, эволюция популяции происходит в присутствии стохастических факторов, мутация и генетический дрейф, которые увеличивают беспорядок, и детерминистического фактора естественного отбора. Стандартной мерой беспорядка является энтропия S , которая определяется как логарифм числа возможных конфигураций, $S = \log N_{\text{conf}}$. Противодействующей силой является естественный отбор, который уменьшает беспорядок и увеличивает приспособленность.

Как уже упоминалось в главах 2 и 3, распределение приспособленности W в многолокусном режиме представляет собой узкий пик при $W = W_{\text{max}}(t)$, который медленно движется во времени [7, 11, 12, 72]. Гипотеза квазиравновесия состоит в том, что энтропия, вычисляемая по ансамблю возможных последовательностей, в каждый момент, успевает приблизиться к текущему максимуму, при том ограничении, что приспособленность всех геномов задана и равна $W_{\text{max}}(t)$. Эта гипотеза подтверждается в широком диапазоне параметров с помощью компьютерного моделирования [24] (ниже). При этом предположении, максимальное значение S зависит только от $W(t)$, а не от t в явном виде, $S(t) = S[W(t)]$. Примеры функции $S(W)$ для разных типов эпистатических сетей представлены в табл. 5.1.

Таблица 5.1. Выражения для приспособленности, энтропии и частот гаплотипов для различных эпистатических сетей (рисунок 5.7). В выражениях для энтропии предполагается сильное неравенство $k_i \ll L$.

Топология сети	Приспособленность $-W/s_0$	$\exp(S)$	Доля гаплотипов f_{10}, f_{11}
Изолированные пары	$k_1 + 2k_2(1 - E)$	$C_{L/2}^{k_2} C_L^{k_1}$	$f_{10} = k_1/L$ $f_{11} = 2k_2/L$
Двойные арки	$k_1 + 2k_2(1 - E) + 2k_3(3 - 4E)$	$C_{L/3}^{k_3} C_{L/3}^{k_1} C_{L/3}^{k_2} 2^{k_2} 3^{k_1}$	$f_{10} = f - f_{11}$ $f_{11} = 3(k_2 + k_3)/(2L)$ $f = (k_1 + 2k_2 + 3k_3)/L$
Тройные арки	$k_1 + 2k_2(1 - E) + 3k_3(1 - 2E)$	$C_{L/3}^{k_3} C_{L/3}^{k_1} C_{L/3}^{k_2} 3^{k_2} 3^{k_1}$	$f_{10} = (k_1 + k_2)/L$ $f_{11} = (k_2 + 3k_3)/L$
Цепочка	$k_1 + \sum_{i=2}^l (i - 2(i - 1)E)k_i$	$\prod_{i=1}^l C_L^{k_i}$	$f_{10} = \sum_{i=1}^l k_i/L$
Бинарное дерево		$\prod_{i=1}^l C_{L_i}^{k_i} (A_i)^{k_i}$	$f_{11} = \sum_{i=2}^l (i - 1)k_i/L$
Неравные двойные арки	$k_1 + 2k_2(1 - E) + 3k_3(1 - E) + k_2'(2 - E)$	$C_{L/3}^{k_3} C_{L/3}^{k_1} C_{L/3}^{k_2} C_{L/3}^{k_2'} 3^{k_1}$	$f_{10} = f - f_{11}$ $f_{11} = 3(k_2 + k_2' + k_3)/(2L)$ $f = (k_1 + 2k_2 + 2k_2' + 3k_3)/L$

Рассматриваются все последовательности в классе гаплотипов, например, 10 (рисунок 5.1). Другие локусы за пределами этой пары могут быть полиморфны. Вероятность гаплотипа f_{10} , по определению вероятности, представляет собой отношение количества конфигураций в классе 10, обозначенного $\exp(S_{10})$, к общему количеству возможных конфигураций. Согласно приведенной выше гипотезе, энтропия каждого класса гаплотипов является функцией приспособленности остального генома, $S_{10} = S(W - W_{10})$. Средняя частота, с которой каждый гаплотип встречается в популяции, пропорциональна числу его конфигураций, $f_{10} \propto \exp(S_{10}) = \exp[S(W - W_{10})]$. Поскольку предполагается, что геном длинный, $L \gg 1$, мы знаем что W_{10} намного меньше, чем W . Следовательно, можно аппроксимировать энтропию линейным разложением Тэйлора

$$S(W - W_{10}) \approx S(W) - \beta W_{10} \quad (5.4)$$

где введено обозначение $\beta \equiv \frac{dS}{dW}$. Объединяя (5.3) и (5.4), частоты 3х гаплотипов принимают вид

$$\begin{aligned} f_{10} &= f_{00} e^{-\beta s_1} \\ f_{01} &= f_{00} e^{-\beta s_2} \\ f_{11} &= f_{00} e^{-\beta(s_1+s_2)(1-E)} \end{aligned} \quad (5.5)$$

Здесь f_{00} находится из условия нормировки $f_{00} + f_{01} + f_{10} + f_{11} = 1$. Исключая β, s_1, s_2 из (5.5), получается

$$\frac{f_{11}}{f_{00}} = \left(\frac{f_{10} f_{01}}{f_{00}^2} \right)^{1-E}$$

откуда

$$E = 1 - \frac{\log(f_{11}/f_{00})}{\log(f_{01}f_{10}/f_{00}^2)} \quad (5.6)$$

Формула (5.6) может быть использована для оценки силы взаимодействия E из набора последовательностей ДНК, РНК или белка. В отличие от существующих корреляционных методов, эта мера эпистаза не только обнаруживает сам факт аллельной корреляции, но и дает значение E без запутывания других параметров. Она также имеет изрядную степень универсальности, как показано ниже и поэтому называется «Универсальный След Эпистаза» (UFE).

Результат (5.6) и гипотеза квазиравновесия были проверены с помощью моделирования Монте-Карло, описанного в § 5.1.1, для длинного генома, состоящего из изолированных эпистатических пар с величиной E . Оценка (5.6) сравнивалась с фактическим значением E , установленным при моделировании (рисунок 5.2). Результат показывает, что (5.6) представляет собой достаточно точную оценку E после $\sim 1/s$ поколений (рисунок 5.2), что одновременно является временем установления бегущей волны (§ 5.4 ниже). Аналогичные результаты

получаются в широком диапазоне параметров, включая гораздо большие N (рисунок 5.2). Однако на больших временах, когда вредная мутация становится важной и почти уравнивает естественный отбор, возникают сильные отклонения UFE от E (рисунок 5.2). Таким образом, оценка (5.6) не работает вблизи истинного равновесия (стационарного состояния), она применима только в процессе медленной адаптации.

5.1.3 Длинный геном изолированных пар

Чтобы рассмотреть явный пример выражения для энтропии S , как функции приспособленности, и еще раз проверить справедливость оценки (5.6), поучительно рассмотреть в явном виде простой пример целого генома. Пусть геном состоит из изолированных эпистатических пар. Для простоты предполагается, что все коэффициенты отбора и эпистатические силы одинаковы для всех оснований и пар, $s_i = -s_0, E_{ij} = E_0$. Аналитический метод описан ниже § 5.1.5, а подробный вывод в работе [24]. Алгоритм вывода следующий.

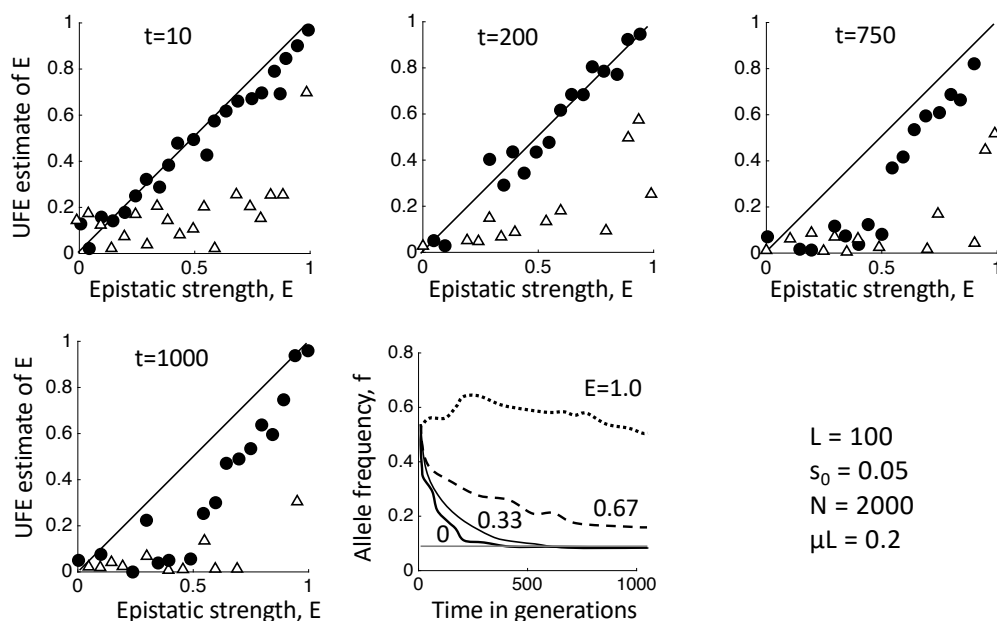


Рисунок 5.2. Оценка E по формуле (5.6) нарисована, как функция истинного значения E . Показаны времена в поколениях. Внизу справа: временное изменение мутантной частоты f . Каждая точка — один прогон Монте-Карло. Исходная популяция содержит случайно расположенные аллели с $f = 0.5$. Частоты гаплотипов усреднены по парам локусов. Черные кружки: заданные в программе эпистатические пары. Белые треугольники: такое же количество случайно выбранных пар. Параметры: $L = 100, s_0 = 0.05, N = 2000, \mu L = 0.2$, одно взаимодействие на локус. По работе [24].

Связные кластеры невыгодных аллелей группируются по размеру и подсчитываются: k_1 одиночек и k_2 пар (рисунок 5.3). Приспособленность W и энтропия S и средние частоты гаплотипов 10 и 00 выражаются через k_1 и k_2 (§ 5.1.5). Чтобы получить наиболее вероятное состояние, k_1 и k_2 выбираются из условия, что энтропия $S(k_1, k_2)$ максимальна при ограничении,

что приспособленность W дана, откуда получается соотношение между долями гаплотипов, идентичное (5.6). Наконец, можно использовать условие $f_{10} + f_{11} = f$, чтобы выразить среднюю частоту вредных аллелей f через E и $f_0 \equiv \frac{|W|}{sL}$.

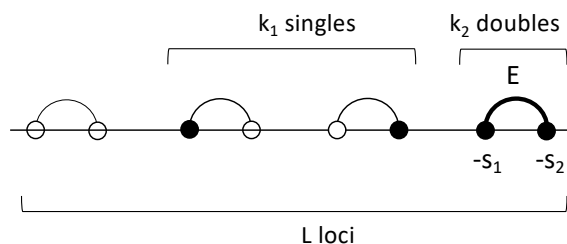


Рисунок 5.3. Длинный геном взаимодействующих пар. Взаимодействующие пары с разными гаплотипами. По работе 23.

В точке половинной компенсации $E = 0.5$, пары аллелей и одиночки имеют одинаковую приспособленность и, в результате, одинаковую частоту, $f_{11} = f_{10}$. Вне этой точки, одна группа гораздо многочисленнее другой над другой. При слабом эпистазе $E < 0.5$, одиночки многочисленнее двоек, а при сильном эпистазе $E > 0.5$, двойки доминируют над одиночками.

Таблица 5.2. Коэффициенты корреляции и критические точки для различных примеров топологии эпистатической сети (рисунок 5.5).

Топология	Количество связей	E_c	E_{UFE}	Интервал E	D_{11}	D_{10}	$f(E)$
Арка	$b_2 = 1$	1	1	$E < \frac{1}{2}$	f^{-2E}	1	f_0
				$\frac{1}{2} < E < 1$	f^{-1}	$f^{\frac{E-1/2}{1-E}}$	$\frac{f_0}{1-E}$
Тройная арка	$b_3 = 3$	$\frac{1}{2}$	$\frac{1}{4}$	$E < \frac{1}{4}$	f^{-2E}	1	f_0
				$\frac{1}{4} < E < \frac{1}{3}$	f^{1-6E}	1	f_0
				$\frac{1}{3} < E < \frac{1}{2}$	f^{-1}	$f^{\frac{E-1/3}{1/2-E}}$	$\frac{f_0}{1-2E}$
Двойная арка	$b_3 = 2$	$\frac{3}{4}$	$\frac{1}{2}$	$E < \frac{1}{2}$	f^{-2E}	1	f_0
				$\frac{1}{2} < E < \frac{3}{4}$	f^{-1}	$\frac{1}{3} f^{\frac{2E-1}{3-4E}}$	$\frac{f_0}{1-4E/3}$
Цепочка	$b_i = i - 1$	$\frac{1}{2}$	$\frac{1}{2} - \frac{1}{\log \frac{1}{f_0}}$	$\frac{1}{2} - E \gg \frac{1}{\log \frac{1}{f_0}}$	f^{-2E}	1	f_0
				$\frac{1}{2} - E \ll \frac{1}{\log \frac{1}{f_0}}$	f^{-1}	$\frac{f_0}{f}$	$\frac{f_0}{(1-2E) \log \frac{1}{f_0(1-2E)}}$
Двоичное дерево	$b_i = i - 1$	$\frac{1}{2}$	$\frac{1}{2} - \frac{\log 4}{\log \frac{1}{f_0}}$	$E_{UFE} - E \gg \frac{1}{\log \frac{1}{f_0}}$	f^{-2E}	1	f_0
				$E_{UFE} - E \ll \frac{1}{\log \frac{1}{f_0}}$	f^{-1}		

Здесь b_i — количество связей для кластера из i аллелей, E_c — точка полной компенсации, E_{UFE} — максимальная эпистатическая величина, при которой оценка UFE все еще точна, D_{11} и D_{10} — коэффициенты парной корреляции, f — частота невыгодных аллелей, $f_0 \equiv -\frac{W}{sL}$.

Параллельно с UFE, полезно вычислить стандартные корреляторы

$$D_{11} = f_{11}/f^2, \quad D_{10} = f_{10}/[f(1-f)].$$

Если частоты аллелей статистически независимы, $D_{11} = D_{10} = 1$. Их значения выше единицы показывают положительную корреляцию, а значения ниже единицы отрицательную корреляцию. Корреляторы можно выразить через среднюю частоту мутантов f , а ее — через заданные параметры E и $f_0 = -W/(sL)$ (табл. 5.2).

Аналитические результаты показаны при $f_0 = 1/10$ и $1/100$ на рисунках 5.4 и 5.6, соответственно. Как и ожидалось, частота аллелей f , равная f_0 при отсутствии эпистаза, расходится вблизи точки полной компенсации $E = 1$, пока не достигает значения 0.5 (рисунки 5.4C и 5.6C, таблица 5.2). Коэффициент корреляции D_{11} достигает максимума при $E = 1/2$, и уменьшается до значения 2 в точке компенсации $E = 1$ (рисунок 5.4A), где $f_{00} = f_{11} = 1/2$. Коэффициент корреляции D_{10} снижается ниже 1 при $E > 0.5$ (рисунок 5.4B). Моделирование подтверждает аналитические результаты (рисунок 5.4.). Самое главное, что оценка $UFE \approx E$ точна во всем интервале E (рисунки 5.2 и 5.6D).

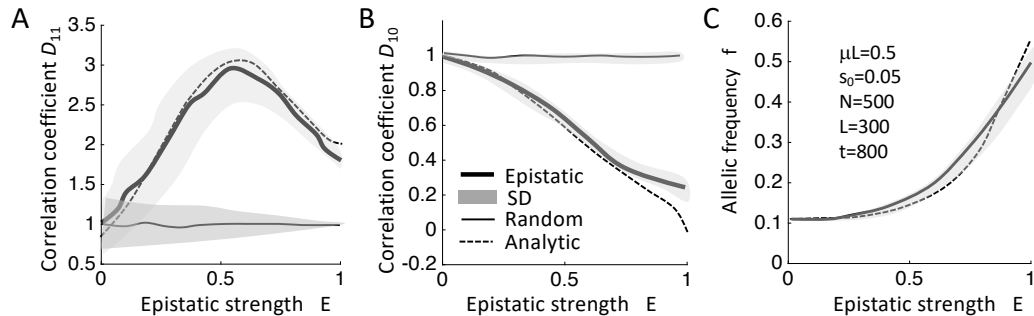


Рисунок 5.4: Эпистаз вызывает сильные корреляции (анализ и численное моделирование). (А, В) Коэффициенты корреляции D_{11}, D_{10} и (С) частота мутаций f показаны как функция E . Корреляторы D_{11}, D_{10} рассчитываются по формуле (5.6) с использованием смоделированных значений f_{ij} и f , усредненных по парам локусов. Сплошные кривые соответствуют среднему значению по 300 прогонам, а затененные области показывают стандартное отклонение среди прогонов для эпистатических пар (толстые кривые) и такого же количества случайных пар (тонкие кривые). Пунктирная кривая — аналитический результаты. Параметры: $N = 500, s_0 = 0.05, L = 300, \mu L = 0.5, t = 800, f_0 \approx 0.1$. Исходная популяция рандомизирована с $f = 0.5$. По работе [24].

5.1.4 Полная компенсация и интервал UFE

Топология реальной эпистатической сети может быть более сложной, чем набор изолированных пар (рисунок 5.5B-E). Конкретные примеры обсуждаются в § 5.1.5 ниже, но сначала полезно

вывести две важные точки в E для любой топологии, используя общее выражение для приспособленности (5.1). Точка полной компенсации определяется из формулы [24]

$$E_c = \min_i [i/(2b_i)] \quad (5.7)$$

где b_i определяется как (среднее) количество связей в кластере вредных аллелей размера i . Точка $E = E_c$ представляет собой порог, где теряется генетическая стабильность всей популяции (табл. 5.2). Порог впервые достигается для критического размера кластера i , для которого отношение числа аллелей на число связей b_i самое маленькое из всех размеров кластеров. При приближении к порогу, число критических кластеров быстро возрастает, пока частота аллелей не перестанет быть малой. Если мутации имеют важный фенотип, такой как устойчивость вируса (опухоли) к лечению или иммунному ответу, этот порог и есть тот момент, когда вирус приобрел устойчивость к лекарству (опухоль начала расти). Эволюционный уход вирусов от иммунного ответа в отсутствие эпистаза рассматривается в §§ 7.1 и 7.2.

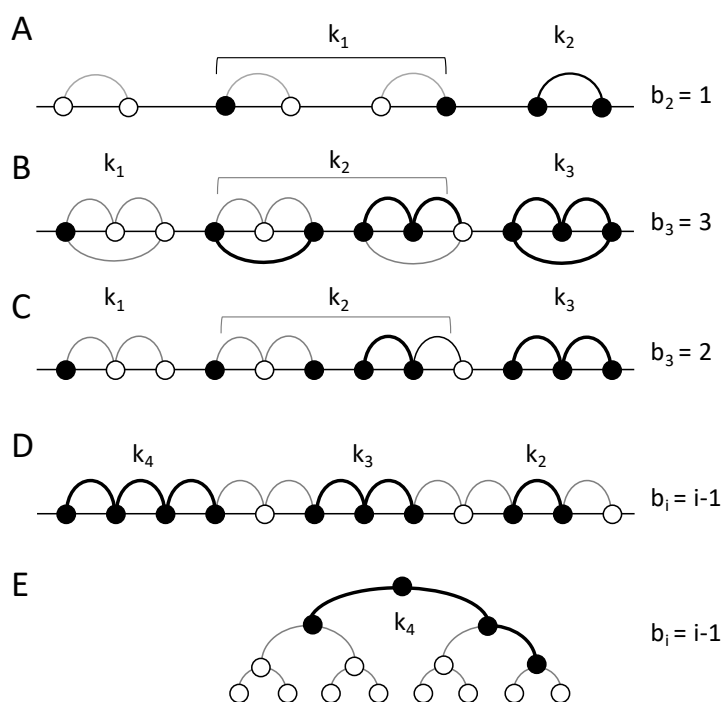


Рисунок 5.5. Примеры эпистатической сети. Черные и белые кружки обозначают вредные и полезные аллели, соответственно. Существующие и потенциальные эпистатические связи между локусами показаны толстыми и тонкими дугами соответственно. Толстые дуги соединяют кластеры вредных аллелей. k_i — число кластеров с i локусами, b_i — число связей на кластер. Разные топологии: А) изолированные пары, В) изолированные тройные дуги, где каждый основание имеет двух эпистатических партнеров, С) двойные дуги, где три локуса участвуют в двух связях, D) длинной цепи, где каждый локус имеет две связи. (Е) Двоичное дерево. (А-Е) показывают топологию связей, но не расположение локусов вдоль генома. По работе [24].

Для изолированных пар, уравнение (5.7) дает $E_c = 1$, как и должно быть. В общем случае, точка компенсации E_c может быть меньше 1, поскольку каждая мутация может быть компенсирована более чем одной мутацией, в зависимости от количества связей b_i .

Вторым важным результатом является интервал $E < E_{UFE}$, в котором Универсальный След (5.6) является точной оценкой эпистатической величины, где

$$E_{UFE} = \min_{i>2} \left[\frac{i-2}{2(b_i-1)} \right] \leq 0.5 \quad (5.8)$$

$E < E_{UFE}$ представляет собой условие, при котором троек вредных аллелей и еще больших кластеров очень мало по сравнению с одиночками и двойками.

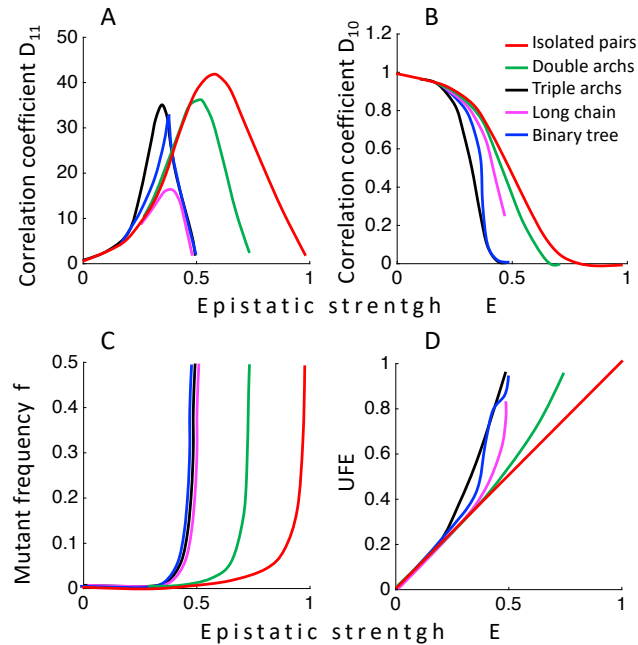


Рисунок 5.6: Универсальный след UFE сохраняет точность для различных топологий при умеренной величине эпистаза, и требует топологически-зависимой поправки при большей. Зависимости от E показаны для пяти топологий на рисунке 5.5. (A, B) Коэффициенты корреляции D_{11}, D_{01} . (C) Доля вредных аллелей f . (D) UFE по формуле (5.6). $f_0 = 1/100$. Таким образом, UFE является точным для изолированных пар и завышает E при больших E для других топологий при $E > E_{UFE}$, формула (5.8). Система неустойчива за пределами точки полной компенсации (5.7). Асимптотические выражения приведены в табл. 5.2. По работе [24].

5.1.5 Влияние топологии сети

Были рассмотрены несколько примеров более сложной топологии сети (рисунок 5.5b-e). При больших эпистатических силах, кластеры из более чем двух вредных аллелей становятся более многочисленными, чем одинокие аллели. Отправной точкой является выражение энтропии, приспособленности и частот гаплотипов в терминах числа кластеров разного размера. Эти выражения и окончательные результаты приведены в таблицах 5.1 и 5.2 соответственно. Общий метод и окончательные результаты для UFE и коэффициентов корреляции для различных топологий обсуждаются ниже. Подробные выводы приведены в оригинальной работе [24].

Общий метод для однородных s и E . Рассмотрим популяцию с долей вредных аллелей $f \ll 1$, и предположим, что все коэффициенты отбора и эпистатическая величины везде одинаковы, $s_i = -s_0 < 0$, $E_{ij} = E > 0$. В этом случае, геном можно полностью охарактеризовать набором чисел кластеров аллелей k_i разного размера i с b_i связями (рисунок 5.5). Это рассмотрение ограничено топологиями, в которых b_i принимает одно значение для каждого размера кластера i (рисунок 5.7В). Приспособленность (5.1) можно представить в виде суммы по кластерам разного размера

$$W \equiv -s_0 f_0 L = -s_0 \sum_{i=1}^{i_{max}} k_i (i - 2Eb_i) \quad (5.9)$$

Здесь f_0 имеет смысл доли вредных аллелей при отсутствии эпистаза и с той же приспособленностью W . Число связей b_i для размера кластера $i > 2$ зависит от конкретной топологии (рисунок 5.5); для любой топологии $b_1 = 0, b_2 = 1$.

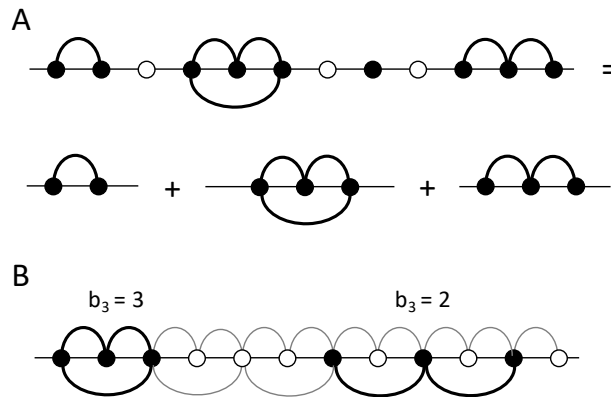


Рисунок 5.7. Варианты топологии эпистатической сети. (А) Неоднородную топологию, состоящую из смеси подграфов с рисунка 5.5, эквивалентна композиции однородных топологий (то есть одиночных пар, двойных и тройных арок), исключая неэпистатические локусы. (В) Пример топологии, когда два субкластера одного размера могут иметь разное количество связей (в данной работе не рассматривается). По работе [24].

Энтропия в квазиравновесии

Как упоминалось выше, значения k_i определяются условием, что энтропия S максимальна с ограничением, что приспособленность W дана, формула (5.9). По определению, энтропия есть логарифм числа конфигураций

$$e^S = \prod_{i=1}^{i_{max}} C_{L_i}^{k_i} (n_i)^{k_i} \quad (5.10)$$

Здесь L_i — количество возможных позиций для кластера размера i , а n_i — количество его возможных конфигураций при данной позиции. Эти значения определяет топология эпистатической сети. Поскольку вредные мутации предполагаются редкими:

$$f = \frac{1}{L} \sum_i ik_i \ll 1 \quad (5.11)$$

перекрытием кластеров можно пренебречь. Последнее условие нарушается только вблизи точки полной компенсации, $E = E_c$, где f расходится.

Из (5.9) и (5.11), имеем $f(E = 0) = f_0$, где f_0 — заданный входной параметр. При положительном E , $f(E) > f_0$ (таблица 5.2 и рисунок 5.6С). Доли гаплотипов могут быть выражены через количество кластеров k_i

$$f_{11} = \frac{1}{L_{\text{pair}}} \sum_i k_i b_i \quad (5.12)$$

$$f_{10} = f_{01} = f - f_{11}$$

где $L_{\text{pair}} = \sum_{ij} T_{ij}$ — общее количество взаимодействующих пар в геноме. Коэффициенты корреляции D_{11} , D_{10} , определенные выше, находятся из (5.12).

Неэпистатические основания и неоднородная сеть

Невзаимодействующие локусы в геноме можно не учитывать (рисунок 5.7А), поскольку приспособленность W и энтропия S аддитивны по эпистатической и неэпистатической части, а частоты f_{11} , f_{10} , f и f_0 определены только для эпистатических оснований. Следовательно, можно разделить геном на две части и максимизировать значения их энтропии по отдельности. По той же причине, разнообразную смесь конечных эпистатических графов можно разбить на независимые части с однородными сегментами (рисунок 5.7А).

Топология сети: тройные дуги

Рассмотрим теперь конкретную топологию более интересную чем отдельные пары: периодическую последовательность трехузловых графов, соединенных тремя связями (рисунок 5.5В). Энтропия и приспособленность зависят от количества одиночных мутаций, k_1 , двойных мутаций, k_2 , и тройных мутаций, k_3 (таблица 5.1). Согласно общему методу, энтропия (5.10) максимизируется по этим трем переменным, при сохранении величины (5.9).

Аналитические результаты, полученные в [24], таковы. Точка полной компенсации $E_c = 0.5$ (рисунок 5.6А, В; таблица 5.2). Зависимость коэффициентов корреляции D_{10} , D_{11} от E различается в трех следующих интервалах эпистатической силы E . При слабом эпистазе $E < 1/4$, связанных троек аллелей очень мало по сравнению со связанными двойками, $k_3 \ll k_2$. В этом интервале, UFE вычисленный по формуле (5.6) является точным (рисунок 5.6D). При умеренном эпистазе $1/4 < E < 1/3$, тройки побеждают двойки, $k_3 \gg k_2$, так что взаимодействующие пары мутантных локусов обнаруживаются преимущественно в тройках. В результате, UFE предсказывает большее значение E , чем фактическое значение (рисунок 5.6D),

что приводит к более крутому увеличению коррелятора D_{11} с E , чем для топологии изолированных пар. При сильном эпистазе $1/3 < E < 1/2$, тройки аллелей превосходят по количеству как двойки, так и одиночки. При приближении к точке полной компенсации $E = 1/2$, накопление троек вызывает расхождение средних доли аллелей f и снижение D_{11} (рисунок 5.6а, в). Таким образом, наличие дополнительных связей, по сравнению с топологией изолированных пар, создает положительную поправку к UFE выше определенного значения $E = 1/3$, что необходимо учитывать при измерении E из геномных данных.

Все остальные топологии (рисунок 5.5) анализируются аналогично, через определение порядка доминирования кластеров различного размера в разных интервалах значений E . Результаты для всех топологий даны в таблице 5.2 и на рисунке 5.6, а для примера неоднородной сети, на рисунке 5.8. Детали выводов содержатся в исходной работе [24].

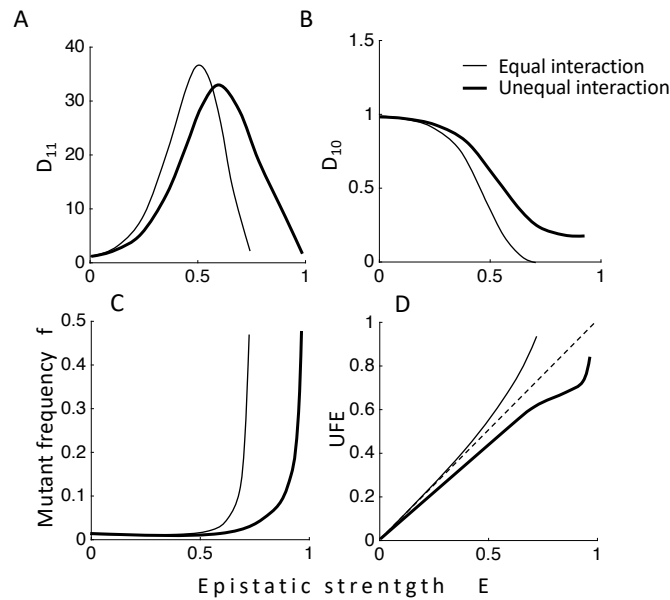


Рисунок 5.8. Результаты для сети из двойных дуг (рисунок 5.5с) с неравной эпистатической силой. Ось Y: коэффициенты корреляции, средняя частота и UFE. Ось абсцисс: эпистатическая сила E . Два случая представляют собой одинаковую эпистатическую силу (тонкие кривые) и неравные силы E и $E/2$ (толстые кривые); $f_0 = 1/100$ — входной параметр. По работе [24].

5.1.6 Заключение

С помощью анализа и моделирования Монте-Карло бесполой популяции, получена оценка эпистаза (UFE) через частоты гаплотипов пар локусов f_{11}, f_{10}, f_{00} найденных из геномных данных, независимая от других параметров системы. При умеренной величине эпистаза, значение UFE не зависит от топологии и равно E , если имеется достаточное усреднение по ансамблю или одинаковым парам взаимодействующих локусов. Для простейшей топологии, когда эпистатические пары изолированы друг от друга, можно показать что $UFE = E$ во всем

интервале E , вплоть до точки полной компенсации, где система становится неустойчивой. Для более сложных сетей взаимодействующих локусов, модель предсказывает значение E , выше которой UFE имеет поправки к $UFE = E$, зависящие от топологии сети и вычисленные в этой работе. Точка полной компенсации, т.е., граница генетической устойчивости по E , получена как функция топологических параметров однородной сети, а именно, среднего числа взаимодействий на взаимодействующий локус. Это предсказание теории можно использовать для идентификации порогов генетической стабильности в предраковых клетках или предсказания эволюции устойчивости патогенов к лекарствам [113-115, 117, 132].

5.2 Обнаружение эпистатических пар локусов в одной отдельно взятой популяции: миссия невыполнима

Важным ограничением работы [24] описанной в предыдущем § 5.1 является то, что частоты гаплотипов используемые для вычисления статистической ассоциации аллелей усредняются по эпистатическим парам локусов, которые заранее известны и заданы в программе. В реальных наборах геномных данных, расположение эпистатических пар обычно неизвестно, и их довольно трудно обнаружить [105-107]. Проблема обнаружения эпистатических пар рассматривается в данном и следующем разделе.

Взаимозависимая эволюция разных участков генома обусловлена не только эпистазом. Общее наследование аллелей в близко расположенных локусах или при отсутствии рекомбинации также вызывает аллельные корреляции, даже при отсутствии эпистаза. Аллельную ассоциацию, возникающую в результате этого генетического сцепления, легко спутать с эффектами эпистаза. Этот эффект сцепления является вероятностным из-за случайной выборки геномов (генетический дрейф) и случайных мутаций. Эффекты сцепления становятся незначительными только в гигантских популяциях, экспоненциально увеличивающихся с номером локуса L [7]. В реальных наборах ДНК, часто неясно, как отличить эффект генетического сцепления от эпистатического взаимодействия, и какой из двух эффектов доминирует в данном наборе. Таким образом, обнаружение эпистаза по геномным данным остается сложной задачей [105-107].

Чтобы решить проблему, нужно найти ее причину. Настоящий раздел посвящен работе [26], в которой получено механистическое объяснение хронических неудач с обнаружением эпистаза. С этой целью, искусственные последовательности генерированы путем моделирования методом Монте-Карло. Несколько мер аллельной корреляции, а также их двойные и тройные

комбинации, сопоставлены между эпистатическими и неэпистатическими парами, в широком диапазоне параметров модели. В одной отдельной популяции, эффекты генетического сцепления преобладают над эпистазом, при почти всех значениях параметров и временах. Этот неутешительный эффект объяснен из случайной природы филогенетического дерева.

5.2.1 След эпистатических пар не отличим от генетического сцепления

Модель популяции с бинарными последовательностями такая же, как и в § 5.1. Чтобы охарактеризовать статистическую корреляцию между аллелями в разных локусах, именуемое в популяционной генетике «неравновесным генетическим сцеплением» между парами локусов, LD, ниже используются два показателя.

Пусть $f_{\alpha\beta}^{ij}$ — средняя частота выбранного бинарного гаплотипа $\alpha\beta$ из четырех возможных, $\alpha\beta \in [00,01,11,10]$, для локусов i и j , а f_{α}^i — доля двух возможных аллелей, $\alpha \in [0,1]$, на локусе i . Одним из широко используемых показателей корреляции является коэффициент Пирсона [418]

$$r_{\alpha\beta}^{ij} = \frac{f_{\alpha\beta}^{ij} - f_{\alpha}^i f_{\beta}^j}{\sqrt{f_{\alpha}^i (1 - f_{\alpha}^i) f_{\beta}^j (1 - f_{\beta}^j)}} \quad (5.13)$$

Коэффициент Пирсона меняется между -1 до 1 , демонстрируя либо отрицательную, либо положительную корреляцию. Другой коррелятор — это введенный выше универсальный след эпистаза [24]

$$\text{UFE} = 1 - \frac{\log(f_{11}/f_{00})}{\log(f_{01}f_{10}/f_{00}^2)} \quad (5.14)$$

Как показано в § 5.1, UFE равен степени компенсации E для изолированной взаимодействующей пары вредных аллелей, если частоты четырех гаплотипов усреднены по ансамблю. Значение $E = 0$ означает отсутствие компенсации, $E = 1$ — полную взаимную компенсацию, $E < 0$ — отрицательное взаимодействие. Формула (5.14) имеет сингулярность при $f_{01}f_{10} = f_{00}^2$, что не влияет на результаты, обсуждаемые ниже.

В литературе существует много других показателей LD. В оригинальной работе использовались еще два [26], с очень похожими результатами. В § 2.2, использовалась частота наименее представленного гаплотипа $\min(f_{00}, f_{11}, f_{10}, f_{01})$. Общим для всех показателей является то, что они исчезают, если $f_{\alpha\beta}^{ij} = f_{\alpha}^i f_{\beta}^j$, что подразумевает отсутствие аллельных корреляций между двумя локусами. Кроме того, попарная корреляция — не единственный метод обнаружения ассоциации аллелей. Были разработаны общие методы, основанные на

филогенетическом дереве [419] и на приближении «равновесия квази-сцепления» QLE [148, 150]. Обзор методов дан в работе [107]. Однако, как станет ясно ниже, окончательные выводы не зависят от метода обнаружения ассоциации аллелей, и два показателя LD, определенные в (5.13) и (5.14), достаточно представительны.

Гистограммы трех показателей LD по всем парам участков были рассчитаны в разные моменты времени моделирования методом Монте-Карло. Первые два — это коэффициенты Пирсона для двух гаплотипов r_{11}, r_{01} (5.13), а третья — UFE (5.14). Распределения (гистограммы) этих показателей, как для заданных эпистатических пар, так и для всех возможных пар локусов, показаны на рисунке 5.9. Площадь под каждым распределением — это общее количество таких пар. При $t = 0$, оба распределения имеют резкий пик около нуля из-за случайного начального расположения аллелей (рисунок 5.9, ряд 1).

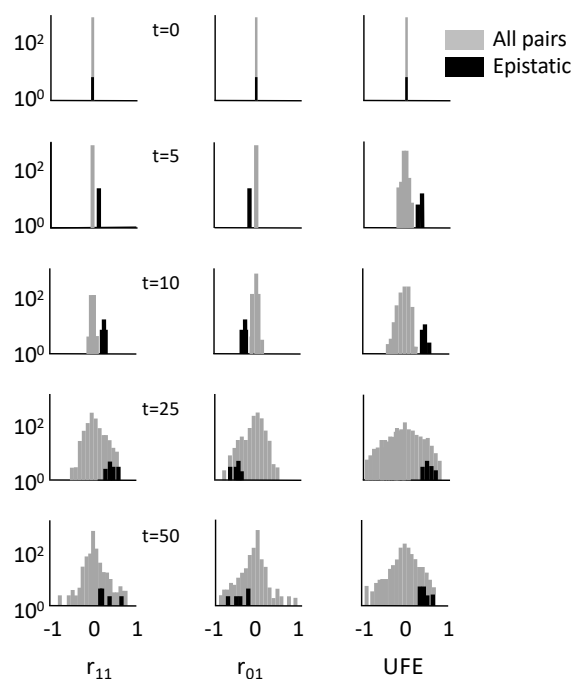


Рисунок 5.9. Показатели корреляции позволяют обнаружить эпистаз только в узком временном окне. В столбцах, показан профиль распределения r_{11}, r_{01} (5.13) и UFE (5.14). Разные строки соответствуют разным моментам времени: $t = 1, t = 5, t = 10, t = 25$ и $t = 50$. Разные оттенки соответствуют гистограммам для всех пар локусов (серый цвет) и для заданных эпистатических пар (черный). Площадь пропорциональна количеству пар каждого подмножества. Параметры: $N = 2 \cdot 10^4, s_0 = 0,1, L = 50, E = [0,1], \mu L = 7 \cdot 10^{-2}$. Топология сети: каждый нечетный локус взаимодействует со своим соседом справа (1-2, 3-4, 5-6,...) с величиной $E = 0.75$. Начальное расположение вредных аллелей случайное, с долей $f = 0.4$. В отрицательном контроле с $E = 0$, темный и светлый цвет всегда перекрываются ([26], рисунок S1). По работе [26].

На ранней стадии эволюции (рисунок 5.9, ряды 2 и 3), происходит разделение двух гистограмм, при этом эпистатическая перемещается за пределы гистограммы для всех пар. Позже, распределение неэпистатических пар расширяется и поглощает распределение эпистатических пар (рисунок 5.9, $t = 25, t = 50$). Неэпистатические пары, в силу стохастического характера мутации и случайного генетического дрейфа, обладают корреляцией случайного знака. Таким

образом, за исключением узкого временного окна, отличить эпистатические пары от других пар в одной популяции невозможно.

5.2.2 Комбинация нескольких корреляторов не улучшает обнаружение эпистаза

Можно было бы надеяться улучшить шансы на обнаружение, комбинируя вместе различные показатели LD. Несколько комбинаций двух и трех мер LD были использованы в попытке обнаружить взаимодействующие пары для $E = 0.75$ в двух временных точках, но без какого-либо улучшения обнаружения. Пример показан на рисунке 5.10, где значения r_{11} и UFE показаны, для одних и тех же пар локусов, по осям ординат и абсцисс, соответственно.

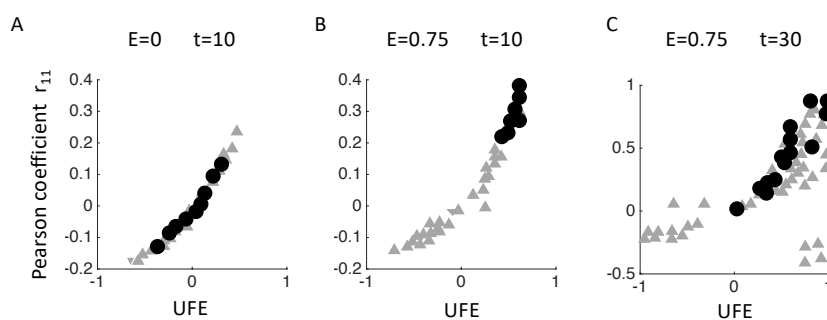


Рисунок 5.10. Комбинация двух разных показателей LD не позволяет улучшить обнаружение эпистаза. (А) Отсутствие эпистаза. Эпистатические пары (черные точки) и случайные пары (серые треугольники) перекрываются. (В) В промежуточный момент времени $t = 10$, эпистатические пары (черные кружки) группируются отдельно и, следовательно, поддаются обнаружению. (С) Позже, существенное перекрытие с не взаимодействующими парами препятствует обнаружению. Для оптимизации обнаружения, используется порог обнаружения для каждой статистики, который минимизирует следующую величину « $DET + a \cdot FPSO$ », где a — параметр подбора, DET — процент обнаружения, а $FPSO$ — процент ложноположительных результатов, основанный на априорном знании эпистатических пар. Параметры как на рисунке 5.9. По работе [26].

Чтобы получить наилучшие, теоретически возможные условия для обнаружения эпистаза, был рассчитан оптимальный порог, отделяющий взаимодействующие пары от не взаимодействующих, на основе априорного знания эпистатических пар (рисунок 5.10). Метод, использующий принцип машинного обучения, не дал существенного улучшения обнаружения ([26], таблица S1). Для реального набора данных, такие априорные знания недоступны, так что обнаружение может быть только хуже. Другие комбинации двух и трех мер LD дают аналогичные результаты ([26], таблица S1).

5.2.3 Анализ чувствительности параметров подтверждает результаты

Таким образом, обнаружение эпистатических пар возможно лишь в узком временном окне. Зависимость ширины этого окна от параметров модели было изучено следующим образом.

Коэффициент отбора

Процент обнаруженных эпистатических пар и процент ложных пар были рассчитаны на основе моделирования при различных значениях коэффициента отбора s_0 (рисунок 5.11). Ширина окна обнаружения обратно пропорциональна s_0 . Однако, при очень малом s_0 , когда эволюция селективно-нейтральна, эпистаз вообще никогда не обнаруживается.

Распределенный коэффициент отбора

Далее, было рассмотрено изменение s между локусами, в соответствии с полугауссовым распределением. Результаты похожи на случай с постоянным s , хотя частота ложно предказанных пар еще выше ([26], рисунок S5).

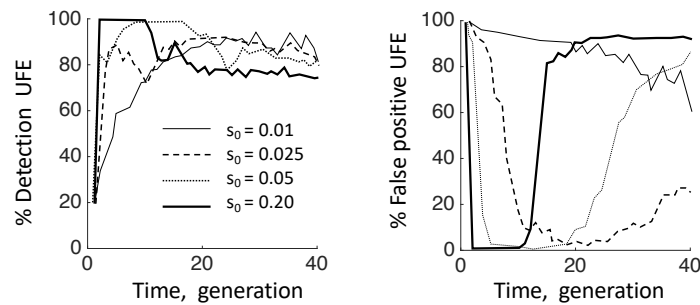


Рисунок 5.11. Обнаружение эпистаза ограничено временным интервалом, контролируемым средним коэффициентом отбора. Процент обнаруженных эпистатических пар и процент ложных предсказаний показаны, как функции времени. Усреднение выполняется по 25 случайным запускам моделирования для каждого значения s_0 (показано). Показателем LD является UFE (5.14). Протокол обнаружения, который максимизирует разницу между процентом обнаружения и процентом ложных предсказаний, путем настройки порога обнаружения, показывает ту же тенденцию. Момент времени $t = 1.5/s_0$ является началом уменьшения обнаружения. В начале $f_0 = 0.45$, остальные параметры как на рисунке 5.9. По работе [26].

Длина генома и размер популяции

В соответствии с предыдущими результатами [106, 107, 145], увеличение длины генома (количества локусов) L ухудшает выявляемость эпистаза ([26], fig. S5). Этот факт известен как «проклятие размерности»: количество возможных пар локусов увеличивается с длиной генома как L^2 , в то время как количество эпистатических пар увеличивается только как L , так что найти «иголку в стоге сена» становится все сложнее.

Численность населения

Окно времени обнаружения расширяется логарифмически с увеличением размера популяции N ([26], fig. S5). Для многолокусных эволюционных моделей [3], где всевозможные наблюдаемые величины имеют логарифмическую зависимость от N , этот результат очевиден.

Начальная вариация

Окно обнаружения во времени существует только тогда, когда начальная популяция достаточно полиморфна, пропорция невыгодных аллелей $f(0) > 10\%$ для параметров в подписи к рисунку 5.9 ([26], fig. S5).

Рекомбинация несколько улучшает обнаружение

До сих пор, эволюция была бесполой. Роль рекомбинации в обнаружении эпистаза была также изучена. Дополнительные параметры модели представляют собой среднее число случайных перекрещиваний на геном, M , и вероятность рекомбинации генома, r . Результаты показали, что умеренная скорость рекомбинации помогает обнаружить эпистаз, разрушая генетическое сцепление в большей степени, чем эпистатический след. При наборе параметров как на рисунке 5.9 наблюдалось значительное уменьшение флуктуаций сцепления, начиная с $r = 20\%$ и $M = 5$ (рисунок 5.12). Таким образом, LD менее чувствительно к рекомбинации, чем скорость адаптации, которая возрастает при гораздо меньших значениях r [15, 76]. Слишком частая рекомбинация нарушает эпистатические корреляции и делает эпистаз необнаружимым. Таким образом, существует оптимальное окно частот рекомбинации для эпистатического обнаружения.

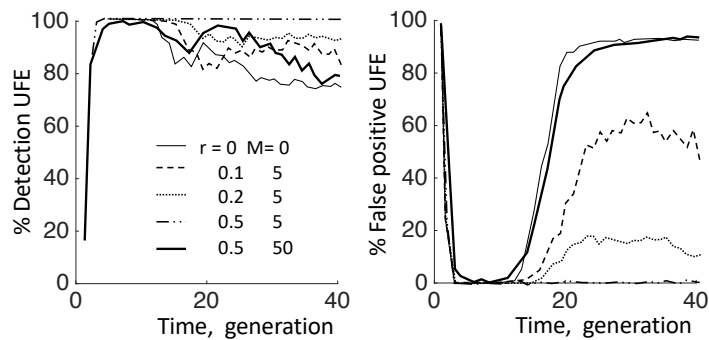


Рисунок 5.12. Изменение временного окна обнаружения при рекомбинации. Процент обнаружения эпистаза и доля ложных предсказаний, как функция времени, усредненных по 25 случайным прогонам моделирования. Похожие результаты получены для других показателей корреляции ([26], рисунок S6-S10). Параметры: $E = 0.75$, остальные как на рисунке 5.9. Значения s выбраны из полугауссовского распределения. По работе [26].

5.2.4 Разбегание популяций ответственно за сильные эффекты сцепления

Чтобы понять причину сильных эффектов сцепления, скрывающих эпистатический отпечаток, было построено филогенетическое дерево при $r = 0$, с использованием алгоритма иерархической кластеризации (рисунок 5.13). Исходная популяция имеет случайное расположение аллелей и, следовательно, звездообразную дерево. С течением времени, последовательности становятся все более близкими генетически, и дерево становится однобоким, с уменьшающимся числом предков. В конце концов, дерево превращается в дерево типа Больтхаузена-Шнитмана (BS) с одним общим предком, как и положено для режима бегущей волны [89, 90, 257].

Возникновение дерева типа BS происходит примерно в то же время, когда случайные ассоциации невзаимодействующих пар локусов возрастает (рисунок 5.9). Причина — движения популяции в случайном направлении, как показано деревьями трех независимо эволюционировавших популяций в отдельных прогонах моделирования (рисунок 5.13). Генетическая дистанция между их самыми недавними общими предками линейно увеличивается во времени, из-за фиксации полезных мутаций в случайных локусах. Поскольку конфигурации гаплотипов общих предков наследуются всеми членами популяций, взаимная дивергенция популяций приводит к сильному LD со случайным знаком. Такое филогенетическое поведение делает очевидным, что любой эпистатический отпечаток безвозвратно теряется на фоне случайного конаследования аллелей на разных локусах, независимо от используемого метода.

5.2.5 Использование большого числа популяций спасает след эпистаза

Поскольку флуктуации сцепления возникают из-за стохастической дивергенции популяций, естественной идеей является усреднение частот гаплотипов по нескольким популяциям. Чтобы проверить это, несколько популяций были созданы при моделировании с одним и тем же набором параметров. Частоты гаплотипов в формулах (5.13) и (5.14) усреднялись по популяциям, для каждой пары локусов. Как и ожидалось, использование достаточного количества популяций приводит к существенному уменьшению шума, а интервал обнаружения расширяется [26]. В § 5.3, это усреднение рассмотрено более подробно.

5.2.6 Заключение

С помощью моделирования эволюционной динамики бесполой популяции методом Монте-Карло, рассчитана плотность распределения различных количественных мер аллельных корреляций и их комбинаций, отдельно для эпистатических и неэпистатических пар локусов. За исключением узкого промежутка времени, аллельные корреляции для неэпистатических пар локусов имеют случайный знак, и их нельзя отличить от корреляций из-за эпистаза. Эффекты случайного генетического сцепления преобладают над эффектами эпистаза в широкой области параметров. Увеличение аллельных корреляций со временем совпадает с установлением генеалогии типа Болтхаузена-Шнитмана, предсказанного для режима бегущей волны [89, 90,

257]. Все последовательности в популяции имеют высокую степень сходства с последовательностью их недавнего общего предка, которая расходится линейно по времени с общими предками других независимых популяций (рисунок 5.13). Это филогенетическое расхождение и является причиной случайных корреляций маскирующих след эпистаза в геномных данных.

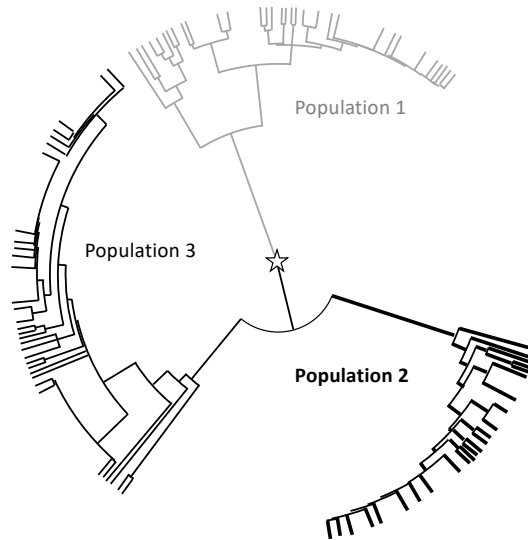


Рисунок 5.13. Филогенетические деревья независимо эволюционирующих популяций растут в разные стороны. Филогенетическое дерево реконструировано для трех популяций, независимо эволюционирующих из одного и того же исходного состояния, в течение 30 поколений. $N = 20\,000$ геномов, начальная аллельная частота $f_0 = 0.40$, остальные параметры как на рисунке 5.9. По работе [26].

5.3 Обнаружение эпистатических пар локусов методом трехлокусной корреляции

Таким образом, случайная ковариация из-за генетического сцепления полностью скрывает эпистатический отпечаток в геномных образцах [26]. Это ограничение не может быть устранено никаким методом обнаружения эпистатических пар, включая широко-используемый подход QLE [148, 150]. Единственный способ помочь обнаружению — усреднить частоты гаплотипов по многим независимым популяциям, развивающимся в сходных условиях. Без множественных популяций, в принципе невозможно обнаружить эпистаз. Но сколько независимых популяций нужно для надежного усреднения? Оказывается, что даже в коротком геноме из 40 локусов, двухсот популяций недостаточно, чтобы подавить флуктуации и засечь эпистаз [28]. Для более длинных геномов из тысяч локусов, необходимое количество популяций быстро растет и становится нереалистично большим.

Метод устранения шума при весьма скромном числе независимых популяций разработан в работе [28], которая рассматривается в данном разделе. Интуитивная идея метода состоит в том, чтобы, при вычислении корреляции для пары локусов, брать для усреднения только последовательности, которые имеют консенсусный аллель на одном из локусов, соединенных с этой парой локусов подозреваемой эпистатической связью. Затем, выбирается другой соседний локус и, таким образом, находится наиболее важный соседний локус, который прерывает цепочку ложного взаимодействия. Такой фильтр, основанный на чувствительности парной корреляции к соседним аллелям, разбивает геном на статистически-независимые блоки.

Для демонстрации высокой точности этого метода обнаружения, используются два параллельных метода: (а) аналитический вывод на основе приближения квазиравновесия (§ 5.1) и (б) моделирование методом Монте-Карло. Метод работает как против не прямых взаимодействий, так и генетического сцепления. Моделирование рассматривает случай положительного (компенсаторного) эпистаза, $0 < E < 1$, но метод работает при любом значении E , включая отрицательный эпистаз. Как и в § 5.1, предполагается, что система находится в процессе адаптации и все еще далека от стационарного состояния. Коэффициент отбора должно быть либо постоянным, либо меняться медленно, т.е., на шкалах времени больших, чем его обратное значение.

После тестирования на моделированных геномах, метод применен к белковым последовательностям нейроминидазы вируса гриппа А. Используя более 8000 последовательностей вирусов гриппа, полученных из разных географических регионов до и после пандемии 2009 года, эпистатические взаимодействия отделены от эффектов сцепления. Данный белок в новом штамме изменился на 20% последовательности, но в какой-то момент в прошлом, у двух штаммов был общий предок. Из этого 20% изменения, в данной работе выявлены первичная мутация и компенсирующие мутации, которые позволили новому штамму вытеснить старый.

5.3.1 Компьютерная модель для создания последовательностей для тестирования метода

Модель популяции та же, что и в § 5.1 и включает факторы случайной мутации, случайного генетического дрейфа, и постоянного направленного отбора. Рекомбинации нет. Некоторые пары локусов отобраны для взаимодействия с относительной положительной амплитудой $E = 0.75$. Для первой демонстрации, выбрана простая эпистатическая сеть двойных дуг (рисунок 5.14В). Эта сеть производит три типа межлокусных аллельных корреляций: (i) эпистатические взаимодействия, (ii) не прямые взаимодействия через третий локус и (iii) генетическое сцепление

из-за наличия общего предка (§ 5.2). Ниже также рассмотрена более сложная топология с петлями.

5.3.2 Первый шаг: усреднение по популяциям

Первая задача — обнаружить все потенциальные эпистатические взаимодействия с помощью корреляционного анализа. Для каждой пары оснований (i, j) мера корреляции, введенная в § 5.1, имеет вид

$$UFE_{ij} = 1 - \frac{\log\left(\frac{f_{11}}{f_{00}}\right)}{\log\left(\frac{f_{01}f_{10}}{f_{00}^2}\right)} \quad (5.14)$$

где $f_{00}, f_{10}, f_{01}, f_{11}$ — частоты гаплотипов, усредненные по результатам моделирования.

Другие показатели корреляции, такие как коэффициент Пирсона r^2 , который использовался в § 5.2, порождают аналогичный стохастический шум (рисунок 5.9). По сравнению с этой статистикой, UFE_{ij} имеет то преимущество, что не только обнаруживает ассоциацию аллелей, а также напрямую измеряет степень их взаимной компенсации, $UFE = E$ (если взаимодействующая пара взаимодействует с другими локусами, то с вычисляемой поправкой) (рисунок 5.6d, § 5.2). Поскольку первая часть (5.14) имеет особую точку, когда одна из четырех частот гаплотипов равна нулю, рассматриваются только такие пары оснований, что $f_{ij} > f_{cut}$, где f_{cut} — нижний порог, установленный здесь при $f_{cut} = 0.05$. Кроме того, ниже будут рассматриваться только пары с большой корреляцией, $UFE_{ij} > 0.5$. Истинное значение E устанавливается в программе, $E = 0.75$.

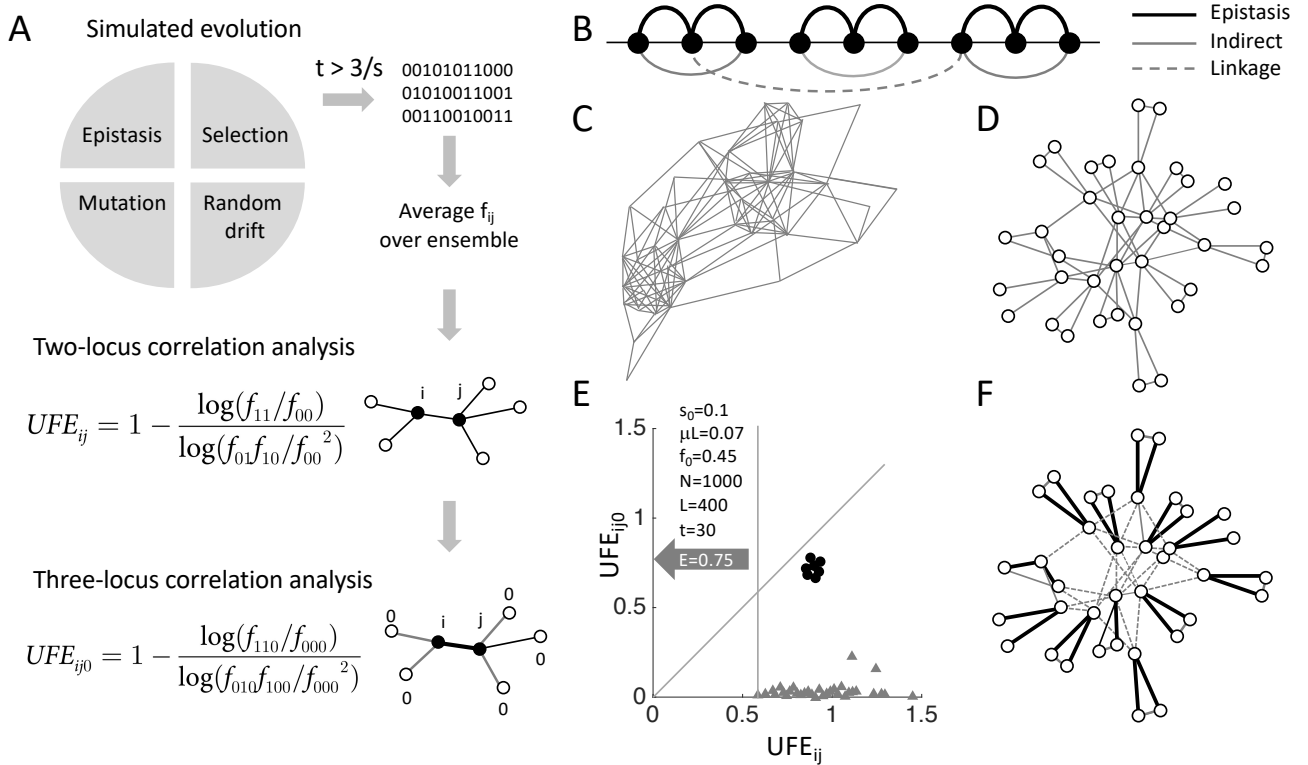


Рисунок 5.14. Схематическая диаграмма метода и его тестирование на смоделированных данных последовательности. А. Компьютерная модель бесполой эволюции включает факторы случайной мутации, отбора, эпистаза и случайного генетического дрейфа. Частоты парных гаплотипов f_{ij} усредняются по результатам моделирования (независимые популяции). Мера парной корреляции UFE_{ij} рассчитывается по формуле (4.14). Косвенные связи и остаточная связь обнаруживаются и отфильтровываются с помощью показателя тройной корреляции UFE_{ij0} по (5.15). В. Предусмотренная эпистатическая сеть на 40 оснований. Черные толстые кривые: реальные эпистатические взаимодействия. Серые тонкие линии: косвенные взаимодействия. Серая пунктирная линия: пример связи. С и D. Сеть сильных ($UFE_{ij} > 0.5$) возможных эпистатических взаимодействий предсказанная (С) для одной популяции и (D) после усреднения по 200 популяций. Е. Диаграмма рассеяния трехстороннего гаплотипа, $\min(UFE_{ij0})$, показанная против UFE_{ij} для пар, идентифицированных в (D). Заштрихованный сектор соответствует прямым взаимодействиям. Верхняя пунктирная линия — это диагональ, $UFE_{ij} = UFE_{ij0}$, а нижняя пунктирная линия отделяет прямые (черные кружочки) от непрямых взаимодействий и генетического сцепления (серые треугольнички). F. Прогнозируемая сеть точно такая же, как и сеть, заданная при моделировании. Параметры: исходная частота аллеля $f_0 = 0.45$, частота мутаций на геном $\mu L = 0.07$, фиксированный коэффициент отбора $s = 0.1$, $N = 1000$, $L = 40$, эпистатическая сила $E = 0.75$. По работе [28].

Без усреднения f_{ij} по прогонам симуляции, граф содержит сотни ложных взаимодействий, которые совершенно прячут 24 истинные взаимодействия, как несколько иголок в стоге сена (рисунок 5.14C). Для более длинного генома, относительное количество ложных связей будет еще больше, пропорционально L^2 . При усреднении f_{ij} по 200 запускам моделирования (независимым популяциям), количество ложных связей уменьшается на порядок (рисунок 5.14D). Однако около половины оставшихся связей по-прежнему ложные, так как их общее количество, по-прежнему, значительно превышает 24 (рисунок 5.14Б).

5.3.3 Второй этап: трехсторонняя корреляция

Остаточные ложные связи вызваны недостаточным усреднением долей гаплотипов f_{ij} по ограниченному ансамблю. Чтобы их удалить, можно пойти в двух направлениях. Либо геномные данные должны быть отобраны из тысяч независимых популяций, либо нужно найти какой-то лайф хак. Первый подход непрактичен и даже невозможен, в большинстве случаев. Такие данные часто отсутствуют и, поскольку относительный шум из-за генетического сцепления увеличивается с длиной генома L , также возрастает и количество независимых популяций, необходимых для усреднения.

Предложенный протокол и является подобным «домашним средством». Он перерезает окольные пути вокруг пары локусов, которые создают ложноположительные связи (рисунок 5.14D). Окольные пути бывают двух видов: (i) не прямые взаимодействия, т.е., цепочки эпистатических взаимодействий [152], и (ii) аллельные комбинации с общими предками (генетическое сцепление) (рисунок 5.14B). Чтобы сократить потенциальный обход интересующей пары оснований, для анализа выбираются только геномы, в которых соседний локус пары имеет аллель 0 (наиболее приспособленный или консенсус). Остальные геномы игнорируются. Затем пересчитывается признак корреляции. Если соседний локус находится на важном окольном пути, эта процедура существенно уменьшит косвенную корреляцию. Корреляции за счет прямых эпистатических взаимодействий, разумеется, не чувствительны к обходным путям и их перерезанию.

Для данного плана, для каждой остаточной связи i, j (рисунок 5.14D) вычисляется трех-локусная корреляция

$$UFE_{ij0} = 1 - \frac{\log(f_{110}/f_{000})}{\log(f_{010}f_{100}/f_{000}^2)} \quad (5.15)$$

где 0 в третьем положении обозначает консенсусный аллель на одном из локусов, связанном с локусом i или j . Все связанные локусы рассматриваются как 0-узел, один за другим, и для каждого вычисляется величина (5.15). Затем вычисляется минимальное значение по всем 0-узлам, $\min(UFE_{ij0})/UFE_{ij}$, для обнаружения наиболее важного окольного пути, если их несколько. Ложные связи определяются как связи с низким отношением $\min(UFE_{ij0})/UFE_{ij}$. Процедура повторяется для каждой потенциальной связи (рисунок 5.14D).

Точечная диаграмма подтверждает, что, для ложных пар, $\min(UFE_{ij0})$ в несколько раз меньше, чем UFE_{ij} (рисунок 5.14E). Для прямых эпистатических взаимодействий, как и ожидалось, эти два показателя почти одинаковы. Точное значение порога в $\min(UFE_{ij0})/UFE_{ij}$ не имеет значения, если частоты гаплотипов усредняются, по крайней мере, по 20 популяциям (для выбранного набора параметров). В итоге, восстановленная эпистатическая сеть в точности совпадает с сетью, заданной при моделировании, что означает 100% обнаружения и 0% ложных

связей (рисунок 5.14В). В дополнение к идеальному обнаружению, использование UFE дает довольно точные оценки эпистатической силы: $UFE \approx E$ в пределах 15% точности (рисунок 5.14F, цифры по ссылкам). Таким образом, настоящий метод обеспечивает 100% достоверность при $L = 40$ локусах или менее, при числе повторных популяций 200 и эпистатической силе $E = 0.5 - 0.75$. Точность обнаружения становится куда хуже при малом $E < 0.25$.

5.3.4 Аналитическая проверка для косвенных взаимодействий

Исключение косвенных взаимодействий было исследовано аналитически, в предположении что доли гаплотипов хорошо усреднены по популяциям. Для локусов 1 и 3 одной из парных дуг на рисунке 5.14В, было показано [28]

$$UFE_{ij} \approx \frac{1}{4(1-E)}$$

$$UFE_{ij0} \approx 0$$

что и объясняет «выпадание в осадок» косвенных взаимодействий на рисунке 5.17Е. Аналитический вывод [28] использовал приближение адаптивного квазиравновесия [24], введенное в § 5.1 (не путать с «квазиравновесием сцепления» в присутствии рекомбинации [148]).

Таким образом, косвенная ковариантность разрушается путем выборки только тех последовательностей, которые имеют консенсусный аллель между двумя интересующими нас локусами. Действительно, по формуле (5.1), взаимодействуют только миноритарные аллели. Как только в середине окольного пути создающего косвенную ковариантность затребована аллель консенсуса, путь прерывается, и ложное взаимодействие исчезает.

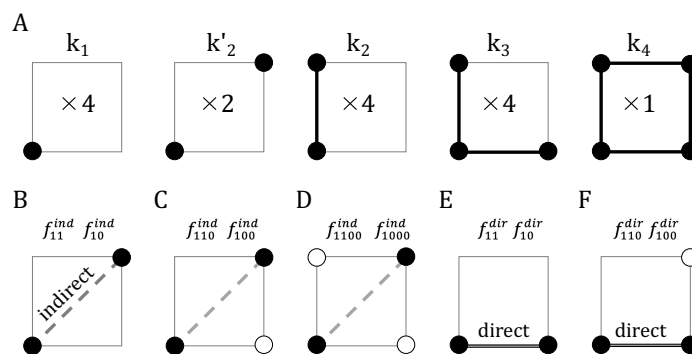


Рисунок 5.15. Вывод UFE для замкнутой квадратной топологии эпистатической сети. А) Возможные конфигурации и их симметрия. В) Косвенное взаимодействие. С) Косвенное взаимодействие с фиксированным нулем в узле. D) Косвенное взаимодействие с двумя фиксированными нулями. Е) Прямое взаимодействие. F) Прямое взаимодействие с фиксированным нулем. По работе [28].

Хотя вывод выше получен для косвенных взаимодействий, тот же метод оказывается столь же эффективным и для устранения эффектов генетического сцепления (сравните рисунок 5.14В с рисунком 5.14Е). Хотя причину еще предстоит понять, вероятно, этот метод разрывает цепочки миноритарных аллелей с общим предком так же, как он прерывает цепочки взаимодействий.

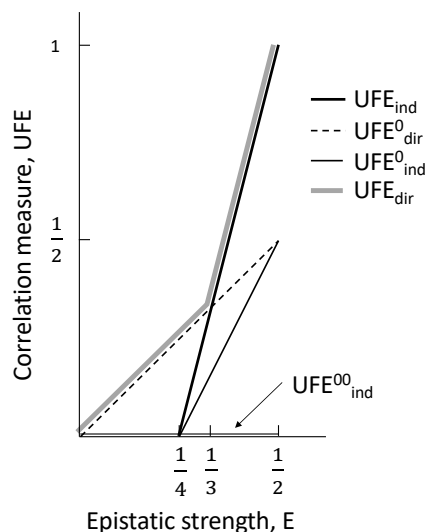


Рисунок 5.16. Зависимость пяти вариантов аллельной корреляции (рисунок 5.15В-Е) от величины эпистаза E , предсказанная для сети из замкнутых квадратов. По работе [28].

По той же методике, была также рассмотрена несколько более сложная сеть, состоящая из $L/4$ изолированных замкнутых квадратов с величиной эпистаза E (рисунок 5.15А) [28]. В этом случае, между противоположными углами квадрата, имеется косвенное взаимодействие, а локусы стороны взаимодействуют напрямую. Основное отличие от сети парных дуг в том, что квадрат образует петлю, так что противоположные углы соединяются двумя путями (топологическая связность 2). Результаты показывают, что при $E > 1/3$, прямые и косвенные корреляции имеют одинаковую амплитуду. Использование трехлокусной корреляции вместо двухлокусной снижает обе корреляции, но косвенную в большей степени (рисунок 5.15В, табл. 5.3). В отличие от сети со связностью 1, косвенная корреляция полностью не исчезает и сохраняет тот же порядок величины, что и прямая. Этот результат интуитивно ожидаем, потому что есть два равных обходных пути, соединяющих углы, и оба должны быть перерезаны для получения сильного эффекта. Естественный способ решить эту проблему — добавить еще один 0 и измерить корреляцию по 4 локусам, чтобы прервать оба обходных пути. Этот обобщенный метод действительно полностью устраняет косвенные взаимодействия, $UFE_{ind}^{00} \equiv 0$ во всем интервале E (рисунок 5.16, таблица 5.3).

Таблица 5.3. Прямые и не прямые значения UFE. Индекс 0 соответствует консенсусу на дополнительном локусе, 00 — на двух дополнительных локусах.

Interval	$E < \frac{1}{4}$	$\frac{1}{4} < E < \frac{1}{3}$	$\frac{1}{3} < E < \frac{1}{2}$
UFE_{ind}	0	$4E - 1$	$4E - 1$
UFE_{ind}^0	0	$\frac{4E-1}{2}$	$\frac{4E-1}{2}$
UFE_{ind}^{00}	0	0	0
UFE_{dir}	E	E	$4E - 1$
UFE_{dir}^0	E	E	E

В общем случае эпистатической сети со связностью выше двух, количество дополнительных нулей, необходимых для исключения косвенного взаимодействия, равно числу независимых обходов, т. е. топологической связности сети. На практике связность сети априори неизвестна, и необходимо итеративно добавлять дополнительные нули, пока предполагаемая сеть не перестанет меняться. Для вирусных данных рассмотренных ниже, эта процедура не используется, потому что полученная сеть представляет собой почти что дерево (связность чуть выше 1) уже после теста с тремя локусами.

5.3.5 Применение к вирусу гриппа А

После того как метод обнаружения был протестирован как аналитически, так и в компьютерной модели, он был использован для обнаружения эпистатической сети из геномных образцов реальной популяции. Один из двух поверхностных белков гриппа А H1N1, нейраминидаза (НА), является детерминантой инфекционности вируса и мишенью лекарственной терапии. Вместе с белком гемагглютинином, этот белок обеспечивает проникновение вируса в клетку-хозяина. Задача была в том, чтобы выявить конкретные мутации и их взаимодействия, вызвавшие появление пандемического штамма 2009 года. Для этого, последовательности нового штамма, изолированного в разных точках мира, нужно было сравнить с последовательностями старого штамма, которые имеют 80% взаимной гомологии. Предполагалось, что оба штамма в начале эволюционировали почти независимо и параллельно в разных географических точках, с очень медленной миграцией вируса между ними. На самом деле, вирус в разных местах представляет собой одну метапопуляцию, но это обычное приближение для работы с данными по мировым пандемиям.

Около 8000 последовательностей нейраминидазы вируса гриппа А H1N1 за период 2000-2010 гг были загружены из базы данных <https://www.fludb.org>. Одинаковое количество последовательностей было отобрано из первого и второго штаммов, несколько сотен раз для усреднения. Была также исследована устойчивость результатов к балансу размера выборки между двумя штаммами ([28], рисунок 2F). Было обнаружено, что старый и новый штаммы генетически различны, и новый штамм постепенно вытесняет старый штамм, в течение

нескольких лет. С точки зрения теории бегущих волн [3], этот процесс можно представить с помощью двух бегущих волн с перекрывающимся распределением приспособленности, одна из которых постепенно ослабевает, а другая усиливается и ее вымещает.

Выровненные последовательности белка нейраминидаза были бинаризованы путем обозначения консенсусного аллеля как 0 и любого неконсенсусного аллеля как 1. Более двух аминокислотных вариантов в одном кодоне встречались редко. Были включены только локусы с гетерозиготностью $> 5\%$. Пандемический штаммы до и после 2009 г. различались примерно в 100 аминокислотах белка.

Показатели корреляции были рассчитаны по формулам (5.14) и (5.15). Чтобы избежать сингулярности, были включены только те пары, где все четыре гаплотипа присутствовали в количестве, превышающем порог, $\langle f_{ij} \rangle > 0.05$. Результат инференции — это сеть с 16 эпистатическими взаимодействиями (рисунок 5.17). Локус 248 в белке NA является носителем первичной мутации, связанным с множественными компенсаторными мутациями, что подразумевает прохождение вирусом через долину приспособленности (рисунок 5.17). Тот же вид сети наблюдается, например, для ВИЧ в связи с выработкой устойчивости к лекарствам [6, 113, 289] или ускользанием от иммунного ответа (§ 7.2).

Возможное опасение состоит в том, что предполагаемые парные мутации на рисунке 5.17 не являются эпистатическими взаимодействиями, а скорее представляют собой главную выгодную мутацию с 16 мутациями-пассажирами, едущими на вторгшемся штамме. Однако, как легко показать, в этом случае данный метод не предсказал бы вообще никакой сети. Действительно, можно рассмотреть предельный случай, когда популяция вируса представляет собой смесь из двух однородных штаммов, до и после 2009 года. В 20% генома эти два штамма различаются. Предположим, что их последовательности равны 101100... и 010011..., причем эти различия включают в себя как главную мутацию, так и пассажиров. Как легко видеть, для любой пары оснований присутствуют только два гаплотипа f_{ij} из четырех. Например, для первого и второго локусов, наблюдаются гаплотипы 10 и 01, а гаплотипы 11 и 00 отсутствуют. При расчете корреляций UFE_{ij} и UFE_{ij0} по формулам (5.14) и (5.15), эта пара локусов и все остальные будут исключены, потому что протокол требует присутствия всех четырех гаплотипов, $f_{ij} > 0.05$. Таким образом, предлагаемый протокол не просто измеряет разницу между старым и новым штаммами, а обнаруживает тонкую ассоциацию между *флуктуациями* частот аллелей.

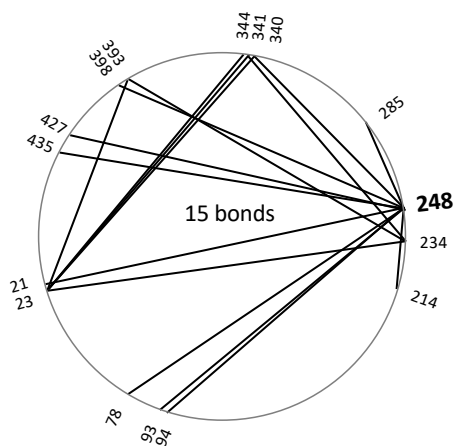


Рисунок 5.17. Эпистатическая сеть (долина приспособленности), ответственная за пандемию 2009 года и предсказанная из последовательностей белка нейраминидазы гриппа А H1N1, полученных в период с 2005 по 2010 год. По работе [28].

5.3.6 Структурная интерпретация

Чтобы понять биохимическую природу предполагаемых компенсаторных участков в белке NA (рисунок 5.17), они были помечены на трехмерной структуре белка (рисунок 5.18). Активный карман NA (RBD, фиолетовый цвет) выполняет функцию связывания рецептора вируса, сиаловой кислоты, на поверхности клетки-мишени. Обратите внимание, что предполагаемая первичная мутация 248 расположена на краю этого кармана. Все предсказанные компенсаторные мутации расположены на поверхности белка внутри α -спиралей, соединяющих и определяющих пространственную ориентацию β -листов. Ни одна не расположена внутри белка или в β -листе. Можно предположить, что их биохимическая роль заключается в восстановлении правильного расположения β -листов, нарушенных первичной мутацией 248. Это явный случай аллостерического взаимодействия, не связанного с прямым связыванием.

В отличие от эпистатических связей (рисунок 5.17 и 5.18), ложные связи, устраняемые данным методом, не привязаны к какой-либо конкретной биохимической структуре. Следствие родословной движимой естественным отбором и случайными процессами, генетическое сцепление в биохимическом смысле неразборчиво, и любая полиморфная пара в базе данных может образовать ложную связь (рисунок 5.14С).

5.3.7 Заключение

Разработан и протестирован метод обнаружения немногочисленных взаимодействующих пар локусов в геномных данных, на фоне шума из-за генетического сцепления и не прямых взаимодействий. Этот метод основан на приближении адаптивного квазиравновесия [24] (§ 5.1). Высокая точность метода продемонстрирована аналитически и с помощью компьютерного моделирования для заранее известной эпистатической сети.

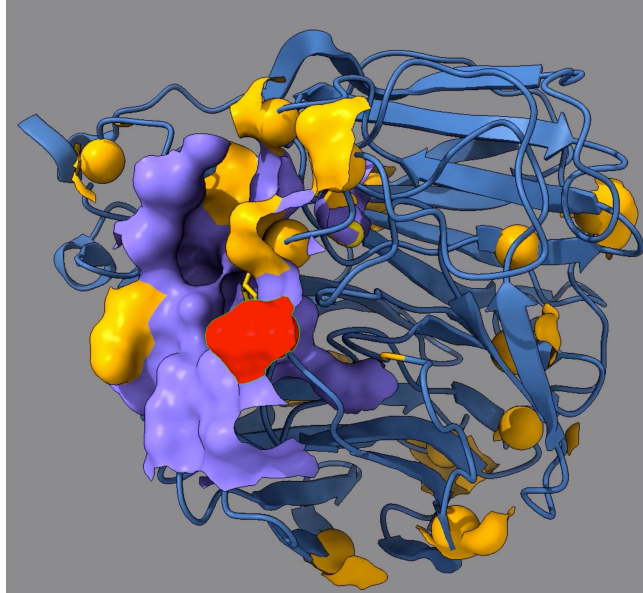


Рисунок 5.18: Структурное расположение предсказанной сети эпистаза для появления нового штамма нейраминидазы вируса гриппа в 2009 году. На рисунке показана трехмерная структура нейраминидазы гриппа А H1N1 (идентификационный код PDB 4QVZ). Цветные сферы представляют предсказанные эпистатические узлы (рисунок 5.17). Красная сфера: предсказанная первичная мутация (остаток 248). Оранжевые сферы: компенсационные узлы (рисунок 5.17). По работе [28].

Этот метод использован для выявления первичных и вторичных мутаций, ответственных за пандемию «свиного гриппа» 2009 года. Результаты предсказывают первичную мутацию и 16 сильных компенсаторных мутаций, что напоминает ситуацию с штаммами ВИЧ устойчивыми к антиретровирусным препаратам [6, 113, 289] (§ 2.1). Предполагаемая первичная мутация 248 наблюдалась после 2009 г. в различных географических точках [420-422]. Она повышает инфекционность вируса через расширения диапазона допустимой кислотности среды [423]. Хотя эмпирические доказательства за или против предполагаемой сети пока неизвестны, для проверки этого предсказания подошел бы метод, так называемого, «глубокого сканирования» [424, 425].

Важно подчеркнуть, что другие существующие методы не подходят для устранения корреляций по сцеплению. Например, очень популярный метод, предназначенный для фильтрации косвенного взаимодействия, основан на приближении квазиравновесия сцепления (QLE), предполагающем, что рекомбинация и эпистаз являются единственными противодействующими силами, а эффекты сцепления отсутствуют, что справедливо только для огромных популяций [148, 150]. Ключевым ограничением метода QLE является игнорирование

эффектов сцепления, которые существуют в реальной жизни даже для разных видов, поскольку любые два вида имеют общего предка, а часть общих белков всегда конаследуется. Следовательно, этот метод можно применять только к данным о разных видах животных, которые разошлись миллионы лет назад, и только к недавним аллелям [152, 153]. Метод трехлокусных корреляций не требует ограничений в виде больших размеров популяции и работает даже для бесполок популяций. Он не предполагает отсутствия эффектов сцепления, как QLE, а отфильтровывает их вместе с косвенными взаимодействиями. Метод с тремя локусами будет работать с геномами и белками, отобранными у одного и того же вида, если доступно хотя бы 20-100 независимо-эволюционирующих популяций.

5.4 Оценка коэффициентов отбора отдельных локусов из геномных данных

§§ 5.1 - 5.3 были посвящены обнаружению и измерению эпистатических взаимодействий между локусами. Адаптивный ландшафт включает также эффекты отдельных мутаций, коэффициенты отбора. Вместе с эпистазом, они необходимы для вероятностного прогнозирования эволюционной траектории популяции патогенов, например, при развитии лекарственной устойчивости или ускользании от иммунной системы (*глава 7*). Поэтому на разработку различных методов инференции коэффициентов отбора из геномных данных были затрачены большие усилия. Существующие методы основаны на различных упрощенных моделях, таких как модели с одним локусом (§ 2.1). Как обсуждалось в *главах 2 и 3*, эффекты многолокусного сцепления полностью модифицируют эволюционную динамику, делая динамику отдельных локусов взаимно зависимой даже при отсутствии эпистаза.

В работе [29] впервые разработан метод измерения коэффициентов отбора для отдельных локусов, независимый от сильных эффектов генетического сцепления. Этот метод применим, когда популяция находится в процессе адаптации и пока еще далека от конечного стационарного состояния (мутационно-селекционного равновесия). Отправной точкой послужило экспериментальное наблюдение, что плотность распределения коэффициентов отбора s между локусами, $DFE(s)$, наблюдаемая в экспериментах над вирусами и бактериями растущими в клеточных культурах, часто имеет простую экспоненциальную форму [91-95] (рисунок 5.19). В работе [29] показано, с помощью компьютерной модели и анализа, что экспоненциальная форма возникает естественным образом, в процессе адаптации. Результаты работы подчеркивают разницу между распределением DFE , наблюдаемым в экспериментах по эволюции вируса, и «внутренним» распределением, измеряемым в эксперименте по мутагенезу на выбранном локусе.

В то время как первое часто имеет универсальную экспоненциальную форму, второе может иметь любую форму и меняется от белка к белку, от вируса к вирусу, от биологического вида к виду. На основе этого теоретического результата, разработан [29] метод измерения эффекта приспособленности отдельных мутаций, нечувствительный к эффектам сцепления и создано программного обеспечение [31].

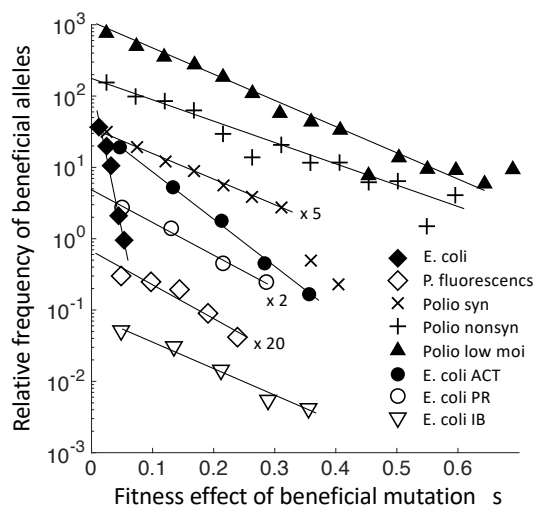


Рисунок 5.19. Эксперименты над вирусами и бактериями растущими в клеточной культуре, показывающие экспоненциальное распределение оснований генома по коэффициенту отбора s . Ордината: доля полезных мутаций $DFE(s, t)\varphi(s)$, определенная в формуле (5.16). Абсцисса: увеличение приспособленности из-за полезной мутации (коэффициент отбора). Символы представляют результаты, полученные для разных участков генома в экспериментах с *Escherichia coli* [91], *Pseudomonas fluorescens* [92], синонимичными мутациями полиовируса, несинонимичными мутациями полиовируса [93], полиовирусом с низким МОИ [94], ацетиамидом *E. coli* (ACT), пропионамидом (PR) и изобутирамидом (IB) [95]. По работе [29].

5.4.1 Экспериментальное распределение коэффициентов отбора имеет универсальный вид

Ключом к пониманию экспоненциальной формы DFE (рисунок 5.19) является то, что полезные мутации могут возникать только на локусах, занятых невыгодными аллелями. (Как и везде, предполагается два возможных аллеля на локус.) Хорошо адаптированная популяция имеет выгодные аллели на большинстве локусов. Следовательно, наблюдаемый DFE определяется оставшимися локусами, которые имеют невыгодные аллели. Их количество уменьшается со временем t , при этом локусы с большими s исчезают первыми. Средняя частота невыгодных аллелей в локусе с коэффициентом отбора s , обозначаемая $f(s, t)$, представляет собой частоту локусов, доступных для полезных мутаций. Если система полностью адаптирована, мы имеем $f = 0$, и дальнейшая полезная мутация невозможна.

В ходе экспериментов, исследователи переносят размножившиеся бактерии или вирусы из клеточной культуры в следующую, и так много раз, и наблюдают за появлением и ростом

новых выгодных мутаций, одной за другой по отдельности. Форма кривой роста во времени подгоняется к однолокусной модели для оценки значений s . Например, группа [93] использовали метод секвенирования *CirSeq* для мониторинга выгодных аллелей вируса полиомиелита и так оценивали их s . Другая лаборатория [91] проделала аналогичную процедуру с кишечной палочкой. Затем исследователи подсчитывают количество таких аллелей, принадлежащих заданному интервалу s (абсцисса на рисунке 5.19).

Первая задачей было понять, что же именно наблюдают эти эксперименты, в количественном смысле. Предположим, на менее приспособленном локусе, случилась мутация с преимуществом s . Из-за случайного генетического дрейфа, новый аллель будет скорее всего быстро потерян, но он может быть и выжить и начать расти в популяции, с некой вероятностью $\varphi(s)$ [16]. Если новые аллели растут по отдельности и не перекрываются, $\varphi(s) = s$ (§ 2.2). Поэтому, эксперимент (рисунок 5.19) измеряет мутации с наблюдаемой плотностью распределения в виде произведения

$$DFE_{\text{obs}}(s, t) = DFE(s, t)\varphi(s) \quad (5.16)$$

В формуле (5.16), $DFE(s, t)$ само состоит из двух множителей

$$DFE(s, t) \equiv f(s, t)g(s) \quad (5.17)$$

где $g(s)$ обозначает «внутреннюю плотность распределения» локусов с коэффициентом отбора s . Функция $g(s)$ может быть измерена в эксперименте путем проведения мутагенеза для каждого основания, одного за другим, и измерения изменения в приспособленности до и после мутации [424, 425]. Таким образом, наблюдаемая и внутренняя плотность локусов по s качественно различаются [81].

В формуле (5.17), средняя доля вредных аллелей $f(s, t)$ зависит от состояния популяции и времени. Внутреннее распределение $g(s)$, наоборот, является свойством самого вируса и клеточной культуры. Оно не зависит ни от времени, ни от состояния популяции, не универсально, и широко варьируется между вирусами и белками. В этом разделе предполагается, что средняя по ансамблю частота мутантов $f(s, t)$ предварительно эволюционировала какое то время до эксперимента в похожих условиях. Предполагается, что популяция еще далека от стационарного состояния (равновесия мутации и отбора), т. е. еще недостаточно адаптирована. В этом случае, вредными мутациями можно пренебречь (§ 5.1).

Как показано ниже четырьмя способами, в долгосрочной перспективе, $f(s, t)$ экспоненциально зависит от s , причем производная по s линейно растет со временем. А вот внутреннее распределение $g(s)$, для данного вида, задано раз и навсегда. Поэтому, рано или поздно, $f(s, t)$ будет меняться с s гораздо быстрее, чем $g(s)$, и, следовательно будет доминировать зависимость DFE от s в логарифмическом масштабе. Это и объясняет экспериментальные результаты (рисунок 5.19).

5.4.2 Модель

Популяционная модель и метод компьютерного моделирования — те же что использовалась в §§ 5.1–5.3, за исключением эпистатического члена в (5.1), который теперь отброшен, предполагая что вклад локусов в логарифм приспособленности аддитивен. Парные эпистатические эффекты рассматривались в §§ 5.1–5.3. Данное приближение работает на умеренных временных промежутках. В этом случае, эпистатические взаимодействия только перенормируют текущие значения коэффициенты отбора s_i . В краткосрочной перспективе, немногие основания являются полиморфными и еще меньше оснований взаимодействуют, так что, для поставленной цели, это хорошее приближение. Если белок эволюционирует долго, значительная часть его последовательности изменится, что повлияет на значения s_i [6, 424-427].

Внутренняя плотность распределения $g(s)$ является входным параметром модели. Как уже упоминалось, ее форма широко варьируется между организмами и белками, и на конечный ответ она не влияет. В компьютерном моделировании ниже, использована положительная половина нормального распределения

$$g(s) = \frac{2}{s_{av}\pi} \exp\left(-\frac{s^2}{\pi s_{av}^2}\right), \quad s > 0 \quad (5.18)$$

где s_{av} — среднее значение. В аналитическом выводе данном ниже (§ 5.4.4), не делается никаких предположений относительно формы $g(s)$.

5.4.3 Результаты Монте-Карло

Первые 100 последовательностей популяции показаны в разные моменты времени (рисунок 5.20А-Е). На оси абсцисс, геномные локусы i ранжированы в порядке возрастания коэффициента отбора s_i слева направо, начиная с $s_i \approx 0$. В процессе адаптации, плотность вредных аллелей перераспределяется по s_i . Вредные аллели с высокой s_i вымирают первыми, что и объясняет наблюдения, на качественном уровне (рисунок 5.19). Можно заметить два различных интервала времени:

(i) *Естественный отбор ранее существовавших аллелей*

На ранней стадии эволюции $t \ll 1/s_{av}$, где s_{av} — среднее значение $g(s)$, новые мутации не важны, и естественный отбор ранее существовавших аллелей является единственным действующим фактором эволюции (рисунок 5.20А, В). Такой режим наблюдался экспериментально для вируса везикулярного стоматита, развивающегося в культуре [13] и

полиовируса в мышах [22]. В этом временном интервале, распределение аллелей по геномам имеет случайный вид.

(ii) *Бегущая волна*

На поздней фазе, $t > 1/s_{av}$, новые полезные мутации имеют решающее значение, поскольку они создают новые геномы, более приспособленные, чем самые приспособленные из начальных последовательностей (рисунок 5.20B-E).

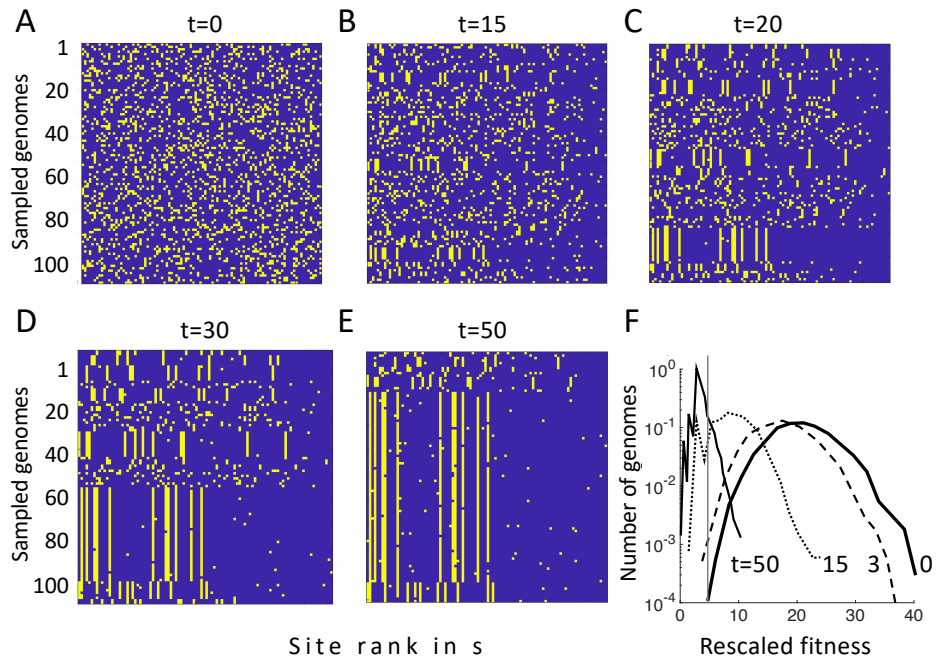


Рисунок 5.20. Вредные аллели с большим эффектом, в процессе адаптации, истощаются первыми. (A-E) Эволюция выборки из 10^2 последовательностей. Фиолетовые точки: полезные аллели, желтые точки: вредные аллели. Ось абсцисс: ранг коэффициента отбора s . Значения $s > 0$ распределены по формуле (5.18), со средним значением $s_{av} = 0.05$. Геномные локусы упорядочены по значению s . Ось ординат: номер генома в образце. Исходная популяция рандомизирована со средней частотой вредных аллелей $f_{in} = 0.2$. Показаны моменты времени в поколениях для одного прогона. (F) Эволюция распределения генома по приспособленности. Ось ординат: эффективное количество вредных аллелей, определенное как $k = -W/s_{av}$, где W — приспособленность генома. Показаны четыре момента времени. Вертикальная серая линия показывает наиболее приспособленный класс геномов при $t = 0$. Появление клональной структуры в (A-E) показывает переход от режима естественного отбора ранее существовавших последовательностей к режиму бегущей волны с новыми мутациями. Параметры: размер популяции $N = 10^4$, количество локусов $L = 100$, частота мутаций на геном $\mu L = 0.05$. По работе [29].

На больших временах, группы идентичных последовательностей (клонов) растут из высокоприспособленного края бегущей волны (рисунок 5.20F). Клоны хорошо видны на рисунке, как вертикальные столбцы сцепленных вредных аллелей (рисунок 5.20B). Волна движется в сторону высокой приспособленности. По мере продвижения волны, клональная структура становится более выраженной (рисунок 5.20C-E). В конце концов, все геномы популяции имеют одного общего предка, который, в начальной популяции, был одним из наиболее приспособленных геномов.

5.4.4 Аналитический вывод универсального DFE

Рассмотрим усредненную по ансамблю частоту мутаций $\langle f(s, t) \rangle$ и предположим, что популяции еще не близки к устойчивому состоянию, так что можно пренебречь событиями вредных мутаций. Экспоненциальная форма $\langle f(s, t) \rangle$ выводится тремя независимыми способами [29], со следующими результатами.

(i) Ранняя эволюция

В ранние времена $t < s_{av}$ единственным фактором является естественный отбор ранее существовавших аллелей, и мутациями можно пренебречь. Рассмотрим локус с коэффициентом отбора s . Остальные коэффициенты отбора в геноме неизвестны и распределены на локусах независимо друг от друга, согласно плотности распределения $g(s)$. В исходной популяции, менее приспособленные аллели случайным образом распределены между аллелями и особями, со средней частотой f_{in} . Результат, для средней доли невыгодных аллелей на этом локусе, имеет вид [29]

$$\langle f(s, t) \rangle = \frac{f_{in}}{(1 - f_{in}) e^{ts} + f_{in}} \quad (5.19)$$

Обратите внимание, что уравнение (5.19) не зависит от вида $g(s)$. При малом f_{in} , его можно аппроксимировать экспонентой

$$\langle f(s, t) \rangle \approx f_{in} \exp(-ts), \quad t < t_0 \sim 1/s_{av}, \quad f_{in} \ll 1$$

Экспоненциальный наклон распределения вредных аллелей

$$\beta = -\frac{\partial \log f}{\partial s} \quad (5.20)$$

в этом случае приблизительно равен времени, $\beta = t$ (рисунок 5.21B). Этот результат объясняет результаты эксперимента уже на количественном уровне (рисунок 5.19), предполагая, что вирус предэволюционировал перед измерительным экспериментом только в течение короткого времени. В случае более длительного периода предэволюции, нужен совсем другой подход.

(ii) Режим бегущей волны

В режиме бегущей волны, который начинается примерно при $t_0 \sim 1/s_{av}$, вывод более сложный. Во-первых, следует учитывать полезные мутации, потому что они позволяют фронту волны продвинуться вперед. Во-вторых, разные реализации популяции сильно различаются. Большая часть вредных аллелей находится в редких мономорфно-вредных локусах (вертикальные желтые столбцы на рисунке 5.20), которые очень различаются между отдельными популяциями.

Величина $\langle f(s, t) \rangle$ получается только усреднением по многим популяциям, и ее уменьшение происходит за счет закрепления новых полезных аллелей на этих мономорфно-вредных локусах. При этом усреднении, немногими полиморфными локусами можно пренебречь. Уравнение динамического баланса, описывающее этот процесс, имеет вид

$$\frac{\partial \langle f(s, t) \rangle}{\partial t} = -\mu N \varphi(s) \langle f(s, t) \rangle, \quad t > t_0 \sim 1/s_{av} \quad (5.21)$$

Начальное условие для уравнения (5.21) задается ранней временной фазой (5.19)

$$\langle f(s, t_0) \rangle \approx f_{in} e^{-t_0 s} \quad (5.22)$$

Решение уравнения (5.21) с этим начальным условием имеет вид

$$\langle f(s, t) \rangle = f_{in} e^{-t_0 s - \mu N \varphi(s)(t-t_0)} \approx f_{in} e^{-\beta(t)s} \quad (5.23)$$

$$\beta(t) = t_0 + \mu N \varphi'(0)(t - t_0) \quad (5.24)$$

Аргумент экспоненты в формуле (5.23) заменен его линейным разложением по s , так как времена эволюции t и $\beta(t)$ велики, а типичное s , соответственно, мало. Вероятность фиксации $\varphi(s)$ уже была найдена в работе [16] (§ 3.2), откуда

$$\varphi'(0) = \frac{x_c}{2Nv}$$

где скорость адаптации v и лидер x_c выражаются через N , частоту полезных мутаций на геном за поколение, U_b , и начальную плотность распределения s , равную внутренней плотности $g(s)$.

В качестве конкретного примера $g(s)$, можно рассмотреть половинку нормального распределения (5.18). При больших размерах популяции, таких что $v \gg s_{av}^2$ или $\log(Ns_{av}) \gg \log^2(s_{av}/U_b)$, получается [16]

$$\varphi'(0) = \frac{1}{2\sqrt{\pi} N s_{av}} \log\left(\frac{s_{av}}{U_b}\right) \quad (5.25)$$

что затем подставляется в уравнение (5.24).

Таким образом, при долговременной адаптации (но до полного равновесия), модель предсказывает что плотность невыгодных аллелей $\langle f(s, t) \rangle$ имеет экспоненциальную форму с наклоном, увеличивающимся во времени, (5.22) и (5.24). Это предсказание объясняет данные на рисунке 5.19, при условии, что эти вирусы и бактерии предварительно эволюционировали в похожих условиях, в течение достаточно долгого времени.

(iii) Квазиравновесный аргумент

Выше экспоненциальная форма DFE была получена для конкретной модели популяции. Тот же результат можно получить и в более общем случае. Как проверено в § 5.1, в режиме бегущей волны $NU_b \gg 1$, когда адаптация медленная, распределение аллелей находится в состоянии квазиравновесия, где энтропия S максимальна и является функцией текущей приспособленности

W [24]. Вероятность вредного аллеля на локусе определяется первой строчкой в формуле (5.5), которую можно переписать в виде

$$\langle f(s, t) \rangle = [1 - \langle f(s, t) \rangle] e^{-\beta(t)s} \quad (5.26)$$

где $\beta \equiv \partial S / \partial W$. Формула (5.26) обобщает результаты (5.19), (5.20), (5.24) и (5.25), которые содержат выражения для конкретных модели популяции и временного интервала. Обратите внимание, что $\beta(t)$ во всех этих выражениях зависит от времени (линейно), поскольку предполагается, что адаптирующаяся популяция все еще далека от истинного стационарного состояния, так что вредные мутационные события не учитываются.

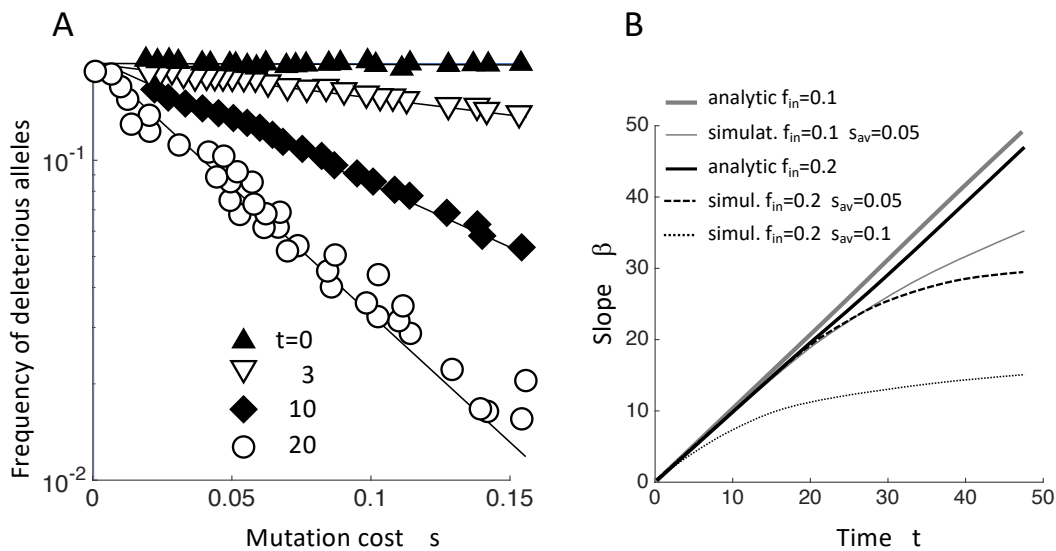


Рисунок 5.21: Доля вредных аллелей затухает экспоненциально с их эффектом приспособленности, с увеличением наклона во времени. (A) Аналитический предсказание частоты вредных аллелей (5.19) согласуется с моделированием Монте-Карло. Абсцисса: коэффициент отбора (стоимость мутации) s . Ордината: доля вредных аллелей на таком локусе, $f(s)$, усредненная по 20 случайным запускам моделирования. Прямые линии: линейная интерполяция. Различные символы показывают результаты моделирования в разные моменты времени. Линии — это аналитические предсказания. (B) Наклон распределения вредных аллелей β в зависимости от времени t : краткосрочное предсказание формулы (5.19) (толстые серые или черные линии) и сглаженный результат моделирования (тонкие, пунктирные или пунктирные линии). Логарифмический наклон для смоделированных кривых доли вредных аллелей в (A) получен экспоненциальной подгонкой. Отклонение смоделированного наклона от (5.19) совпадает с установлением бегущего режима (рисунок 5.20F). На длинных временах, предсказание для бегущей волны дается формулой (5.24). Диагональ показывает $\beta(t) = t$. Параметры такие же, как на рисунке 5.20, если не указано иное (f_{in} и s_{av}). По работе [29].

Таким образом, экспоненциальная зависимость наблюдаемого DFE от s интерпретируется тем, что наиболее быстро меняющимся членом DFE является частота вредных аллелей $\langle f(s, t) \rangle$, которая превращается в экспоненту в ходе эволюции. Результат довольно универсален, потому что он устойчив к врожденной плотности $g(s)$, начальным условиям и модели популяции. Компьютерное моделирование подтверждает этот аналитический результат (рисунок 5.21).

5.4.5 Оценка коэффициентов выбора из набора последовательностей

Эти теоретические результаты имеют важное практическое применение, реализованное в работе [31] входящей в диссертацию, а именно, метод оценки коэффициентов отбора на который не влияют вездесущие эффекты сцепления. Измерение производится в единицах $1/\beta$. Предлагаемый метод достаточно нечувствителен к деталям системы, но необходимо, чтобы система удовлетворяла нескольким условиям, вытекающим из предыдущих подразделов. Выполняются ли они, становится самоочевидно в конце процедуры, так что метод является самотестирующим и позволяет также оценить ошибку измерения. Условия такие:

(i) Естественный отбор является направленным и либо постоянным, либо мало изменяющимся в масштабе времени $t \sim 1/s_{av}$.

(ii) Имеется достаточное количество независимых популяций (разные географические местоположения) для усреднения $\langle f(s, t) \rangle$.

(iii) Популяции далеки от стационарного состояния, $\langle f(s, t) \rangle \gg f_{eq}$

Первым шагом процедуры является получение из базы данных онлайн выровненных последовательностей ДНК, РНК или белка в трех временных точках (третья для проверки). Последовательности бинаризуются, как в § 5.3. Частота миноритарных аллелей для каждого локуса, $\langle f_i(t) \rangle$, рассчитывается для всей базы данных, включая все популяции. Вставки и делеции исключаются, потому что они не могут быть описаны одним коэффициентом отбора для каждого локуса, если только они не имеют фиксированной длины и могут рассматриваться как бинарная замена. (В принципе, метод легко обобщается на небинарный вариант.)

Используя (5.26), относительное значение коэффициента отбора s_i на локусе i можно оценить как

$$\beta(t)s_i = -\log \left[\frac{\langle f_i(t) \rangle}{f_{norm}} \right] \quad (5.27)$$

где предполагается $\langle f_i(t) \rangle \ll 1$. Ниже угловые скобки, обозначающие усреднение по ансамблю $\langle \dots \rangle$, опущены.

Наличие дополнительного фактора f_{norm} в (5.27) отсутствующего в (5.26) связано с изменением определения $f_i(t)$. Выше f_i определена как частота менее приспособленного аллеля, а s_i всегда положительное. В реальных геномных образцах, наиболее подходящая последовательность неизвестна, и вместо этого в качестве эталонной последовательности приходится использовать консенсус. Следовательно, антиконсенсусный аллель может быть лучше приспособлен, и в этом случае $s_i < 0$. Локусы с отрицательным значением s_i — это локусы с $f_i > f_{norm}$, поэтому необходим коэффициент нормализации $f_{norm} < 1$.

Левая часть (5.27) представляет собой произведение двух факторов: $\beta(t)$, который зависит только от времени, и s_i , который зависит только от номера локуса, i . Этот факт указывает на

существование фиксированной точки $s_i = 0$, которая определяет нормировочный коэффициент f_{norm} следующим образом:

Для каждого момента времени t , локусы i ранжируются в порядке убывания величины $\beta(t)s_i$ из (5.27), для получения отображения $i \rightarrow m_i$, где m_i — номер локуса после ранжирования. Результат отображается в виде монотонной кривой $s_{rank}(m, t)$.

Надо заметить, что этот способ ранжирования хорош, только если усреднение очень хорошее. Для маленькой выборки популяций, локусы ранжируются один раз, уже после усреднения $\langle f_i(t) \rangle$ по трем временам, применения формулы (5.27), и получения m_i . В этом случае, три кривые $s_{rank}(m, t)$ не будут в точности монотонны.

Затем, находят точку пересечения кривых $s_{rank}(m, t)$, полученных в три разных момента времени t (рисунок 5.22). Точнее, находят центр треугольника который получается из этих кривых. Внутренняя плотность распределения $g(s)$, введенная в § 5.4.1, определяется как $g(s) = [ds/dm]^{-1}$.

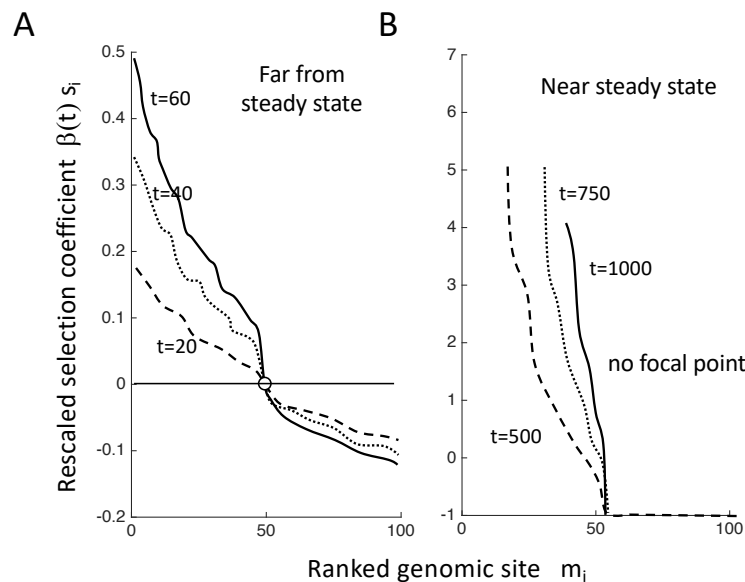


Рисунок 5.22. Ранжированная кривая $s_i(m_i)$ для трех моментов времени. А) Успешное разделение кривых с одной точкой пересечения. В) Отказ метода из-за близости к стационарному состоянию. Параметры: (А) $\langle s \rangle = 0.05$, $L = 100$, $N = 1000$, $t = 60$, $f_0 = 0.1$, 100 прогонов, $\mu L = 0.1$, $t_1 = 20$, $t_2 = 40$, $t_3 = 60$; (В) $t_1 = 500$, $t_2 = 750$, $t_3 = 1000$, $t = 1000$, а остальные параметры как в (А). По работе [31].

Для проверки метода, три кривые должны иметь четкое разделение и единую точку пересечения (рисунок 5.22А), в противном случае одно из трех вышеперечисленных условий не выполняется. Если кривые имеют много пересечений или коллапсируют друг на друга (рисунок 5.22Б), такое поведение свидетельствует либо о недостаточном усреднении по популяциям, либо о стационарном состоянии, либо о слишком близком расположении временных точек, либо о зависящем от времени отборе. В любом из этих случаев, метод не сработает.

Затем, значение f_{norm} в (5.27) подгоняется, чтобы получить $s_{\text{rank}} = 0$ в точке пересечения. Наконец, ранжированные основания переупорядочиваются в обратном порядке, $i \leftarrow m_i$, и βs_i отображаются в виде столбцов разной высоты относительно их фактических оснований, i .

В заключение, метод оценивает коэффициенты отбора в относительных единицах $1/\beta$. Масштабный коэффициент β , одинаковый для всех участков генома, должен быть найден из независимых наблюдений. Програмное обеспечение находится в онлайн приложении к работе [31].

5.4.6 Заключение

В эксперименте по эволюции вируса или бактерии в культуре клеток наблюдается экспоненциальное распределение коэффициентов отбора для полезных мутаций, $DFE(s)$ (рисунок 5.19). Приведенные выше результаты дают простую эволюционную интерпретацию этого наблюдения, а именно то, что более вредные аллели уходят из популяции первыми. На основе этих теоретических результатов, разработан метод оценки коэффициентов отбора на каждом полиморфном локусе на основе наборов геномных последовательностей взятых из большого числа адаптирующихся популяций, в три момента времени. Внутренняя плотность распределения измеряемая в эксперименте по направленному мутагенезу, $g(s)$, сильно отличается от наблюдаемой в эксперименте и может принимать любую форму, в зависимости от вируса и белка.

Выражение (5.26) было впервые получено в физике, для энергетического распределения фермионов в термодинамическом равновесии. Используя аналогичный термодинамический аргумент, оно было получено для различных биологических систем с двумя состояниями [428-430]. Причина, по которой равновесная статистика оказывается справедливой в неравновесном случае, заключается в низкой скорости адаптации [24].

Все результаты главы 4 получены для режима адаптации вдали от стационарного состояния (мутационно-селекционного баланса, § 2.1 и 3.1), когда новыми вредными мутациями можно пренебречь. Вблизи этого состояния, вредная мутация важна, так как она уравнивает эффект естественного отбора, поэтому все выводы данной главы становятся недействительными. В этом случае, экспоненциальная форма получается для другой величины [430], а именно, для отношения плотности полезных мутаций к плотности вредных мутаций $DFE(s)/DFE(-s)$. В отличие от рассмотренного здесь режима адаптации, где экспоненциальный наклон $DFE(s, t)$ увеличивается во времени, экспоненциальный наклон, предсказанный для $DFE(s)/DFE(-s)$ в

стационарном состоянии, является постоянным и равным среднему времени до недавнего общего предка пары геномов [430].

Ранние попытки объяснить вездесущую экспоненциальную форму DFE не использовали никаких эволюционных аргументов и основывались на теории экстремальных значений [157-159]. Их цель состояла в том, чтобы доказать экспоненту для внутреннего распределения коэффициента отбора, измеренного методом мутагенеза, $g(s)$, в пределе больших s . Настоящая теория предсказывает, что DFE является экспоненциальным в широком диапазоне s из-за меняющейся аллельной частоты $\langle f(s, t) \rangle$. Еще одно отличие состоит в том, что указанный подход предсказывает, что $DFE(s)$ постоянен во времени, в то время как настоящий анализ, в соответствии с экспериментом, демонстрирует, что наклон β не является постоянным и изменяется во времени (см. рисунок 5.19 и 5.21). Третье отличие состоит в том, что экспонента предсказывается для всех значений s , тогда как авторы [157-159] предсказывают ее только в пределе больших s .

Таким образом, экспоненциальная форма DFE, наблюдаемая у многих бактерий и вирусов, возникает автоматически в процессе адаптации. В дополнение к оценке эпистаза из геномных данных в §§ 5.1-5.3, данная теория предлагает общий метод измерения всего набора коэффициентов отбора из геномных данных, что и завершает главу о инференции адаптивного ландшафта.

Глава 6. Эволюционная роль признака

Ничто в биологии не имеет смысла, кроме как в свете эволюции

Феодосий Добжанский

К сожалению, далеко не все в биологии имеет смысл даже в свете эволюции. Свойства организмов развились в ходе длительного процесса, когда естественный отбор действовал на фоне стохастических факторов, таких как мутация, генетический дрейф, и рекомбинация [3]. Следовательно, многие признаки возникли как случайности на пути эволюционной истории и не дают организму селективного преимущества. Иногда эволюционное преимущество данного свойства существует но заранее неизвестно, и необходимы серьезные исследования, чтобы выяснить, какую пользу организму или вирусу оно может принести. В этой главе исследуются три примера, когда признак, казавшийся случайным и биологически бессмысленным, на самом деле дает важное преимущество, обнаруженное после исследования.

6.1 СПИД, как следствие адаптации ВИЧ к хозяину

У 99.5% лиц, инфицированных ВИЧ и не принимающих лекарств (к счастью, таких уже мало), развивается синдром приобретенного иммунодефицита, связанный с истощением CD4 Т-клеток, что приводит к появлению симптомов СПИДа в течение 2–20 лет. В то же время, как видно из ряда фактов, инфицирования CD4 клеток недостаточно для прогрессирования к СПИД. ВИЧ-инфекция представляет собой устойчивое стационарное состояние с быстрым оборотом и заменой всех соответствующих типов клеток. Это стационарное состояние почти точное, но не совсем, так как число CD4 клеток медленно снижается. Существование стационарного состояния предсказываются простыми математическими моделями типа «охотник-жертва» [181, 190, 191]. Однако все эти модели предсказывают только идеальное стационарное состояние. Причины медленного снижения количества CD4 клеток на временных масштабах, на два и три порядка превышающих время оборота всех типов клеток, остаются неясными.

Этот раздел посвящен работе [202], которая объясняет причину постепенного снижения числа CD4 клеток в ВИЧ инфекции из адаптации вируса к индивидуальному хозяину. В основе работы объединение моделей эволюции в главах 2 и 3 с очень простыми моделями иммунного ответа. Альтернативная (и весьма популярная) гипотеза постулирует, что причиной постепенного истощения CD4 клеток является повреждение гомеостатических механизмов [431].

Ниже это альтернативная идея воплощена в виде двух моделей, и предсказания всех моделей сравниваются с данными и между собой (рисунок 6.1). Предложены новые экспериментальные проверки и показано, как эти теоретические результаты позволяют сделать прогноз СПИДа для индивидуальных пациентов, которые либо не лечатся, либо находятся на субоптимальном режиме терапии.

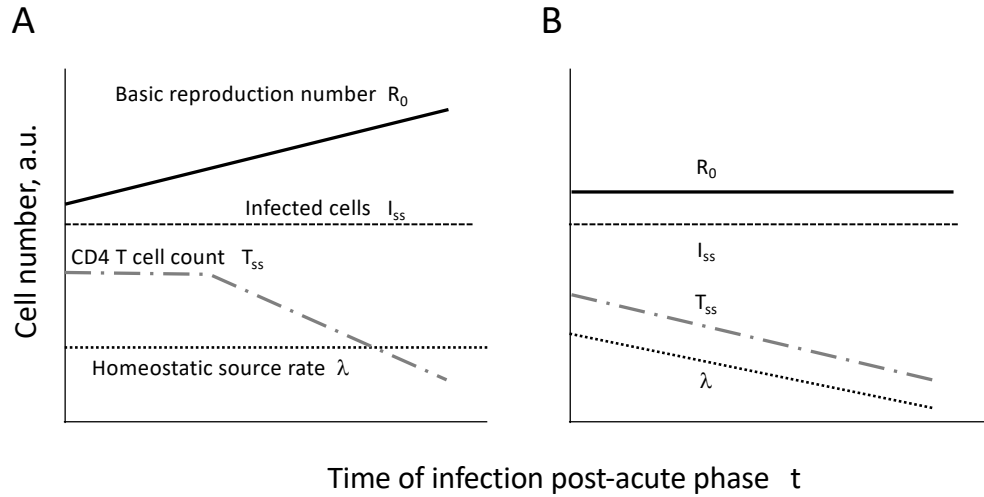


Рисунок 6.1: Две альтернативные гипотезы о природе постепенного снижения числа CD4 Т-лимфоцитов, вызывающего СПИД. Штрихпунктирная линия показывает квазистационарный уровень CD4 Т-клеток T_{ss} , предсказанный формулой (6.9) если (А) базовый коэффициент воспроизводства R_0 изменяется вследствие постепенной вирусной адаптации, (В) R_0 остается постоянным, но гомеостатический источник CD4 Т-клеток λ со временем ослабевает. По работе [202].

6.1.1 Модель вирусной динамики

Модель 1: отсутствие иммунного ответа

Отправной точкой являются две простейшие модели вирусной динамики, предсказывающие точное устойчивое состояние, с иммунным ответом и без него. Эволюция вируса будет включена в модель попозже. У младенцев в возрасте до 1 года, клеточный ответ Т-клеток является незрелым, и им можно пренебречь. Простейшая модель вирусной динамики включает в себя неинфицированные клетки-мишени и инфицированные клетки [180, 191]

$$\frac{dT}{dt} = \lambda - (d_T + kI)T \quad (6.1)$$

$$\frac{dI}{dt} = kIT - d_I I \quad (6.2)$$

где T — количество неинфицированных клеток-мишеней с фенотипом CD4+ CCR5+ Т, I — количество инфицированных клеток, λ — постоянная скорость гомеостатического восполнения клеток-мишеней из предшественников, d_T — обратная продолжительность жизни клеток-мишеней, d_I — скорость гибели инфицированных клеток, k — параметр инфекционности. Динамика вирионов явным образом не включена, потому что вирионы очень недолговечны в

теле, около 6 часов [432], и их количество в любое время пропорционально количеству инфицированных клеток.

Модель предсказывает точное устойчивое состояние, которое соответствует хронической, клинически латентной фазе ВИЧ-инфекции [180-183, 372]. Уравнения (6.1) и (6.2) имеют устойчивое стационарное состояние где количество клеток каждого типа не меняется со временем

$$T_{ss} = \frac{T_0}{R_0} \quad (6.3)$$

$$I_{ss} = \frac{d_T}{k}(R_0 - 1) \quad (6.4)$$

Здесь T_0 обозначает число клеток CD4 у неинфицированных пациентов ($I = 0$), а R_0 представляет собой базовое репродуктивное число, определяемое формулой

$$T_0 \equiv \frac{\lambda}{d_T}, \quad R_0 \equiv \frac{k\lambda}{d_I d_T} > 1 \quad (6.5)$$

Чтобы произошло заражение, R_0 должен быть больше 1.

Модель 2: CD8 T-клеточный иммунный ответ

Простейшая модель с цитотоксическим иммунным ответом имеет вид [190, 191]

$$\frac{dT}{dt} = \lambda - (d_T + kI)T \quad (6.6)$$

$$\frac{dI}{dt} = kIT - d_I I \left(1 + \frac{E}{E_0}\right) \quad (6.7)$$

$$\frac{dE}{dt} = aE \frac{I}{I + I_0} - d_E E \quad (6.8)$$

Новое уравнение (6.8) описывает динамику эффекторных CD8 T клеток-убийц, где a — максимальная скорость пролиферации, I_0 — характерное число инфицированных клеток, при котором достигается полумаксимальная экспансия иммунных клеток, d_E — смертность эффекторных клеток. По сравнению с (6.2), в уравнении (6.7) появился дополнительный член, описывающий уничтожение инфицированных клеток эффекторными клетками, E . Новый параметр модели E_0 представляет собой число эффекторных клеток, которое увеличивает скорость гибели клеток в два раза. Полагая правые части уравнений (6.6)-(6.8) нулю, получаются стационарные значения

$$\begin{aligned}
T_{ss} &= \frac{T_0}{1 + \frac{R_0}{R^*}}, R^* \equiv \frac{T_0 d_T}{I_{ss} d_I} \gg 1 \\
I_{ss} &= \frac{I_0 d_E}{a - d_E} \\
E_{ss} &= E_0 \left[\frac{R_0}{1 + \frac{k d_E I_0}{d_T (a - d_E)}} - 1 \right]
\end{aligned} \tag{6.9}$$

с T_0 и R_0 определенными в формуле (6.5).

Эта модель не включает латентно-инфицированные клетки. В слизистой фазе инфекции (§ 6.2), у контроллеров ВИЧ и при антиретровирусной терапии, латентные клетки имеют решающее значение [19, 224, 432]. Модель также не включает ответ хелперных Т-клеток, что очень важно для пациентов с низкой вирусемией и вирусных контроллеров [174, 274, 433]. Как показано в работах автора по иммунологии не вошедших в данную диссертацию, гонка вооружений между хелперными Т-клетками и вирусом может быть причиной персистенции ВИЧ и его успешного контроля в некоторых пациентах [223, 224, 375, 434, 435]. Поскольку основное внимание уделяется прогрессированию СПИДа у пациентов с высокой или умеренно низкой вирусемией, ответом хелперных клеток и латентными клетками можно пренебречь.

6.1.2 Скорость адаптации ВИЧ к хозяину

Следующая задача состоит в том, чтобы рассчитать скорость адаптации вируса, а затем связать ее со снижением количества CD4 Т-лимфоцитов (§ 6.1.2). Скорость адаптации определяется как скорость, с которой базовое число репродукции вируса R_0 изменяется в единицу времени, определяемое формулой $V = d \log R_0 / dt$. Величина R_0 , представляющее собой среднее число вторично инфицированных клеток на одну зараженную клетку, эквивалентно приспособленности вируса. Его значение выражается через параметры модели вирусной динамики, формула (6.5).

Скорость адаптации V , в общем виде, может быть выражена через параметры эволюционной модели (главы 1 и 2). Это выражение зависит от главных факторов эволюции [7, 11, 12, 14, 16, 72-74, 76]. В сценарии бесполой многолокусной эволюции когда ВИЧ далек от эволюционного равновесия, вредными мутациями можно пренебречь, а рекомбинация отсутствует (§ 3.1). При этих предположениях, верных для пациентов с низкой вирусной нагрузкой $< 10^5$ копий РНК на мл плазмы [15], скорость адаптации определяется выражением [11]

$$V = 2s^2 \frac{\log(I_{SS}\sqrt{sU_b})}{\log^2 \frac{s \log I_{SS}}{U_b}} \quad (6.10)$$

где U_b — частота полезных мутаций на геном на поколение инфицированных клеток (1 день), s — средний коэффициент отбора, т.е., увеличение $\log R_0$ на мутацию, а I_{SS} — число инфицированных клеток при хронической инфекции. Число I_{SS} определяется формулой (6.4) or (6.9), в зависимости от иммунокомпетентности больного на момент заражения (т.е. возраста).

Аргументы обоих логарифмов в (6.10) должны быть намного больше 1, что означает, что I_{SS} не может быть слишком маленьким, $I_{SS}\sqrt{sU_b} \gg 1$. Это условие может не выполняться у «элитных контроллеров», редких пациентов с аномально низкой вирусной нагрузкой, у которых полезные мутации возникают на фиксацию редко и не перекрываются во времени. В этом случае, (6.10) заменяется на более простую формулу $V = sI_{SS}U_b$ (§ 3.1). Тем не менее, даже в контроллерах ВИЧ, значение I_{SS} редко бывает ниже 100-1000 клеток, while $s \sim 0.01$, $U_b = 0.01 - 0.1$ [6], так что формула (6.10) почти всегда справедлива.

Пока предполагалось, что значение коэффициента отбора s одинаково для всех локусов. В реальной жизни, оно варьируется в зависимости от локуса, согласно плотности распределения. Например, экспоненциальное распределение часто наблюдается в эксперименте [16] (§ 6.4). При достаточно больших значениях I_{SS} , этот случай сводится к фиксированному коэффициенту отбора, используя эффективные параметры [16], которые проще оценить напрямую из образцов вирусной РНК в пациенте (§§ 2.1 и 2.3).

У пациентов с высокой вирусемией, $> 10^5$ копий РНК/мл плазмы, рекомбинация ВИЧ происходит достаточно часто, чтобы ускорить адаптацию в несколько раз [8, 15], и для оценки частоты замен V необходимо использовать модели, включающие рекомбинацию (глава 4). Эффективная вероятность рекомбинации на геном r зависит от количества встроенных вирусов на клетку и может быть также оценена из образцов вирусной РНК в пациенте (§ 2.3). Заметим, что долгосрочная адаптация ВИЧ в таких пациентах, скорее всего, происходит благодаря совместному действию мутации и рекомбинации [76]. Выражение для скорости адаптации в этом случае было получено другими группами [74, 75, 396].

6.1.3 Прогнозирование скорости прогрессирования СПИДа

Стационарное количество CD4 Т-клеток T_{SS} , определяемое, в зависимости от иммунокомпетентности человека, из формулы (6.3) или (6.9), уменьшается с базовым числом репродукции R_0 , которое медленно увеличивается во времени в процессе адаптации к организму больного [6] (§ 2.1). Важной особенностью является разделение масштабов времени между

генерической адаптацией, которая занимает годы, и временем оборота клеток, дни или недели [180]. Поэтому, при очень медленном увеличении параметра R_0 , система остается все время в динамически-равновесном состоянии, а T_{ss} очень медленно снижается в течение инфекции, пока из-за сильного обеднения клеток не появятся щели в их антигенном репертуаре и не наступит СПИД.

Средний исходный уровень CD4 Т-клеток при хронической ВИЧ-инфекции составляет $T_{ss}(0) = 500$ клеток/мкл крови. Ранние симптомы СПИДа начинаются, когда количество клеток CD4 падает ниже $T_{ss}(t_{AIDS}) = 200$ клеток/мкл крови. Время до появления симптомов СПИДа можно рассчитать следующим образом.

$$t_{AIDS} = \frac{\log R_0(t_{AIDS}) - \log R_0(0)}{V}$$

$$V \equiv \frac{d \log R_0}{dt} \quad (6.11)$$

в котором предполагается, что скорость адаптации V в формуле (6.10) или аналогичной формуле постоянна во времени (на самом деле, она меняется, но логарифмически медленно). При отсутствии Т-клеточного иммунного ответа у младенцев, из (6.3) и (6.11) получаем

$$t_{AIDS} = \frac{\psi}{V}, \quad \psi \equiv \frac{\log(T_{ss}(0))}{T_{ss}(t_{AIDS})}$$

Обратите внимание, что ψ в среднем близок к 1, но может варьироваться у разных пациентов. У взрослых людей с Т-клеточным иммунным ответом, согласно формулам (6.9) и (6.11), время до СПИДа есть

$$t_{AIDS} = \frac{1}{V} \left\{ \psi + \log \left[1 + \frac{R^*}{R_0(0)} (1 - e^{-\psi}) \right] \right\} \quad (6.12)$$

что в несколько раз дольше, чем при отсутствии иммунного ответа, $t_{AIDS} = \psi/V$. Это предсказание модели согласуется с тем наблюдением, что среднее время до СПИДа у нелеченых взрослых дольше (10 лет), чем у младенцев (от 1 до 2 лет).

Модель объясняет отрицательную корреляцию между временем до СПИДа и вирусной нагрузкой

Простым тестом предсказания (6.12) является отрицательная корреляция времени до СПИДа и вирусной нагрузкой v наблюдаемой в пациентах [436]. Измеренная вирусная нагрузка v пропорциональна общему количеству инфицированных клеток, $I_{ss} = v\theta$. Прогнозируемое время до СПИДа у взрослых t_{AIDS} , формула (6.12), можно написать как функцию вирусной нагрузки v

$$t_{AIDS} = \frac{A}{\log(Bv)} \left[1 + \log \left(1 + \frac{C}{v} \right) \right] \quad (6.13)$$

где V и R^* выражены через I_{ss} из (6.10) и (6.9), соответственно, и введены новые обозначения

$$A \equiv \frac{1}{s^2 \frac{\log^2(s)}{U_b}}, \quad B \equiv \theta v \sqrt{s U_b}, \quad C \equiv \frac{T_0 d_T}{(\theta d_I R_0(0))(1 - e^{-1})}$$

Используя A и B как подгоночные параметры, формула (6.13) была подогнана к данным наблюдений в пациентах [436] (рисунок 6.2). Параметр C для подгонки не потребовался и был зафиксирован на произвольном значении выше 10^6 копий РНК/мл. Подгонка использовала метода наименьших квадратов и программу `fminsearch.m` в MATLAB™. Наблюдается хорошее соответствие предсказаний модели данным (рисунок 6.2). (Экспериментальные оценки параметров модели из таблицы 6.1, необходимые для индивидуального прогноза пациента, описанного ниже, при данной подгонке не использовались.)

Основная гипотеза о том, что прогрессирование к СПИДу обусловлено адаптацией вируса, также согласуется с положительной корреляцией между прогрессированием и уровнем иммунной активации, наблюдаемой как у людей, так и у обезьян [186-189, 436]. Действительно, адаптация ВИЧ как раз и вызвана ускользанием от иммунного ответа в эпитопах и последующими компенсаторными мутациями (§ 2.1 и § 7.2).

Таблица 6.1. Параметры модели важные в квазистационарном состоянии.

Обозначение	Наименование	Единицы измерения	Диапазон значений
ψ	Уменьшение логарифма количества CD4 Т-клеток при начале СПИДа	1	~ 1
I_{ss}	Число зараженных клеток	клетка/тело	$100 - 3 \cdot 10^9$
L	Общее количество оснований	1	$10 - 1000$
μ	Вероятность мутации	1/поколение/основание	$3 \cdot 10^{-5}$ $3 \cdot 10^{-6} - 10^{-4}$
$U_b = \mu L$	Вероятность мутации на геном	1/поколение/геном	$3 \cdot 10^{-4} - 0.03$
b	Эффективный коэффициент отбора	1	$0.1 - 0.005$
r	Вероятность рекомбинации	1	$0 - 1$
T_0	Нормальное число клеток-мишеней	клетка/мм ³	$700 - 1400$
d_T	Их обратная продолжительность жизни	1/день	$0.1 - 0.01$
d_I	Обратная продолжительность жизни зараженных клеток	1/день	$0.7 - 1.2$
$R_0(0)$	Базовое репродуктивное число	1	$5 - 11$

$R^* = \frac{T_0 d_T}{I_{ss} d_I}$	Составной параметр	1	
------------------------------------	--------------------	---	--

Оценки параметров для прогноза СПИДа в индивидуальных пациентах

В дополнение к объяснению корреляции между вирусемией и прогрессированием до СПИДа, эта модель может быть использована для прогнозирования времени до СПИДа у отдельных пациентов. Параметры модели в формуле (6.12) (Таблица 6.1) можно получить из вирусологических данных отдельных пациентов, которые измеряются непосредственно или выводятся, как показано ниже.

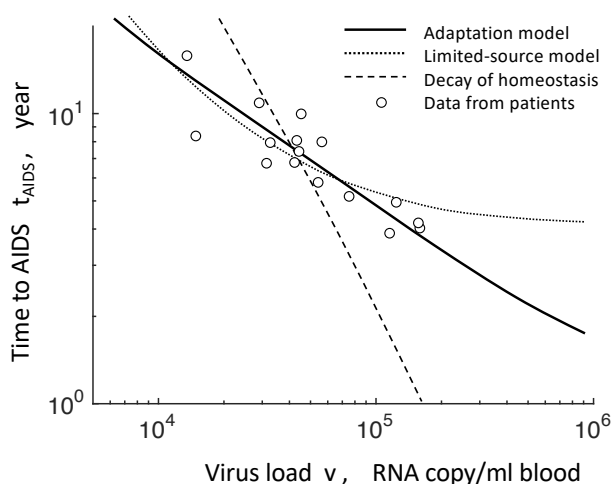


Рисунок 6.2. Время до СПИДа, предсказанное тремя моделями с двумя составными параметрами подбора, подогнанное к данным по ВИЧ+ мужчинам, сероконверсия которых произошла во время исследования [436]. Из 16 пациентов, 11 не получали лечения, а 5 получали неоптимальную антиретровирусную терапию в той или иной форме (не тройные коктейли); на результат это не повлияло. Кружки: время выживания в зависимости от вирусной нагрузки в данных пациентов. Сплошная толстая кривая: наилучшее предсказание модели адаптации (6.13). Пунктирная кривая: предсказание модели ограниченного гомеостатического источника (6.15). Пунктирная линия: предсказание модели распада гомеостатической среды вызванного вирусом (6.16). По работе [202].

Коэффициент ψ находится из уравнения (6.12), где $T_{ss}(0)$ — количество клеток CD4 у больного в начальной стадии хронической инфекции. Количество инфицированных клеток, I_{ss} , можно оценить по вирусемии в копиях РНК/мл крови, измеренной с помощью количественной ПЦР, с использованием конвертера единиц $\theta = 10^4$ клетка·мл/копия РНК [55]. Частота мутаций $U_b = \mu L$, где частота мутаций на основании μ измерялась для каждой пары замен со средним значением $\mu \approx 3 \times 10^{-5}$ [56], а L — количество полиморфных оснований (SNP), которое можно подсчитать непосредственно в вирусе выделенном из данного больного.

Средний коэффициент отбора s и вероятность рекомбинации r , необходимые для пациентов с более высокой вирусемией, могут быть оценены по геномным последовательностям пациентов различными методами [6, 15, 77, 154, 155]. Первые оценки среднего эффективного s у нелеченых пациентов была получена в работе автора [6] и группой [77]. Оценка этих

параметров с помощью многолокусного моделирования [15] (§ 2.3) является наиболее точным, поскольку она учитывает эффекты сцепления, рекомбинацию и любые желаемые факторы. Входные параметры T_0 , d_T , d_I и $R_0(0)$ могут быть заменены хорошо известными средними значениями для пациентов [375, 435], так как их вариация в наблюдаемых пределах мало влияет на результат.

Таким образом, все параметры модели могут быть либо точно определены для индивидуума, либо не являются чувствительными и могут быть заменены средними по популяции. Классификация пациентов и отбор проб описаны в рецензируемой работе [202].

6.1.4 Альтернативные модели развития СПИДа: нарушение гомеостаза

Альтернативным объяснением прогрессирования СПИДа, является дерегуляция гомеостаза [431]. Автор разработал две модели, основанные на этой альтернативной гипотезе и сравнил их результаты с данными.

Выше линейная скорость пополнения клеток λ предполагалась фиксированной. Фактически гомеостатический механизм отключает пополнение, когда T увеличивается до порога, например, так:

$$\lambda(T) = \lambda_0 \left(1 - \frac{T}{T_0}\right) \quad (6.14)$$

где T_0 и λ_0 — два независимых постоянных параметра.

Первый способ ввести нарушение гомеостаза— постулировать, что максимальное число делений конечно из-за теломеров. В этом случае, общее производство клеток-мишеней имеет конечный предел Λ

$$\lambda(T_{ss})t_{AIDS} = \Lambda$$

Подставляя уравнение (6.14) в (6.6) с $dT/dt = 0$, из последнего уравнения получаем

$$t_{AIDS} = \frac{\Lambda (\lambda_0/T_0) + kI_{ss}}{\lambda_0} \equiv a \frac{b + I_{ss}}{I_{ss}} \quad (6.15)$$

Здесь a, b —параметры используемые для подгонки к данным в пациентах (рисунок 6.2). В уравнении (6.15), предполагается сильное неравенство $kI_{ss} \gg d_T$, что эквивалентно $R_0 \gg R^*$. Это неравенство выполняется для среднего нелеченого инфицированного. Величина kI_{ss}/d_T представляет собой увеличение скорости оборота CD4 Т-клеток из-за ВИЧ-инфекции, диапазон которого 3 – 6 [180]. Наилучшая подгонка формулы (6.15) к данным [436] дает изогнутую кривую в полулогарифмическом масштабе, включающую крутой спад, переходящий в плато (рисунок 6.2). Такой кривизны в данных не видно из-за разброса, хотя и полностью исключить ее нельзя из-за короткого интервала времени.

Но главная проблема с этой моделью конечных делений, это отсутствие в ней постепенного спада числа CD4 Т-клеток в течение нескольких лет наблюдаемого в данных. По этой модели, стационарное состояние просто резко обрывается и заменяется СПИДом.

Чтобы решить эту проблему, вместо ограниченного производства клеток-мишеней, прогрессирование СПИДа также может быть вызвано постепенной деградацией гомеостатической среды, из-за накопления вирусных продуктов [431]. Можно постулировать, что вирусные продукты постепенно разрушают максимальную интенсивность гомеостатического источника, λ_0

$$\frac{d\lambda_0}{dt} = -\alpha I_{SS} \lambda_0(t)$$

Решая это простое уравнение, можно получить экспоненциальный спад во времени для интенсивности источника

$$\lambda_0(t) = \lambda_0(0)e^{-\alpha I_{SS} t}$$

что вызывает постепенное снижение T_{SS} , согласно формуле (6.9).

Как уже упоминалось, начало СПИДа происходит, когда количество CD4 падает до характерного значения.

$$T_{SS}(t_{AIDS}) = T_{AIDS}$$

Подставляя $\lambda_0(t)$ вместо λ в (6.5) и (6.9), получаем

$$t_{AIDS} = \frac{1}{\alpha I_{SS}} \log \left(\frac{\lambda_0(0)}{k I_{SS} T_{AIDS}} \right) \equiv \frac{1}{\alpha I_{SS}} \log \left(\frac{\beta}{I_{SS}} \right) \quad (6.16)$$

где по-прежнему предполагается $k I_{SS} \gg d_T$, а новые составные параметры α и β не зависят от I_{SS} и используются для подгонки данных. Подогнанная формула (6.16) для t_{AIDS} слишком резко уменьшается с вирусной нагрузкой по сравнению с фактической корреляцией (рисунок 6.2). Это резкое несоответствие полностью исключает модель «деградации источника».

Таким образом, две модели гомеостатических нарушений, как со счетчиком делений, так и с вирус-индуцированной деградацией, не могут обеспечить соответствие данным с качеством модели адаптации, несмотря на одинаковое количество подгоночных параметров (два) (рисунок 6.2). Причиной хорошего согласия модели адаптации с данными является медленная логарифмическая зависимость скорости эволюции от вирусной нагрузки. Важно отметить, что эта особенность является универсальным свойством большого класса моделей эволюции с множественными эволюционирующими локусами [3] и независима от конкретного выбора модели иммунного ответа. Таким образом, предлагаемое объяснение не только хорошо согласуется с данными, но и устойчиво к вариациям модели.

6.1.5 Заключение

Прогрессирование к СПИДу можно объяснить эволюцией ВИЧ, который со временем становится более приспособленным к хозяину и, тем самым, снижает количество клеток-мишеней находящиеся в стационарном состоянии. Модель гораздо лучше соответствует данным наблюдений, чем широко распространенная гипотеза о нарушении гомеостаза. Таким образом, хотя гомеостаз у ВИЧ-инфицированных, возможно, и нарушен, это не является фактором, вызывающим прогрессирование к СПИДу. Также предложен метод прогнозирования времени до СПИДа по нескольким измеряемым параметрам пациента.

Эта система представляет собой пример, когда адаптация вируса к хозяину приводит к увеличению его вирулентности. Другим таким примером является эволюция полиовируса в организме хозяина [22] рассмотренная в § 6.3. В общем случае, вирусы могут эволюционировать по сложным траекториям в плоскости «приспособленность-вирулентность» [23], и направление эволюции может различаться между биологическими уровнями (§ 7.4).

6.2 Эволюционная роль латентного состояния ВИЧ

Большинство ВИЧ-инфицированных клеток продуцируют большое количество вируса и погибают примерно через день после заражения. Однако вирус, интегрированный в ДНК человека, также может находиться в состоянии покоя, называемом «латентным», когда он не экспрессирует мРНК и белки вируса [203, 204]. Размер латентной клеточной популяции относительно невелик, около одной клетки на миллион CD4 T-клеток [205]. Однако латентный резервуар важен, потому что он заставляет вирус сохраняться во время длительной антиретровирусной терапии (АРТ). Следовательно, латентность является основным препятствием для лечения ВИЧ [206].

В то время как латентность помогает ВИЧ сохраниться во время терапии, его эволюционная польза в течение миллионов лет естественной эволюции лентивирусов в отсутствие АРТ остается неясной. Преобладающая гипотеза ранее состояла в том, что латентность — это случайность эволюции, которая произошла несколько миллионов лет назад и передавалась по филогенетическому древу. Возможное преимущество латентности в приспособленности заключается в том, что она может повысить шансы вируса на выживание в суровых условиях окружающей среды. Существует фаза инфекции, когда вирусная нагрузка всегда низкая: начальная инфекция слизистой оболочки. Системные инфекции устанавливаются, обычно, из единственной РНК генома вируса-основателя в переданной дозе (инокуляте) [228, 229]. Эксперименты на приматах свидетельствуют о том, что слизистая оболочка изначально

неблагоприятна для роста ВИО. Большие инокуляции $\sim 10^9$ инфекционных единиц вымирают в течение 5 дней [232]. Каждая инфицированная клетка живет около 1 дня [234], так что число активно инфицированных клеток через 5 дней пропорционально $(R_0^{muc})^5$, где R_0^{muc} здесь обозначает базовое число репродукции в слизистой оболочке. Поскольку количество активно инфицированных клеток падает ниже уровня обнаружения в течение 5 дней [232], это означает, что $R_0^{muc} < 0.1$.

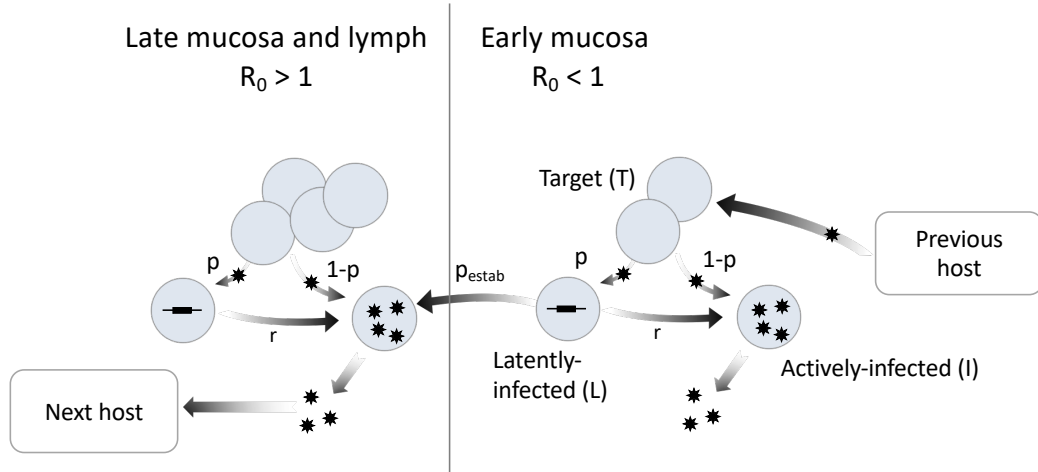


Рисунок 6.3. Латентность ВИЧ — это стратегия троянского коня для оптимизации передачи вируса. Схема двухкамерной модели. Справа налево: ВИЧ проникает в хозяина через слизистую оболочку, но из-за небольшого количества перmissive клеток-мишеней в ранней фазе инфекции слизистой оболочки ($R_0 < 1$). Чтобы успешно установить системную инфекцию, вирус должен сохраниться до момента, когда $R_0 > 1$, ожидая в латентно инфицированных клетках. По работе [19].

В работе [19] рассмотренной ниже, исследована гипотеза о том, что латентность обеспечивает преимущество в приспособленности на стадии передачи между людьми или животными (рисунок 6.3). Идея состоит в том, что латентно инфицированные клетки живут долго и могут активироваться, как только условия в слизистой изменятся на $R_0 > 1$, что начнет системную экспансию вируса. В то же время, увеличение p_{lat} снижает вирусную нагрузку в фазе системной инфекции. Этот эффект снижает количество вирусов, передаваемых новым хозяевам, и, следовательно, вероятность передачи. Баланс между двумя противоположными эффектами приводит к оптимальной вероятности латентности.

6.2.1 Математические модели передачи ВИЧ

Чтобы количественно оценить полное влияние латентности на передачу лентивирусов, ниже проанализированы три математические модели (рисунок 6.4). Каждая модель включает раннюю инфекцию слизистой оболочки с $R_0^{muc} < 1$. Первая модель описывает инфекцию только в слизистой оболочке. Учитывая небольшое количество клеток, инфицированных передаваемым вирусом, инфекция слизистой оболочки, по существу, является стохастическим процессом [437], для описания которого необходимо использовать моделирование Монте-Карло или теорию

вероятностей. Ниже анализируются как стохастическая модель, так и детерминистическое приближение, чтобы найти зависимость вероятности выживания вируса в слизистой оболочке от вероятности латентности, p_{lat} .

Model	Question	Method	Result
Early Mucosa Single compartment Fig 3.3 right	Will early mucosal infections go extinct without latency?	ODE, Gillespie simulation	Latency impedes extinction (Fig. 3.5A, 3.6B)
	What level of latency is required to impede extinction?	Branching process & Wright-Fisher simulation	$p_{lat} > 0.005$ (Fig. 3.6C, D)
Systemic Two compartments Fig. 3.3	How does latency impact establishment p_{estab} ?	ODE, Wright-Fisher simulation	Latency increases p_{estab} (Figs. 3.5 and 3.6)
	How does latency impact viral load I_0 ?	ODE	Latency decreases I_0 (Fig. 3.5B)
	What level of latency optimizes p_{trans} ?	ODE	Optimal $p_{lat} \sim 0.5$ (Fig. 3.5C)
Immune Fig. 3.3 & 3.7	Can the immune model lower $p_{lat}(t)$ in patients and fit other patient data?	ODE	$p_{lat}(t)$ is down by many orders. Data are fit. Optimal $p_{lat}(0) \sim 0.5$ (Fig. 3.8B)

Рисунок 6.4: Блок-схема моделей. По работе [19].

Вторая модель представляет собой двухкомпонентную модель, которая описывает как инфекцию слизистой оболочки, так и системную инфекцию лимфоидной ткани [рисунок 6.3 и уравнения (6.24)]. Чтобы начать системную инфекцию, вирус должен проникнуть в лимфоидную ткань, богатую CD4+ Т-клетками, где находятся 98% CD4+ Т-клеток [355]. Модели слизистой оболочки и системной инфекции идентичны и отличаются только значением одного параметра, R_0 (рисунок 6.3 и табл. 6.2). Как уже упоминалось, R_0^{muc} намного меньше 1 при начальном заражении [232], а при системном заражении $R_0^{LT} \sim 10$ [438]. Динамику в двух компартментах можно анализировать в каждом из них отдельно. Как показано ниже, данная модель предсказывает оптимальный уровень латентной вероятности, $p_{lat}^{opt} \approx 0.5$, что согласуется с результатами экспериментов на клеточной культуре [213, 214] и слизистой обезьян [210]. Однако частоты латентности, наблюдаемые у хронически инфицированных пациентов, куда более низкие [176, 439]. Чтобы учесть это несоответствие, к модели системной инфекции, добавлен иммунный ответ хозяина, уравнения (6.24). Действительно, наличие иммунного ответа является ключевым отличием хронической инфекции в пациентах от экспериментов на клеточных культурах. Все параметры модели либо известны из литературы, либо подгоняются к данным для пациентов (табл. 6.3). В конце концов, такая расширенная модель предсказывает то же самое оптимальное значение до начала иммунного отклика p_{lat} , что и базовая модель.

Исходная модель представляет собой простейшую модель вирусной динамики [191] расширенную за счет включения латентных инфекций [440-443]. В отсутствие иммунного ответа, который мы включим позже, и для больших клеточных популяций, модель описывается обыкновенными дифференциальными уравнениями

$$\begin{aligned}\frac{dT}{dt} &= b - d_T T - kVT \\ \frac{dI}{dt} &= (1 - p_{lat})kVT - d_I I \\ \frac{dL}{dt} &= p_{lat}kVT - d_L L - rL \\ \frac{dV}{dt} &= nd_I I - cV\end{aligned}\quad (6.17)$$

Здесь неинфицированные клетки-мишени T производятся с линейной скоростью b , погибают с частотой d_T и могут быть инфицированы вирусными частицами V с эффективностью k . Зараженные клетки имеют два возможных пути. С вероятностью p_{lat} , они становятся латентно инфицированными L , а с вероятностью $1 - p_{lat}$, они становятся активно инфицированными, I . Латентно инфицированные клетки либо активируются с частотой r , либо погибают с малой частотой d_L . Активно инфицированные клетки производят по n вирусных частиц каждая и погибают с частотой d_I . Вирусные частицы теряют инфекционность или выводятся из циркуляции с частотой c , которая считается очень быстрой. Все значения параметров приведены в таблице 6.2. Параметры расширенной модели, включая иммунный ответ, приведены в таблице 6.3. Уравнения (6.17) решались численно с помощью программы ode15s.m в пакете MATLAB™.

Таблица 6.2: Параметры базовой модели

Обозначение	Определение	Единицы измерения	Значение	Ссылки
R_0^{LT}	Базовое число репродукции при системной инфекции	1	10 на рисунке 6.5 15* на рисунках 6.8 и 6.9	[438]
R_0^{muc}	Число репродукции при инфекции слизистой	1	0.1 на рисунках 6.5A, 6.6A, B 0.25 на рисунке 6.6F	[232, 444]
b_{muc}	Скорость восполнения зараженных клеток в слизистой	клетка/день	$R_0^{muc} = b_{muc} nk / cd_T$	****
b_{LT}	Тот же в лимфоидной ткани	клетка/день	$2 \cdot 10^{11} d_T$	**
k	инфекционность вируса	1/день/копия РНК	$R_0^{LT} = \frac{b_{LT}nk}{cd_T}$	[438]**** [230, 346]
n	Количество вирионов на клетку	копия РНК /клетка		

c	Частота потери вирионов	1/день		
d_T	То же для клеток-мишеней	1/день	0.1	[445]
d_I	То же для вируспродуцирующих клеток	1/день	1.0	[446]
d_L	То же для латентных клеток	1/день	$10^{-3} - 10^{-4}$	[204, 440]
r	Частота активации латентных клеток	1/день	Слизистая: $2 \cdot 10^{-3}$ Лимфоидная: $r(0) = 10^{-3} - 10^{-4}$ $r(E)$: см. (6.23)	[204, 440]
p_{lat} или p	Вероятность латентности	1	[0 1]	
f_{act}	Доля нелатентного переноса при $p_{lat} = 0.5$	1	[0 1]	
f_{nonlat}	Доля нелатентного переноса при $p_{lat} = p_{opt}$	1	[0 1]	

**** Конкретные значения параметров b_{muc} , k , c , n , выбранные при моделировании, не имеют значения при условии, что два экспериментальных значения R_0 верны.

*** Результаты для оптимального p_{lat} (рисунок 6.5С) зависят от одного параметра, R_0^{LT}

** b_{LT} соответствует общему количеству CD4 Т клеток, $2 \cdot 10^{11}$ у неинфицированного человека [355].

* В модели с иммунным ответом, R_0^{LT} увеличено в 1.5 раза, чтобы компенсировать фазу вирусного затмения, которая замедляет вирусную экспансию предказанную моделью^{67,405,416}. Значения параметров, полученные из данных пациентов-людей, одинаковы для лимфоидной ткани и слизистой оболочки, за исключением скорости пополнения клеток-мишеней, b , которая намного меньше для слизистой оболочки чем в лимфоидной ткани, откуда более низкий R_0 [230-232, 444].

6.2.2 Эффект латентности зависит как от слизистой, так и от системной инфекции

Чтобы определить оптимальную вероятность латентности p_{lat}^{opt} , двухкомпонентная модель была использована для расчета эффекта латентности на слизистом и системном этапах инфекции. Вероятность передачи вируса между партнерами во время контакта пропорциональна произведению двух факторов: начального инокулята инфицированных клеток в слизистой оболочке, I_0 , и вероятности того, что инфицированная клетка установит системную инфекцию, p_{estab} , both depending on p_{lat}

$$p_{trans}(p_{lat}) = p_{estab}(p_{lat}) I_0(p_{lat}) \ll 1 \quad (6.18)$$

Заметим, что меньше 1% половых контактов приводит к системному заражению ВИЧ [225-227].

Учитывая уравнение (6.18), общая цель состоит в том, чтобы определить влияние задержки на I_0 и p_{estab} .

Латентность повышает вероятность системной инфекции

Сначала была получена функция $p_{estab}(p_{lat})$. Как упоминалось выше, первые пять дней характеризуются отсутствием обнаруживаемой репликации вируса [232, 444], что доказывает,

что число репродукции в слизистой оболочке R_0^{muc} изначально намного ниже единицы ([19], математическое приложение]).

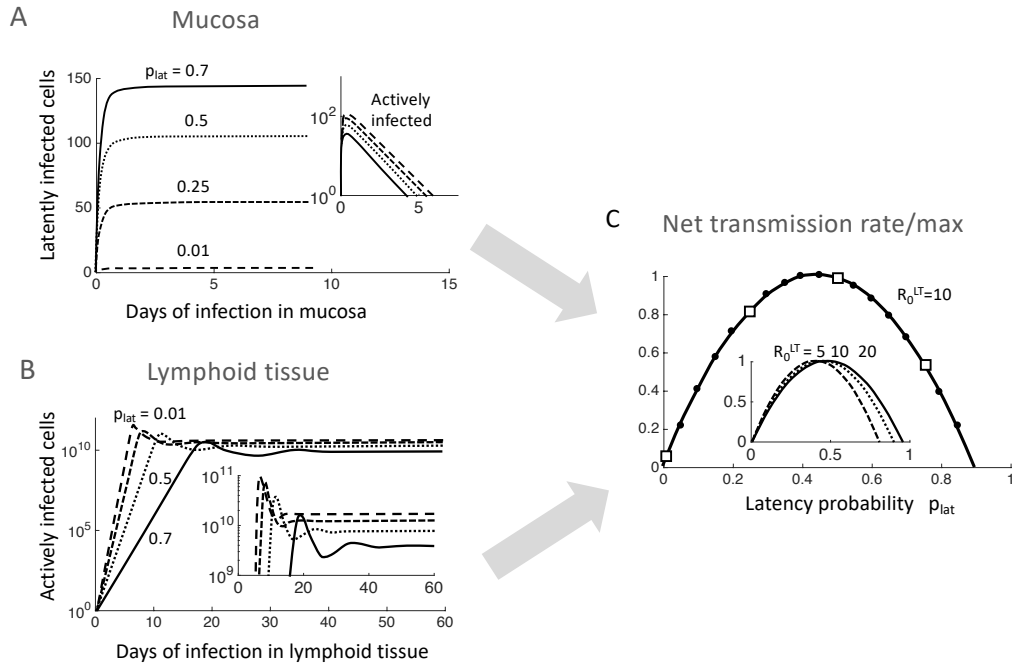


Рисунок 6.5: Вероятность оптимальной латентности оптимизирует увеличение скорости установления инфекции по сравнению с уменьшением заражения между хозяевами. (А) Динамика латентно инфицированных клеток при ранней инфекции слизистой оболочки ($R_0^{muc} = 0.25$). Количество выживших латентно инфицированных клеток увеличивается с p_{lat} . Вставка: Когда вероятность латентности p_{lat} увеличивается, активно инфицированные клетки вымирают быстрее. (В) При системной инфекции ($R_0^{LT} = 10$), вирусная нагрузка (и вирусная доза, передаваемая следующему хозяину) уменьшается с p_{lat} . Результаты на (А, В) получены численно из (6.17) и параметров в таблице 6.2 при $r = 0$. (С) Нормализованная вероятность передачи вируса, p_{trans} , как функция p_{lat} , полученная аналитически из формулы (6.21) (сплошная линия) и численно с использованием уровней плато для активно инфицированных клеток I и латентно инфицированных клеток L из (А, В) (точки). По работе [19].

Как детерминистическая (рисунок 6.5А), так и стохастическая версия модели (рисунок 6.6А,В) предсказывают преимущество приспособленности к латентному периоду, поскольку в его отсутствие все инфицированные клетки вымирают во время инфекции слизистой оболочки (рисунок 6.5А, вставка; 6.6С). Компьютерное моделирование выполнено с $R_0 \leq 0.25$, предполагая несколько десятков изначально инфицированных клеток, как в экспериментах со слизистой оболочкой животных [230-232]. Напротив, даже низкая вероятность латентного периода способствует выживанию вируса (рисунки 6.5А, 6.6D). Чтобы проверить надежность этих прогнозов для всех значений $R_0 < 1$ и $I_0 < 100$, эти результаты были повторены аналитически с помощью метода Марковской цепочки (ветвящегося процесса). Вместо того, чтобы усреднять тысячи симуляций Монте-Карло для каждого R_0, I_0 , этот подход [19] позволяет получить выражение для вероятности исчезновения вируса, как функции времени, в общей форме. Результаты показывают, что при отсутствии латентности, к пятому дню инфекции слизистой оболочки, вероятность исчезновения приближается к 1, если только R_0 не близок к 1 (рисунок 6.6С). Среднее количество выживших инфицированных клеток в слизистой оболочке,

полученное аналитически для процесса Райта-Фишера [[19], Приложение], увеличивается примерно линейно с p_{lat} для каждого значения инокулята I_0 (рисунок 6.6E).

Это рассмотрение было основано на одном компартменте слизистой оболочки. Для проверки, модель процесса Райта-Фишера была расширена до двухкомпонентной модели, которая напрямую связывает слизистую оболочку и лимфоидную ткань ([19], приложение). Между слизистой оболочкой и лимфоидным компартментом различается только значение одного параметра: $R_0^{muc} < 1$, и $R_0^{LT} > 1$ (таблица 6.2). Двухкомпонентная модель подтвердила (рисунок 6.6F), что увеличение p_{lat} увеличивает вероятность установления системной инфекции.

Латентность уменьшает инокулят

В то время как увеличение p_{lat} увеличивает вероятность системной инфекции на переданную зараженную клетку, вероятность передачи также зависит от среднего числа переданных зараженных клеток I_0 , которое считается пропорциональным вирусной нагрузке передающего пациента. Таким образом, было необходимо выяснить, как вирусная нагрузка в стационарной инфекции зависит от вероятности латентности.

С этой целью, детерминистическая модель лимфоидной ткани (6.17) была решена численно при $R_0 = 10$, с начальным условием: одна зараженная клетка, $I(0) = 1$ [228, 229, 366]. Результаты демонстрируют снижение стационарной вирусной нагрузки с увеличением p_{lat} (рисунок 6.5B). Как и ожидалось, увеличение латентности уменьшает инокулят, полученный следующим человеком в цепочке.

6.2.3 Вероятность латентности около 50 % оптимизирует передачу

Используя формулу (6.18) для вероятности передачи вируса, можно вычислить значение p_{lat} , которое максимизирует произведение $p_{estab} I_0$ (и, следовательно, приспособленность вируса на уровне человеческой популяции). С этой целью, p_{lat} и I_0 выражены ниже через p_{lat} .

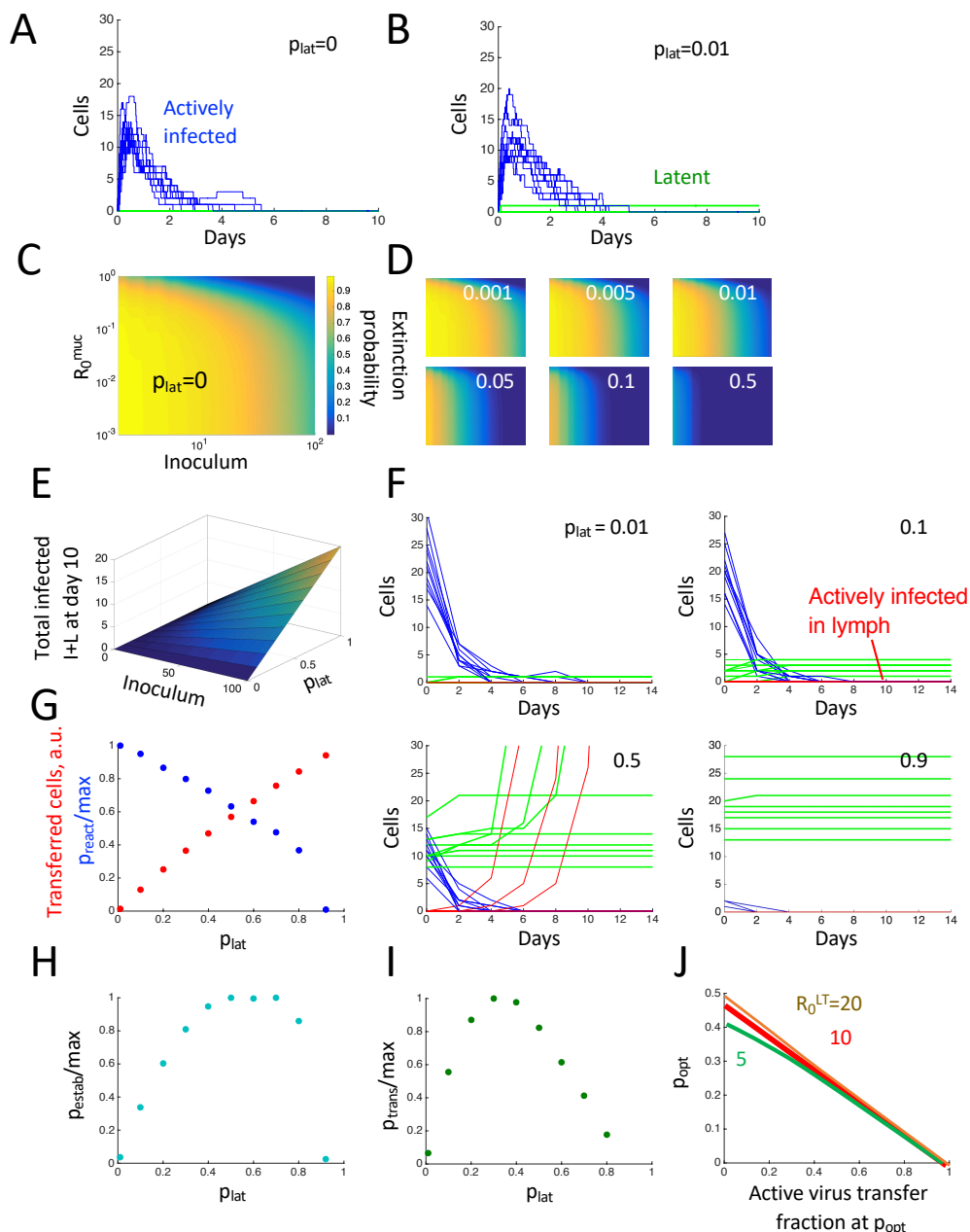


Рисунок 6.6. Стохастическое моделирование демонстрирует, что латентность ВИЧ увеличивает шансы на выживание вируса. (A) 10 репрезентативных прогонов моделирования Гиллеспи однокамерной модели (6.28), при типичных параметрах. (B) То же самое с $p_{lat} = 0.01$. Разница между самыми длинными прогонами в A и B обусловлена случайными флуктуациями. (C) Аналитическая вероятность вымирания на 5-й день после заражения, полученная с использованием метода Марковской цепи, при отсутствии латентности ($p_{lat} = 0$), в зависимости от R_0 и количества первоначально инфицированных клеток, I_0 . (D) Вероятность вымирания для $p_{lat} = 0.001, 0.005, 0.01, 0.05, 0.1, 0.5$ на 5-й день после заражения в зависимости от R_0 и I_0 . (E) Количество выживших клеток на 10-й день при $R_0 = 0.25$, как функция p_{lat} и I_0 , из аналитического процесса Райта-Фишера. (F) Динамика в двух отсеках, бедных ($R_0^{muc} < 1$) и богатых ($R_0^{LT} > 1$) клетками-мишенями, при четырех различных значениях p_{lat} , для 10 прогонов численного моделирования Райта-Фишера. Показаны активно инфицированные клетки слизистой оболочки (синие), латентные клетки лимфоидной ткани (зеленые) и активно инфицированные клетки лимфоидной ткани (красные). (G-I) Моделирование Райта-Фишера в связанных компартментах с непостоянной вероятностью реактивации p_{react} . (G) Доля первоначально инфицированных клеток слизистой оболочки, которые становятся латентными и переносятся в лимфоидную ткань, и вероятность реактивации p_{react} , в зависимости от p_{lat} . (H) Вероятность установления инфекции на начальную зараженную клетку p_{estab} имеет максимум при $p_{lat} \approx 0.6$. (I) Вероятность передачи вируса, включая зависимость передаваемой дозы I_0 от p_{lat} , имеет максимум при $p_{lat} = 0.35$. Каждая точка в (G-I) является средним из 10 000 прогонов с параметрами: $R_0^{muc} = 0.25$, $R_0^{LT} = 0.10$, $d_I = 0.5$, $r = 0.006$. (J) Включение нелатентного переноса между слизистой оболочкой и LT сохраняет высокие значения

p_{opt} . Зависимость p_{opt} от доли переноса активной инфекции f_{latent} пересчитывается апостериорно при $p_{\text{lat}} = p_{\text{opt}}$ По работе [19].

Из-за малого числа репродукции, количество латентно инфицированных клеток в слизистой оболочке $L_{\text{init}}^{R_0 > 1}$ возникает в первом поколении инфекции, $L_{\text{init}}^{R_0 > 1} \approx p_{\text{lat}} I_0$ (рисунки 6.5А и 6.6). Для установления системной инфекции, как минимум одна из латентных клеток должна реактивироваться, что происходит с вероятностью p_{react} . Таким образом, вероятность установления системной инфекции на одну исходную инфицированную клетку равна

$$p_{\text{estab}} = \frac{L_{\text{init}}^{R_0 > 1}}{I_0} p_{\text{react}} \approx p_{\text{lat}} p_{\text{react}} \quad (6.19)$$

Здесь предполагается, что все активные клетки вымерли, и только латентно инфицированные клетки вызывают системную инфекцию (рисунки 6.5А и 6.6).

Далее, средняя инфекционная доза I_0 , переданная новому больному, примерно пропорциональна средней вирусной нагрузке при системной инфекции [191]. Системная инфекция включает начальную вирусную экспансию, пик и стационарное состояние, достигаемое после пика. Аналитический результат для стационарного состояния в хронической инфекции имеет вид [19]

$$I_0 \approx \text{const}(p_{\text{lat}}) \cdot [(1 - p_{\text{lat}})R_0^{\text{LT}} - 1] \quad (6.20)$$

каковая величина уменьшается с вероятностью латентности (рисунок 6.5В). Первый член в уравнении (6.20) постоянен по p_{lat} , но может зависеть от других параметров. Из уравнений (6.18) - (6.20), вероятность передачи вируса как функция p_{lat} определяется выражением

$$p_{\text{trans}} \approx \text{const}(p_{\text{lat}}) \cdot p_{\text{lat}} \cdot [(1 - p_{\text{lat}})R_0^{\text{LT}} - 1] \quad (6.21)$$

которое имеет максимум (рисунок 6.5С) при

$$p_{\text{lat}}^{\text{opt}} \approx \frac{1}{2} - \frac{1}{2R_0^{\text{LT}}} \quad (6.22)$$

Интересно, что, для типичного значения $R_0^{\text{LT}} = 10$, вероятность латентности, которая максимизирует лентивирусную передачу, довольно велика, близка к половине (рисунок 6.5С). В соответствии с аналитическим результатом, численное усреднение инокулята с течением времени также показывает оптимум при $p_{\text{lat}} \approx 0.5$ (рисунок 6.5С). Этот оптимум при $p_{\text{lat}} \approx 0.5$ существует во всем наблюдаемом диапазоне значений R_0^{LT} (рисунок 6.5С). Здесь предполагается, что вероятность реактивации p_{react} не зависит от p_{lat} ; отмена этого допущения несколько снижает $p_{\text{lat}}^{\text{opt}}$ ([19], приложение, рисунок 6.6G).

Результат устойчив к наличию нелатентных маршрутов

Далее оптимальная вероятность латентности была пересчитана для случая, когда часть успешных системных заражений поступает нелатентными путями (например, такими как дендритные клетки Лангерганса, которые тоже экспрессируют CD4 и могут заражаться ВИЧ на низком уровне). Оказалось, что пока доля нелатентных путей меньше 100%, оптимальная вероятность задержки остается очень большой (рисунок 6.6J). Например, если 80% лентивирусных передач осуществляются нелатентными путями, $p_{\text{lat}}^{\text{opt}} = 0.1$ (рисунок 6.6J).

Некоторые изменения в допущениях делают $p_{\text{lat}}^{\text{opt}}$ даже больше, чем в (6.22) [19]. Ослабление предыдущего неявного предположения о том, что латентно инфицированные клетки умирают до реактивации в слизистой, $r \ll d_L$, снижает стоимость латентности для инокулята и, следовательно, увеличивает оптимальную вероятность латентности. В результате, если $r > d_L$, имеем $p_{\text{lat}}^{\text{opt}} = 1$. Далее, если отменить предположение о линейной зависимости между вирусной нагрузкой и скоростью передачи и позволить ей насыщаться при высоких вирусных нагрузках [225], тогда I_0 станет менее чувствительным к латентности, а оптимальное значение, опять же, увеличится до 1 ([19], приложение).

Эксперименты на клеточной культуре и приматах подтверждают предсказания модели

Предсказанная латентная вероятность 50% [213] или выше [214] действительно наблюдалась в клеточной культуре. Это значение также согласуется с исследованием на макаках-резусах, где на 3-й день инфекции слизистой оболочки был зарегистрирован большой резервуар латентно инфицированных клеток [210]. Однако это значение отличается от очень низкой частоты латентных клеток, наблюдаемой у хронически инфицированных пациентов, получающих антиретровирусное лечение. Только одна из $10^6 - 10^7$ CD4 Т-клеток латентно инфицированы [176, 441], что соответствует $p_{\text{lat}} \sim 10^{-5} - 10^{-4}$ [441, 443]. Некоторые другие исследования показывают, что латентная частота клеток в 60 раз выше [439], но это все еще намного ниже 50%. Таким образом, проблема заключается в объяснении разницы между двумя экспериментами: высокой частотой латентности в ранней инфекции или культуре клеток, и низкой частотой в хронической инфекции. Какой фактор может отсутствовать в первом случае, но присутствовать во втором?

6.2.4 Почему латентное состояние встречается редко в пациентах

В отличие от ранней инфекции или клеточной культуры, хроническая инфекция вызывает сильный ВИЧ-специфический иммунный ответ, включая Т- и В-клеточный ответ [278]. Как видно из экспериментов по обеднению CD8 Т-клеток у обезьян [212, 281] и быстрой

генетической изменчивости эпитопов CD8 Т-клеток [18, 286, 356], цитотоксический Т-клеточный ответ сильно ограничивает распространение ВИЧ в организме хозяина. Иммунный ответ должен быть включен в модель вирусной динамики [190, 191], чтобы объяснить 100-1000-кратное снижение вирусной нагрузки между пиком острой инфекции и хронической инфекцией [445]. CD8 Т-клетки убивают активно инфицированные клетки, которые они распознают по специфическому связыванию TCR с комплексом МНС-I-вирусный пептид, и подавляют репликацию вируса в клетках путем секреции цитокинов. Однако, ни один из этих механизмов не вредит латентному вирусу, который не экспрессирует вирусные белки и, следовательно, не может быть распознан CD8 Т-клетками.

Включение иммунного ответа

Выше предполагалось, что p_{lat} и r постоянны. В действительности, оба параметра чувствительны к фазе клеточного цикла и, следовательно, к цитокинам, секретируемым Т-клетками, которые могут активировать клетки. Уровни активации CD4⁺ Т-клеток повышаются при хронической ВИЧ-инфекции, о чем свидетельствует экспрессия трех маркеров активации [444] и повышенная скорость оборота CD4 Т-клеток [180]. Одним из механизмов активации клеток является гомеостатический ответ на истощение CD4 Т-клеток ВИЧ-инфекцией. Другим потенциальным механизмом является секреция CD8 Т-клетками цитокина TNF- α [355] и накопление в клетке факторов транскрипции, таких как NF- κ B, которые активируют промотор ВИЧ и подавляют переход провируса в латентное состояние. Клеточные факторы активации могут резко снижать p_{lat} и, таким образом, активировать транскрипцию ВИЧ [214, 447, 448]. Работа [20] продемонстрировала, что клеточная активация влияет на экспрессию провируса ВИЧ и, следовательно, снижает p_{lat} и увеличивает r за счет изменения относительных временных интервалов латентного и активного состояний.

Поскольку адаптивный иммунный ответ секретирует клеточные факторы активации [214, 444, 447, 448], этот эффект может объяснить пониженный уровень латентности у хронически инфицированных пациентов (рисунок 6.8А). Поскольку цитокины, секретируемые CD8 Т-клетками, недолговечны, можно предположить, что p_{lat} и r являются функциями текущей концентрации эффекторных CD8 Т-клеток, $E(t)$, как показано ниже (рисунок 6.7В).

$$p(E) = p(0) \frac{E_{0L}}{E_{0L} + E}, \quad r(E) = r(0) + d_I \frac{E}{E_{0L} + E} \quad (6.23)$$

Здесь E_{0L} — характерное число иммунных клеток, при котором половина инфицированных клеток получает цитокиновый сигнал выше провирусного порога активации, а значения $p(0) \approx 0.5$ и $r(0) = 10^{-4} - 10^{-3}$ (табл. 6.2) соответствуют отсутствию ВИЧ-специфического иммунного ответа.

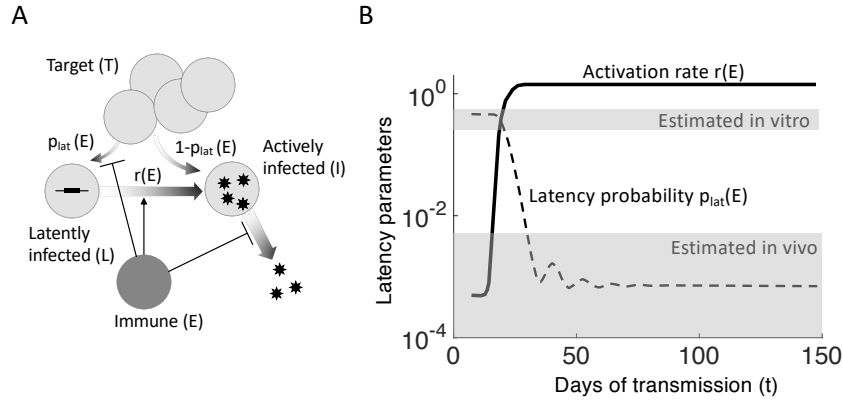


Рисунок 6.7: Включение иммунного ответа объясняет разницу в частоте латентных клеток ВИЧ между моделями клеточных культур и хронически инфицированными людьми. (А) Расширенная модель включает CD8 Т-клетки, E , которые убивают активно инфицированные клетки и активируют латентно инфицированные клетки, (6.23)-(6.24). (В) Вероятность латентности p_{lat} и скорость реактивации r сильно изменяются после пика виремии из-за активации сторонних цитокинов иммунными клетками, формулы (6.23). По работе [19].

К модели системной инфекции (6.17), добавляются параметры латентности (6.23) и динамика антиген-специфического ответа Т-клеток CD8 [190, 449, 450]. При системной инфекции, количество клеток во всех компартаментах очень велико, поэтому мы можем безопасно использовать детерминированное приближение

$$\begin{aligned}
 \frac{dT}{dt} &= b - d_T T - kVT \\
 \frac{dI_E}{dt} &= (1 - p_{lat})kVT - d_{IE}I_E + rL \\
 \frac{dI}{dt} &= d_{IE}I_E - d_I \left(1 + \frac{E}{E_0}\right)I \\
 \frac{dL}{dt} &= p_{lat}kVT - d_L L - rL \\
 \frac{dV}{dt} &= nd_I I - cV \\
 \frac{dE}{dt} &= \frac{aI(E_N + E)}{I + I_{av}} - d_E E \\
 \frac{dE_N}{dt} &= \frac{aIE_N}{I + I_{av}}
 \end{aligned} \tag{6.24}$$

где дополнительные параметры модели указаны в таблице 6.3. По сравнению с моделью (6.17), которая предполагает, что смертность вируспродуцирующих клеток d_I постоянна, в расширенной модели (6.24), цитотоксические CD8 Т-клетки E могут убивать активно инфицированные клетки (термин E/E_0 в строке 3). Два новых уравнения описывают активацию наивных CD8 Т-клеток E_N вирусом, а также пролиферацию и распад эффекторных клеток. Как обычно в моделях, включающих цитотоксический ответ, добавлена задержка между инфекцией и продуцированием вируса, фенотип I_E . Задержка необходима для предотвращения артефакта

резкого падения виремии и гигантских колебаний [223, 375, 435, 450]. Дополнительные параметры в (6.24) приведены в табл. 6.3.

Таблица 6.3. Дополнительные параметры модели включающей иммунный ответ, рисунок 6.7 и формулы (6.23) — (6.24), согласующейся с данными в пациентах (рисунок 6.8)

Обозначение	Определение	Единицы	Определение	Ссылки
a	Максимальная частота пролиферации CTL	1/день	$a - d_E = 0.5$	[451]
d_E	Смертность CTL	1/день	0.75	Подгонка данных*
d_{IE}	Обратная длина фазы затмения	1/день	$1.4/day$	[234, 369]
E_0	Число CTL которое вдвое сокращает время жизни зараженной клетки	клетка	$2 \cdot 10^8$	Подгонка данных*
I_{av}	Порог авидности по числу зараженных клеток	клетка	10^8	Подгонка данных*
$E_N(0)$	Число наивных CTL	клетка	$2 \cdot 10^6$	[355]
E_{0L}	Число CTL которое уменьшает p_{lat} вдвое	клетка	$4 \cdot 10^6$	Подгонка данных*

*Из семи параметров, четыре параметра d_E, E_0, E_{0L}, I_{av} подгоняются чтобы соответствовать четырем экспериментальным плато в пациентах (рисунок 6.8), а три остальных менее чувствительные и фиксированы на значениях взятых из литературы. В частности, параметры E_0, E_{0L}, I_{av} оцениваются по прогнозируемым стационарным уровням E, I, L по сравнению с их экспериментальными оценками (рисунок 6.8): $E = 10^9$ клеток [278, 452], $I = 10^8$ ячеек [55], и $L = 10^6$ ячеек [176]. По оценкам, d_E соответствует L при антивирусной терапии: $L = 10^5$ ячеек [204]. Предполагается, что общее количество клеток $T(0) = b/d_T = 2 \cdot 10^{11}$ как для CD8 T, так и для CD4 T клеток [355]. Эта величина может быть эффективно ниже для CD4 T клеток, из-за гибели клеток из-за abortивной инфекции [453].

Система уравнений (6.24) представляет собой довольно простую модель клеточного ответа CD8. Более сложные модели включают зависимую от хелперных клеток активацию [223, 224, 375, 435] и множественные клоны CTL, распознающие разные эпитопы [18]. Тем не менее, эта модель годится для динамики латентности, так как эффекты CD8 ответа на латентность достаточно сильны чтобы объяснить огромное изменение доли латентного ВИЧ между фазой слизистой инфекции и инфекцией системной. Устойчивость к вариациям модели была проанализирована [[19], Приложение].

До адаптивного иммунного ответа и до вирусного пика, модель предсказывает высокие латентные вероятности ~ 0.5 и низкие показатели реактивации, как и в приведенных выше упрощенных моделях. Однако после пика усиление клеточной активации [444] снижает $p_{lat}(E(t))$ до низких уровней и увеличивает $r(E(t))$ до высоких уровней [рисунок 6.7В и формула (6.23)]. Таким образом, иммунная модель успешно объясняет огромную разницу между латентными частотами в культуре и при системной инфекции.

Модели, включающие иммунный ответ, согласуются с данными в пациентах

Следующая задача — протестировать модель на следующих данных пациентов:

- вирусная нагрузка до антивирусной терапии [225],
- число эффекторных Т-клеток до терапии [278],
- число латентно инфицированных клетки до терапии [176],
- оно же после терапии [454].

Все эти наблюдаемые величины, в конце концов, выходят на плато (рисунок 6.8А). Модель иммунного ответа, уравнения (6.24), позволяет эти значения на плато подогнать.

Динамика системной инфекции рассчитывается численно по (6.23) и (6.24), с использованием программы ODE15S.m, реализованной в программном обеспечении MATLAB™ (рисунок 6.8А). Параметры модели приведены в таблице 6.3. Первоначально $p_{\text{lat}}(0) \sim 0.5$ и $r(0) \ll 1$ (рисунок 6.7В), отсюда и высокий уровень латентных клеток, сходный с уровнем активных клеток (рисунок 6.8А). После пика виремии, иммунный ответ увеличивает r и снижает p_{lat} на несколько порядков и, тем самым, истощает латентный резервуар. Популяция латентных клеток, L , выполаживается на низком уровне, при котором поступление новых латентно инфицированных клеток из компартмента клеток-мишеней и активация латентных клеток взаимно компенсируются (рисунок 6.8А). Следовательно, латентный компартмент L динамически связан с активно инфицированным компартментом I .

Большинство инфицированных получают антиретровирусную терапию (АРТ). Начало АРТ блокирует новые клеточные инфекции. Модель предсказывает снижение количества всех инфицированных клеток, сначала вируспродуцирующих клеток, а затем латентных клеток, поскольку эти два компартмента связаны (рисунок 6.8А). Предсказанный уменьшение латентного резервуара, наблюдаемый у пациентов (рисунок 6.8А), происходит так же быстро, как клетки активируются, $r(t)$. В начале АРТ, из-за высокого уровня CD8 Т-клеток, скорость активации высока, а латентный компартмент быстро обедняется. Истощение антигена (вируса) вызывает сокращение популяции иммунных клеток E . В результате, скорость реактивации $r(t)$ снижается и вскоре возвращается к своему низкому фоновому уровню, $r(0)$, формула (6.6). При длительном АРТ, любое дальнейшее уменьшение латентного резервуара происходит очень медленно, как это и наблюдается у пациентов (рисунок 6.8А). Модель объясняет 10-кратное уменьшение количества ДНК-позитивных клеток, наблюдаемое после АРТ [176].

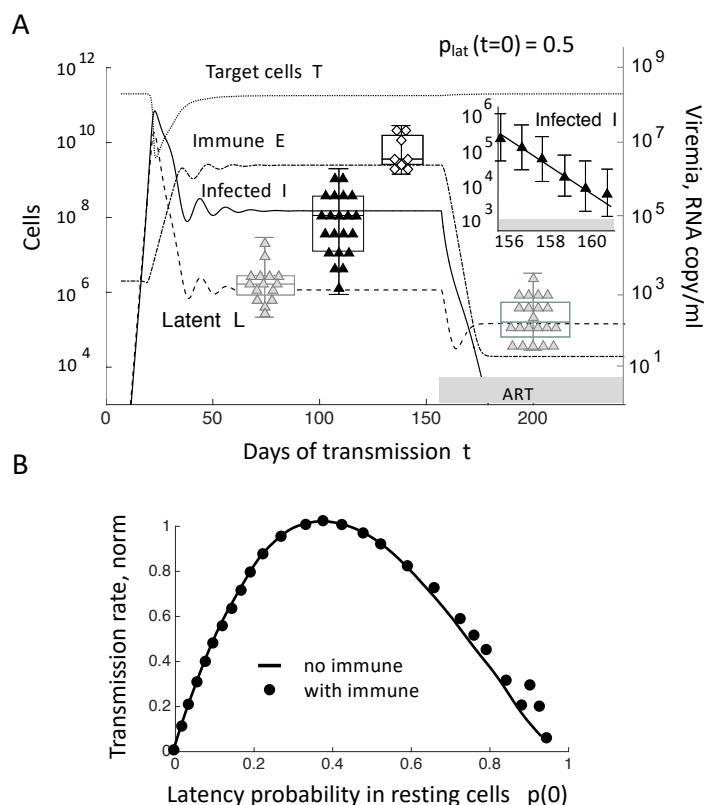


Рисунок 6.8. Расширенная модель иммунного ответа подогнанная к данным в пациентах не изменяет оптимальную вероятность латентности. (A) Динамика клеточных компартментов во время системной инфекции, рассчитанная по уравнениям (6.23) и (6.24). Антиретровирусная терапия (АРТ), начатая во время стационарной инфекции, вызывает снижение латентного резервуара L . Выполживание падения латентного резервуара происходит из-за снижения числа иммунных клеток E во время АРТ. Точки данных по пациентам-людям: вирусная нагрузка связана с количеством инфицированных клеток до АРТ [225] (черные треугольники); латентные клетки до АРТ [176] и после высокоактивного АРТ [454] (серые треугольники); эффекторные CD8 Т-клетки [278] (белые ромбы). Врезка: Вирусная нагрузка после начала АРТ [234] (столбики погрешностей показывают стандартное отклонение). (B) Нормализованная скорость передачи вируса p_{trans} , как функция $p_{\text{lat}}(0)$, рассчитанная из динамики (A) и формулы (6.18). Для сравнения показаны два случая: с иммунными клетками (E , черные кружки) и без иммунных клеток ($E = N = 0$, сплошная кривая). Включение в модель иммунных клеток слабо влияет на предсказание большой оптимальной латентной вероятности для покоящихся клеток, $p_{\text{lat}}^{\text{opt}}(0) \approx 0.5$. Параметры модели на панелях А-В приведены в таблицах 6.2 и 6.3, при этом $p_{\text{lat}}^{\text{opt}} = 15$ и $p_{\text{lat}}(0) = 0.5$ на панели А. См. также рисунок S3 в [19]. По работе [19].

Интересно, что включение иммунного ответа и зависимости параметров латентности от времени не меняет сделанного выше предсказания о том, что оптимальная вероятность начальной латентности $p_{\text{lat}}(0)$ велика, $p_{\text{lat}}^{\text{opt}}(0) \sim 0.5$ (рисунок 6.8B). Предсказание также устойчиво к эпидемиологическим предположениям, таким как монотонная зависимость передачи от вирусной нагрузки или преобладание хронической или острой инфекции в передаче вируса.

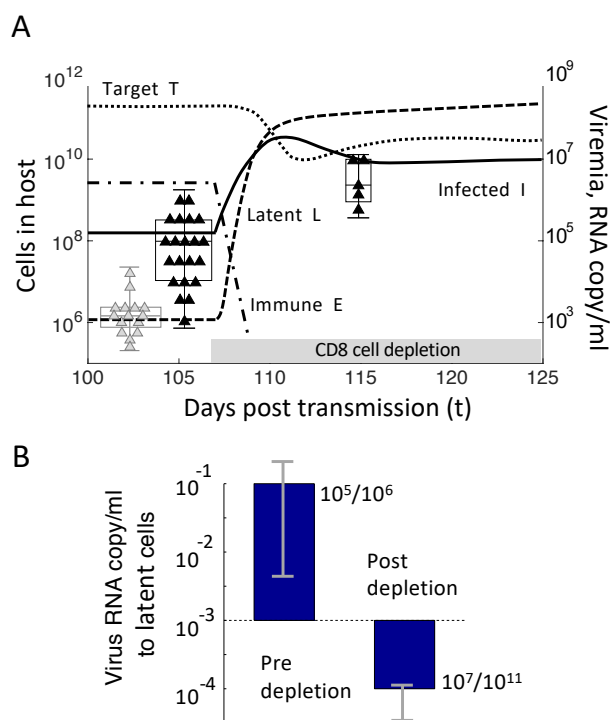


Рисунок 6.9: Предсказание, поддающееся проверке: истощение CD8+ T-клеток у макак, инфицированных ВИО, увеличит латентный резервуар в гораздо большей степени, чем вирусную нагрузку. (A) Динамика системной инфекции для модели с иммунным ответом (6.23), (6.24). Точки данных и параметры такие же, как на рисунке 6.8. (B) Соотношение активно инфицированных и латентно инфицированных клеток будет инвертировано истощением CD8+ T-клеток; «пост-истощение» относится к 125-му дню на (A). Увеличение вiremии наблюдалось в работах [212, 281]. Пунктирная горизонтальная линия 10^{-3} РНК/мл/клетка соответствует соотношению 1:1 латентно и активно инфицированных клеток. Синие полосы соответствуют параметрам и размерам отсеков в (A). Максимальные ожидаемые ошибки (вертикальные столбцы) оцениваются по границам прямоугольника с усами в (A) (два средних квартиля). См. также рисунок S4 в [19]. По работе [19].

6.2.5 Прогноз: истощение T-клеток CD8+ у макак увеличит латентный резервуар

Вышеуказанная расширенная модель позволила сделать следующее, поддающееся проверке предсказание. Поскольку CD8 T-клетки уменьшают латентность, их искусственное истощение антителами против CD8 привело бы к тому, что вероятность латентности p_{lat} вернулась бы к исходно высокому значению 0.5, а скорость реактивации r упала бы до исходного низкого значения. У макак-резусов, инфицированных ВИО, искусственное истощение CD8+ T-клеток вызывает быстрое (10-100)-кратное увеличение вирусной нагрузки [212, 281]. Настоящая модель предсказывает, что истощение CD8 в хронической инфекции должно увеличить латентный резервуар еще выше, на 4-5 порядков (рисунок 6.9A). В то же время, во время ранней предпиковой инфекции, истощение CD8 [455] увеличило бы размер латентного резервуара, который и так велик, только в пару раз [[19], рисунок S4]. Проверка этих прогнозов может пролить свет на динамическую природу образования скрытого резервуара.

Молекулярная инженерия рекомбинантных штаммов SIV с сильно сниженной латентностью могла бы предложить еще один экспериментальный тест на животных. Если бы

эти штаммы вызывали системные инфекции реже, чем дикий тип, этот результат прямо продемонстрировал бы преимущество латентности для передачи вируса.

6.2.6 Заключение

Латентность ВИЧ служит троянским конем, который позволяет вирусу проникать внутрь организма и вызывать системную инфекцию. Даже при наличии нелатентных маршрутов инфекции, основной вывод что латентность помогает передаче вируса на системный уровень остается справедливым (рисунок 6.6J). Предложена прямая экспериментальная проверка этого предсказания (рисунок 6.9).

В случае подтверждения, можно было бы предложить следующую потенциальную терапию (рисунок 6.10). Во время ранней инфекции (первые 3-5 дней), реактивация латентных клеток запускает системную инфекцию (рисунок 6.6F). Эта реактивация является слабым звеном, поскольку она зависит от начальной иммунной стимуляции, например, вызванной антигенпрезентирующими клетками, такими как дендритные клетки Лангерганса [456]. Если бы ранняя реактивация была подавлена, например, супрессорами реактивации транскрипции ВИЧ [215] или контролем миграции клеток, системная инфекция стала бы менее вероятной (рисунок 6.10A-C). Ограничением предлагаемого подхода является необходимость раннего выявления ВИЧ в течение первой недели заражения. Альтернативно, этот метод должен использоваться в качестве профилактики после незащищенного секса. Тем не менее, раннее выявление и раннее лечение ВИЧ применяется на практике [231]. Значительно меньший латентный резервуар $\sim 10^2$ клеток у инфицированного человека потребовал бы реактивации меньшего количества латентных клеток повысив эффективность последующей терапии «активировать и вычищать» [457, 458].

6.3 Роль рекомбинации в адаптации и оптимальная частота мутаций в системе вирус полиомиелита - мышь

Как обсуждалось в главах 2, 3 и 4, генетическая эволюция происходит благодаря совместному действию нескольких биологических факторов. Новые генетические варианты создаются путем случайных мутаций (ошибок копирования) и рекомбинаций разных геномов. Их последующая динамика определяется факторами естественного отбора (включая эпистаз), случайным генетическим дрейфом и эффектами генетического сцепления.

Роль мутации в эволюции двоякая. Хотя более высокая вероятность мутации позволяет организму быстрее адаптироваться, она также приводит к большему количеству вредных мутаций. Модели многолокусной эволюции, в том числе входящие в данную диссертацию, предсказывают существование оптимальной частоты мутаций [7, 258]. Различные модели также показывают, что эволюционная роль рекомбинации заключается в снижении негативных последствий генетического сцепления, ускорении адаптации и устранении вредных аллелей [8, 60, 67-69, 71] (глава 4). Однако экспериментальных подтверждений этих математических предсказаний, в многолокусных условиях, слишком мало. Обсуждаемая ниже работа [22] проделана чтобы заполнить этот пробел.

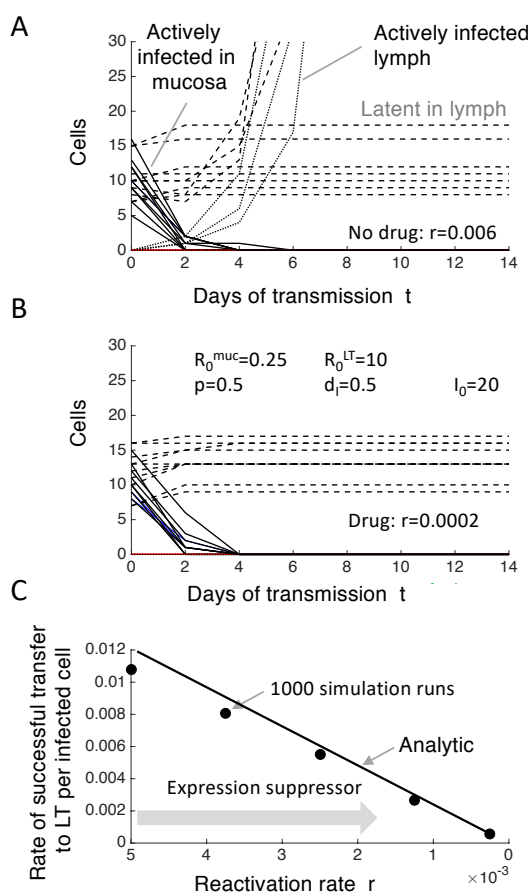


Рисунок 6.10: Уменьшение реактивации в постслизистой фазе инфекции уменьшит размер латентного резервуара, повысив эффективность последующей терапии «активировать и вычищать». (А) Установление системной инфекции между днями 4-6. Стохастическое моделирование процесса Райта-Фишера с частотой реактивации латентных клеток $r(0) = 6 \cdot 10^{-3}$ /сутки. (В) То же, но при уменьшении $r(0)$ в 30 раз. (С) Предказанное подавление передачи ВИЧ супрессором экспрессии провируса. По работе [19].

В совместной работе с экспериментальной лабораторией [22], проведены эксперименты с вариантами вируса с различной скоростью мутации и рекомбинации, как в клеточной культуре, так и на животных (мышах), а затем использовано математическое моделирование для интерпретации экспериментальных результатов. Эксперименты кратко рассмотрены ниже; подробности протоколов можно найти в оригинальной работе [22].

Экспериментаторы [22] синтезировали варианты вируса с высокой и низкой скоростью мутации, а также высокой и низкой скоростью рекомбинации, в любой комбинации. Затем, скорость вирусной адаптации у животного-хозяина (мыши) измерена через тропизм тканей и летальность вируса в мышах. Эти экспериментальные результаты проинтерпретированы с использованием компьютерной модели из работы [15] (§ 2.3). Аналитические предсказания оптимальной вероятности мутации, сделанные ранее для эволюции на большом промежутке времени [7, 258], в данной работе обобщены на случай короткой эволюции и проверены экспериментально.

6.3.1 Эксперименты на штаммах с измененной частотой рекомбинации и мутации

Мутация D79H снижает скорость рекомбинации в 10 раз

Частота мутаций полиовируса зависит от его белка полимеразы [459, 460]. Используя серийные пассажи, экспериментаторы [22] искусственно селектировали мутантный вирус с необычно низкой скоростью рекомбинации, что было подтверждено двумя отдельными методами ([22], рисунок 1). Геном этого варианта содержал единственную аминокислотную замену в полимеразе, мутацию аспартата на гистидин на номере 79, обозначенную D79H. Используя точный метод глубокого секвенирования CirSeq для измерения частоты случайных мутаций, команда проверила, что мутация D79H, отключающая рекомбинацию, не влияет на частоту мутаций на остальных основаниях (рисунок 6.11A). Кроме того, мутация D79H не влияет на приспособленность вируса, измеренную в эксперименте по конкуренции роста штаммов в клеточной культуре ([22], рисунок 3).

Открытие мутации D79H, снижающей скорость рекомбинации в 10 раз, дало возможность изучить эффект одновременного изменения частоты рекомбинации и мутации. Для этой цели, D79H был объединен с ранее обнаруженным высокоточным мутантом G64S (т.е., с низкой частотой мутаций) или низкоточным мутантом H273R [460-462]. Три мутации расположены далеко друг от друга в структуре белка полимеразы и влияют на различные функциональные домены ([22], рисунок 1C). Чтобы связать скорость адаптации с частотой рекомбинации и мутации, экспериментальная группа создала двойные мутанты G64S/D79H (обозначенные ниже для краткости GD или HiFi Rec-) и H273R/D79H (HD, LoFi Rec-) и одиночные мутанты G64S (G, HiFi), D79H (D, Rec-) и H273R (H, LoFi). Вместе с вирусом дикого типа, она получила набор из шести штаммов с любой комбинацией нормальной или низкой частоты рекомбинации с нормальной, высокой или низкой частоты мутации.

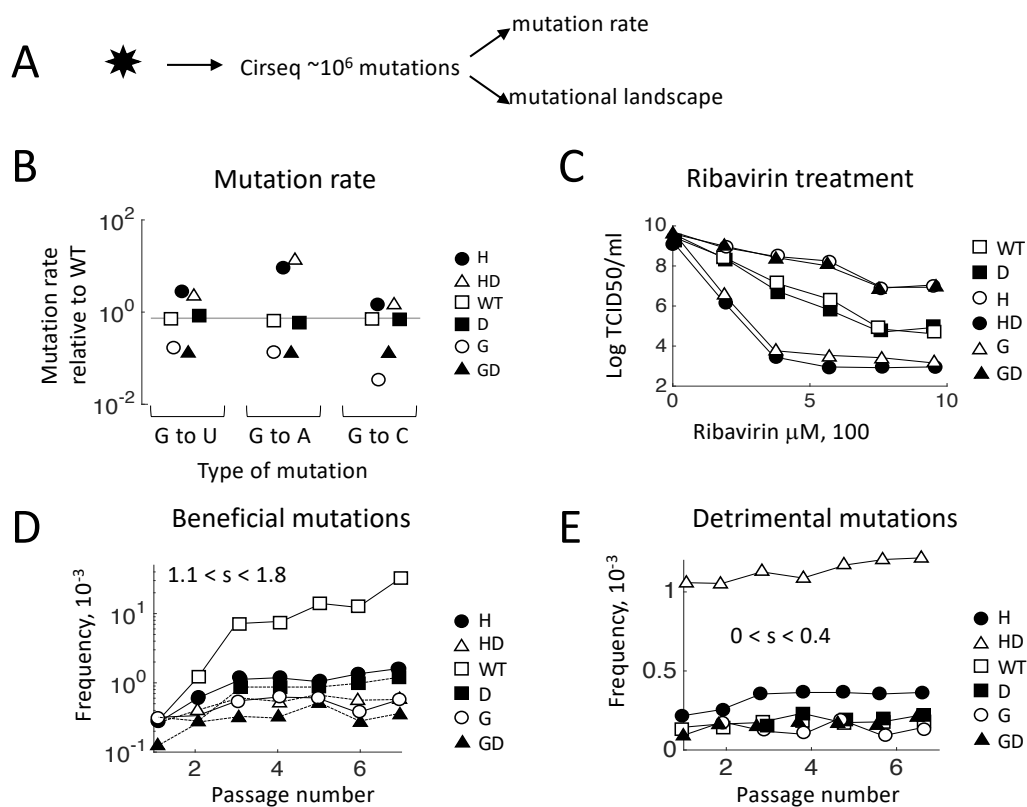


Рисунок 6.11: Мутация с дефицитом рекомбинации не изменяет частоты случайных мутаций, но значительно влияет на скорость накопления полезных и вредных мутаций. (А) Серия пассажей вирусных штаммов в клетках HeLa. Для каждого пассажа, 10^7 монослойных клеток HeLaS3 инфицировали вирусными штаммами при множественности заражения 0.1 с размером популяции 10^6 инфекционных единиц. Частоты мутаций были рассчитаны в соответствии с опубликованным протоколом [93, 463]. Штаммы: дикий тип (WT), мутант с дефицитом рекомбинации D79N (сокращенно D), мутант высокой точностью копирования G64S (сокращенно G), мутант низкой точности копирования H273R (H) и двойные мутанты G64S/D79N (GD) и H273R/D79N (HD). (В) Частота летальных мутаций, которые приводят к бессмысленным кодонам, использовалась для измерения частоты мутаций [93, 463]. (С) Анализ чувствительности вирусных штаммов к мутагену РНК-вируса рибавирина в клетках HeLaS3 предварительно обработанных указанными концентрациями рибавирина (показано на оси X) в течение 4 часов. (D) Накопление полезных мутаций через пассажи вирусных штаммов. Полезные мутации находятся в интервале $0.1 < s < 0.3$, со средним $s = 0.2$. (E) Накопление вредных мутаций в пассажах вирусных штаммов в диапазоне от 0.6 до 1 приспособленности дикого типа. В приведенном ниже моделировании, определяем s определяется через относительную приспособленность $\exp(s) - 1 \approx s$ для малых s . По работе [22].

Мутация D79N замедляет адаптацию вируса в культуре

При серийных пассажах, увеличение частоты полезных мутаций наблюдалось только у вируса дикого типа (рисунок 6.11D). Частота полезных мутаций (в верхних 5% прироста приспособленности) увеличилась за 7 пассажей в 100 раз. Такого увеличения не наблюдалось у двойных мутантов с высокой или низкой частотой мутаций, HD и GD, и лишь незначительное увеличение произошло для одиночных мутантов D (редкие рекомбинации) и H (частые мутации) (рисунок 6.11D). Этот результат указывает на то, что способность отбора к адаптации развилась у WT оптимально и может быть только снижена, либо путем изменения точности полимеразы в любом направлении, либо путем уменьшения рекомбинации. Их сочетание особенно вредно для адаптации (рисунок 6.11D).

Напротив, накопление вредных мутаций повышено у мутантов с высоким уровнем мутаций (H и HD), особенно в варианте с дефицитом рекомбинации и высокой частотой мутаций (HD) (рисунок 6.11E). Последний эффект, повышенное накопление вредных мутаций при низкой рекомбинации, происходит из-за генетического сцепления и называется «храповиком Мюллера» [7, 11, 69, 245, 246] (§ 3.1).

Двойные мутанты нарушают адаптацию вируса у мышей

При заражении животного, вирус подвергается давлению отбора, вызванному иммунным ответом хозяина (§ 7.2) и необходимостью адаптироваться к различным тканям и типам клеток [464]. Таким образом, инфекция *in vivo* обеспечивает окончательную проверку роли мутации и рекомбинации в адаптации вируса. Пять мутантов полиовируса описанных выше и дикий тип использовались для заражения мышей двумя путями: системная инфекция путем внутривенной инъекции или более локальный внутримышечный путь, ведущий в центральную нервную систему (ЦНС) (рисунок 6.12A).

После внутривенного заражения, вирус дикого типа (WT) быстро размножился в организме животного, вторгнулся в ЦНС и убивал всех мышей в течение 6 дней (рисунок 6.12B). Столь же летальным был дефектный по рекомбинации штамм D (Rec-), и маломутабельный штамм G (HiFi) также убивал 80% мышей (рисунок 6.12B). Высокомутабельный штамм H (LoFi) был менее патогенным при внутривенной инокуляции (70% мышей выжили), но столь же летальным при внутримышечной инокуляции, ведущей к поражению ЦНС (рисунок 6.12C).

Использование двойных мутантов с измененной частотой мутации, в любую сторону, и дефективных по рекомбинации резко снижало патогенность: независимо от пути заражения, выживало 90-100% мышей, привитых двойными мутантами GD или HD (рисунок 6.12B и C). Измерение 50% летальной дозы каждого варианта вируса при внутримышечном введении показало, что патогенность двойных мутантов была снижена в 100 раз по сравнению с диким типом ([22], рисунок 4D). Чтобы проверить эффект изменения приспособленности на нейронах мыши, была использована интракраниальная инокуляция двойного варианта GD в головной мозг. Большое количество вируса свидетельствовало о способности GD к репликации в тканях головного мозга [22].

Следует заключить, что снижение скорости рекомбинации в сочетании с сильным изменением скорости мутаций резко снижает способность вируса к адаптации и установлению инфекции в центральной нервной системе, в то же время обеспечивая нормальный (дикий тип) уровень репликации вируса в селезенке и мышцах ([22], рисунок 5).

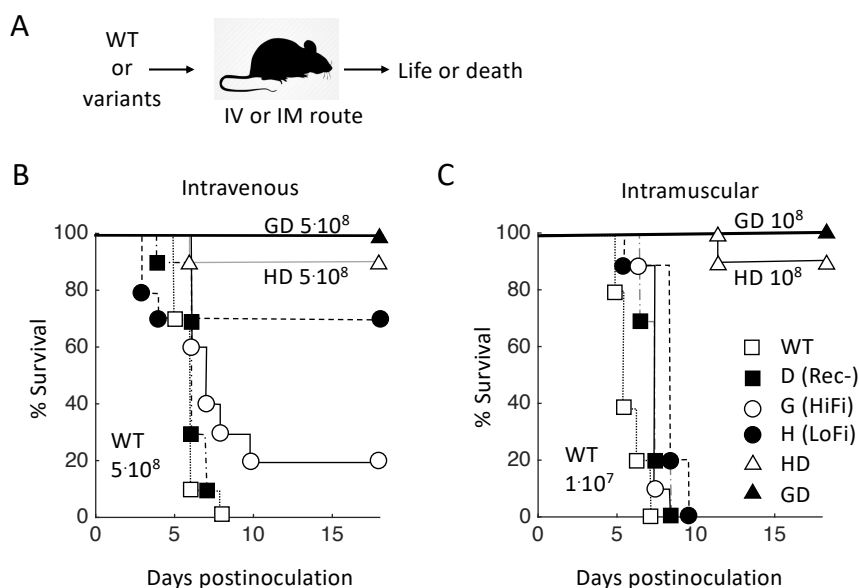


Рисунок 6.12: Дефект рекомбинации и изменение частоты мутаций значительно ослабляют вирулентность полиовируса. (А) Измерялась выживаемость мышей, инфицированных WT или мутантами двумя путями. (В) Процентная выживаемость восприимчивых мышей, инфицированных через хвостовую вену $5 \cdot 10^8$ инфекционных единицами вируса дикого типа (WT), двойными мутантными штаммами, включая штамм высокоточный/дефектный по рекомбинации GD, штаммами с низкой точностью/дефектом рекомбинации HD, и единичными мутантами, включая D, G, и H. (С) Процентная выживаемость восприимчивых мышей, инфицированных внутримышечным путем 10^7 инфекционных единиц штамма WT, D, или G. Мышей инфицировали 10^8 инфекционных единиц штамма HD или GD. По работе [22].

6.3.2 Математическое моделирование и подгонка данных

Вероятность мутации вируса дикого типа максимизирует скорость адаптации вируса в мышцах

Как предсказывают аналитические и вычислительные модели, скорость адаптации бесполой популяции имеет максимум по вероятности мутации, из-за баланса между накоплением полезных и вредных мутаций [7, 258]. Добавление процесса рекомбинации ускоряет адаптацию (глава 4). Модели предсказывают, что в зависимости от временной шкалы эволюции и исходного полиморфизма популяции, рекомбинация и мутация будут влиять на адаптацию либо независимо, либо в синергии. В краткосрочной перспективе и при наличии начальной генетической вариации, рекомбинация и естественный отбор действуют независимо от мутаций *de novo* [8-10, 13, 14]. В долгосрочной эволюции или при отсутствии исходного полиморфизма, два процесса действуют синергически, и геном эволюционирует, как система частично перекрывающихся бесполок блоков [74-76, 396].

Следовательно, следующей задачей было проверить, на основе обсуждавшихся выше экспериментов, действительно ли адаптация имеет оптимум по скорости мутаций и зависит ли оптимальная скорость мутаций от рекомбинации. В этом отношении были рассмотрены две альтернативные гипотезы: либо рекомбинация ускоряет адаптацию, не влияя на оптимальную

скорость мутаций, либо взаимодействие между рекомбинацией и мутациями сдвигает оптимальную скорость мутаций (рисунок 6.13). Чтобы выяснить, какая гипотеза соответствует экспериментальной системе, была использован программа основанная на методе Монте Карло [15] из § 2.3, в которой можно задавать начальные состояния популяции, параметры и факторы модели и промежутки времени.

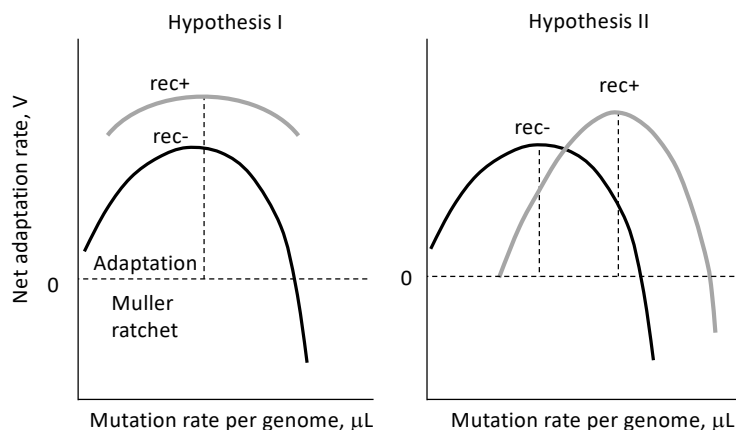


Рисунок 6.13: Две гипотезы о зависимости скорости адаптации вируса от частоты мутации и рекомбинации. Гипотеза 1: скорость адаптации имеет максимум по частоте мутаций, которая повышается за счет рекомбинации независимо от скорости мутаций. Черные и серые линии: схематическая скорость адаптации в отсутствие и в присутствии рекомбинации соответственно. Оптимум представляет собой компромисс между естественным отбором полезных мутаций и накоплением вредных мутаций из-за хrapовика Мюллера. Гипотеза 2: рекомбинация и мутация действуют синергетически, и рекомбинация сдвигает максимум скорости мутации. По работе [22].

Отправной точкой была аналитическая модель [7, 11] без рекомбинации, но включающая полезные и вредные мутации фиксированного эффекта приспособленности, случайный генетический дрейф и эффекты генетического сцепления (§ 3.1). Эволюционные факторы модели включают естественный отбор, случайный генетический дрейф из-за случайной выборки потомства, а также прямые и обратные мутации с одинаковой вероятностью μ на локус на поколение. Эффект мутации на логарифмическую приспособленность либо s (полезный), либо $-s$ (вредный), в зависимости от аллеля на основании. Было проверено, что замена фиксированного s более реалистичным распределением из эксперимента [94] выводов не меняет ([22], рисунок S5).

Следуя обозначениям § 3.1, доля оснований, которые могут иметь полезную мутацию, обозначается α . Таким образом, частота полезных и вредных мутаций на геном есть $\alpha\mu L$ и $(1 - \alpha)\mu L$, соответственно. Цель — вычисление средней частоты замен на геном на поколение $V = V_{ben} - V_{del}$, где V_{ben} и V_{del} обозначают средние скорости замен полезных и вредных мутаций, соответственно. Параметры модели приведены в таблице 6.4.

Таблица 6.4: Входные и выходные параметры модели

Обозначение	Определение	Диапазон	Ссылки
-------------	-------------	----------	--------

L	Количество полиморфных оснований	3000	a b
α	Доля оснований с вредными аллелями	0.08	a b d
μ	Вероятность мутации на основание для штамма дикого типа низконадежного штамма H273R высоконадежного штамма G64S	10^{-4} $3 \cdot 10^{-4}$ $0.15 \cdot 10^{-4}$	a b
s	рибавка/потеря приспособленности	0.2	b
N	Размер популяции (количество зараженных клеток)	$10^3 - 10^4$	a
r	Вероятность рекомбинации на геном	0 – 1	a b
M	Количество кроссоверов на геном	1	a b
$2d$	Начальная гетерозиготность	3	d
T	Общее время (количество поколений)	20	a b c
V_{ben}, V_{del}	Частота полезных (вредных) замен на геном на поколение	результат	
$V = V_{ben} - V_{del}$	Частота замен, чистый выход	результат	

a) Данные эксперименты

b) [93, 94]

c) Использовано 7 пассажей с ~ 5 циклами репликации за 8 часов пассажа. Последние 2 цикла в каждом пассаже исключаются из-за высокой множественности инфекции, поскольку естественный отбор подавляется комплементацией белков между вариантами вируса в клетке.

d) Параметры d и α подгоняются для соответствия данным на рисунке 6.14. Остальные параметры являются оценками из настоящих или опубликованных данных.

Частота замен V в бесполой популяции на больших временах была выведена ранее аналитически для малого коэффициента отбора $s \ll \mu L$ и большого размера популяции, $Ns \gg 1$ [7] (§ 3.1). Было показано, что нормализованная частота замен $V/\mu L$ является функцией составного параметра $(s/\mu L)\log(N/N^*)$ и частоты вредных аллелей α , где $N^* \sim 1/(\alpha\sqrt{s\mu L})$, см. формулу (2.21) с $V = -v\mu L$. Скорость замен V , как функция μL , для значений параметров из таблицы 6.4, демонстрирует наличие максимума при $\mu L \approx 1.5$ (вставка к рисунку 6.14A). Таким образом, как уже сказано, модель бесполой популяции [7] предсказывает наличие скорости мутации, оптимизирующей скорость адаптации, что позже было подтверждено другим автором с помощью численного моделирования [258].

Вывод в работе [7] (§ 3.1) описывает временной диапазон, соответствующий десяткам или сотням поколений, когда бегущая волна в приспособленности устоялась и движется стационарно. Этого еще не происходит при острой инфекции на данной экспериментальной мышинной модели, где число генераций не превышает $t \approx 20$, что соответствует 7 суткам и 3 циклам репродукции в сутки [22]. Модель адаптации при острой инфекции должна учитывать кратковременность процесса. На таких коротких промежутках времени, удобнее использовать численное моделирование методом Монте-Карло.

Был использован алгоритм, разработанный автором ранее [15] для моделирования стохастической эволюции N отдельных геномов в присутствии рекомбинации (§ 2.3). Моделирование подтвердило существование оптимальной скорости адаптации при частоте геномных мутаций $\mu L \approx 0.3$ (рисунок 6.14А), что в 5 раз меньше, чем при длительной эволюции (рисунок 6.14А, вставка). Очень важно, что это оптимальное значение приближенно равно частоте мутаций полиомиелита дикого типа, определенной экспериментально (рисунок 6.11).

Рекомбинация и мутация независимо влияют на адаптацию

Затем в модель бесполой популяции был добавлен эффект рекомбинации. Модели предсказывают, что рекомбинация усиливает вирусную адаптацию многолокусной системы [8-10, 74, 76, 77]. В каждом поколении, случайным образом выбирается часть геномов r для проведения M перекрестов с другим геномом, также выбранным случайным образом. Один из двух возможных рекомбинантов заменяет один из двух родительских геномов. На основе настоящих экспериментов, в среднем происходит один перекрест на геном, $M = 1$. Случай $r = 1$ является предельным случаем сильной рекомбинации, характерным для популяций, размножающихся только половым путем, а $r = 0$ представляет случай бесполого размножения (выше). РНК-вирусы, способные к рекомбинации, включая полиовирус, находятся где-то посередине, так как величина r зависит от частоты коинфицированных клеток и, следовательно, от концентрации вируса [8].

Результаты показывают, что оптимальная частота мутаций μL не изменяется из-за рекомбинации, но все скорости адаптации увеличиваются равномерно для всех скоростей мутаций (рисунок 6.14А). Результаты также подтверждают, что, в краткосрочной эволюции, начальный полиморфизм имеет решающее значение. В вирусах, не имеющих рекомбинации, начальное генетическое разнообразие также важно в краткосрочной эволюции, где новые мутации не играют роли [13].

Начальное распределение числа аллелей по геномам было задано случайным. Полу-гетерозиготность $d = 1.5$ была подогнана, чтобы объяснить данные по вирулентности на рисунке 6.12. Это подогнанное значение находится в пределах диапазона, наблюдаемого для полиовируса в клеточной культуре (рисунок 6.11D), и недалеко от диапазона оцененного для вируса везикулярного стоматита, $d = 0.4 - 0.7$ [13].

Если начальное генетическое разнообразие отсутствует, $d = 0$, эффект рекомбинации при малых частотах мутаций исчезает ([22], рисунок S3), поскольку полезных аллелей для объединения нет, пока не произойдут новые мутации. Второй и последний параметр модели, подогнанный к данным — это доля оснований с вредными аллелями, $\alpha = 0.08$. Только эти основания могут иметь полезную мутацию. Таким образом, два параметра модели d и α найдены

путем подгонки, а остальные параметры ($N = 10^3 - 10^4$, $T = 20$, $s = 0.2$, таблица 6.4) определены прямо из эксперимента (включая проверку на более реалистичное распределение s [94]).

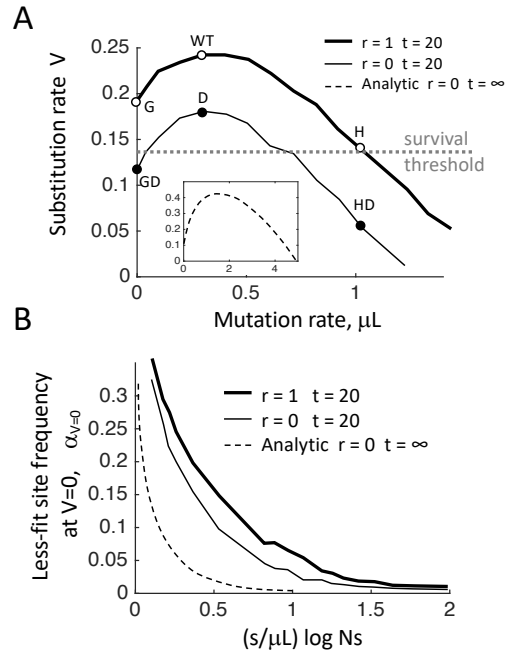


Рисунок 6.14: Скорость эволюции, предсказанная моделью, имеет максимум при оптимальной частоте мутаций, что объясняет различия в смертности мышей между штаммами вируса. (А) Кривые: полная скорость адаптации V , определяемая как количество полезных минус вредных мутаций на геном за цикл репликации, показанная как функция скорости геномных мутаций μL для половых ($r = 1$, толстая кривая) и бесполой ($r = 0$, тонкая кривая) популяций, точка по времени $t = 20$. Моделирование эволюции методом Монте-Карло выполняется с фиксированным s , включая полезные и вредные мутации, на основе программы из работы [15]. Кружки: экспериментальные штаммы вируса, способные вторгаться в ЦНС (над пунктирной линией), неспособные к вторжению (ниже пунктирной линии) или находящиеся на пороге (пунктирная линия), подогнанным для объяснения двойного поведения сильно мутирующего варианта (H), которое зависит от пути заражения (рисунок 6.12). Белые и черные кружки соответствуют рекомбинационно-компетентным и рекомбинационно-дефицитным штаммам, соответственно. Вставка: Скорость адаптации в бесполой популяциях после длительного времени, формула (2.21). (В) Частота вредных аллелей, при которой скорость адаптации V равна нулю, $\alpha_{V=0}$, рассчитанная для разных частот мутаций и нарисованная как функция составного параметра $(s/\mu L) \log(Ns)$, для полового (толстая сплошная кривая) и бесполого (тонкая сплошная кривая) случаев. Пунктирная кривая: равновесие бесполой популяции на бесконечном времени, уравнение (2.22) [также [385], приложение, формула (13)]. Фиксированные параметры: (А) $\alpha = 0.008$; (А, Б) $N = 1000$, $s = 0.2$, $d = 1.5$, $M = 1$, $T = 20$. По работе [22].

Математическое моделирование соответствует данным о выживаемости мышей

Для определения роли мутаций и рекомбинаций в экспериментальном патогенезе, авторы [22] использовали тот факт, что вирус должен измениться генетически, чтобы проникнуть из кишечника (его начальная ткань) в мозг. Отсюда следует, что вирус проходит в мозг, только если он достаточно быстро адаптируется, пока иммунной ответ его не удалит, т. е. если скорость адаптации V превышает некий порог (рисунок 6.14А). При таком разумном предположении, основанном на известных фактах, модель предсказывает правильный порядок патогенеза для всех штаммов. В частности, два способных к рекомбинации штамма (WT и G) имеют V намного выше порогового значения и, следовательно, они убивают всех мышей, как это наблюдалось

экспериментально (рисунок 6.14А). Единственный штамм с дефицитом рекомбинации, скорость адаптации которого выше порога и который может внедряться в мозг, — это штамм с частотой мутаций дикого типа (оптимальной), вариант D. Два двойных мутанта с дефицитом рекомбинации, с высокой или низкой верность (варианты GD и HD), слишком медленно адаптируются к проникновению в мозг и поэтому являются доброкачественными, как это и наблюдается экспериментально (рисунок 6.14А). Высокомутабельный компетентный в рекомбинации мутант, который имеет промежуточную вирулентность (H), должен находиться вблизи порога по V для патогенеза, откуда порог и определялся (рисунок 6.14А). Таким образом, модель объясняет порядок патогенности шести штаммов с разной скоростью рекомбинации и мутации из их упорядочивания по единственной предсказанной величине — их адаптивности. Также эта комбинация модели и эксперимента подтверждают существование ранее предсказанной оптимальной частоты мутаций, объясняет, почему патогенные эффекты вирусов с неоптимальной скоростью мутаций слабее и почему наличие рекомбинации частично компенсирует неоптимальную скорость мутаций.

Чувствительность к параметрам

Также была протестирована чувствительность прогнозируемой скорости адаптации к изменению самых чувствительных параметров s , d и α (S4 Fig in [22]), где d и α подгонялись к данным, а s брался прямо из эксперимента (рисунок 6.11D). Двукратное изменение любого из этих параметров по сравнению с подогнанными значениями (рисунок 6.14А) приводило к какому-то противоречию с экспериментом: либо с данными по вирулентности разных штаммов (рисунок S4 в [22]), либо со относительной скоростью эволюции штаммов в культуре клеток. Таким образом, сделанный выбор параметров d и α уникален и необходим для согласия с экспериментом.

Кроме того, были рассмотрены два предельных случая для частоты рекомбинации r , предполагая либо тотальную рекомбинацию, либо ее отсутствие ($r = 1$ или $r = 0$). Напомним, что значение r зависит от частоты коинфицирования клеток [8]. Чтобы исследовать промежуточные r , моделирование было повторено для нескольких значений r (рисунок S4 в [22]). Для любых значений r не меньше 0.4, предсказания походили на наблюдаемый патогенез. Таким образом, основные выводы мало чувствительны к точному значению r .

6.3.3 Заключение

Как предсказано ранее и подтверждено в данной работе экспериментально, вероятность мутаций вируса полиомиелита эволюционировала миллионы лет, чтобы оптимизировать адаптацию

вируса в организме хозяина и проникновение в мозг. Остается неясным, как это связано с оптимизацией передачей вируса между хозяевами. Возможно что паралич крупных животных, вызываемый проникновением вируса в мозг, приводит к локальной концентрации фекалий облегчающих распространение вируса. Действительно, полиовирус передается из фекалий через рот. Решение этого вопроса выходит за рамки данной работы.

Объединение экспериментальных данных о выживаемости мышей с математическим моделированием объясняет узкий диапазон вероятности мутации на геном, 0.01–0.1, наблюдаемый для многих видов вирусов и патогенов. Модель не поддерживает старую парадигму «катастрофы ошибок», предсказывающую «расплавление» генома выше некоего порога вероятности мутаций [237, 238]. Эта идея основана на устаревших детерминистских моделях эволюции, делающих экзотические предположения о форме адаптивного ландшафта [465, 466]. Реалистичные модели стохастической эволюции предлагают совершенно другое объяснение, а именно, что вирус имеет оптимальную скорость адаптации [7, 258].

Результат того, что вирус эволюционировал для оптимизации вирулентности у мышей, весьма нетривиален из-за эволюционного конфликта между силами отбора, действующими на уровне хозяина и уровне популяции хозяев. Действительно, эволюционная оптимизация вируса происходит в естественной популяции, что поднимает интересные вопросы о связи между патогенными эффектами и числом передачи вируса в популяции. Решение этой проблемы требует отдельного исследования. ВИЧ инфекция представляет собой другой пример, когда приспособленность и вирулентность вируса обе увеличиваются внутри хозяина с течением времени, за исключением того что это происходит на более длинных промежутках времени (§ 6.1).

В целом, направление эволюции вирулентности на уровне популяции хозяев зависит от компромисса между числом размножения и вирулентностью (см. обзор [23] в списке работ диссертации). Дополнительное осложнение — это вероятный эволюционный конфликт между уровнем хозяина и уровнем популяции. Такой конфликт рассмотрен в § 7.4 для другой системы, ВИЧ и его дефектных интерферирующих частиц в группах населения высокого риска.

Глава 7. Наперегонки с враждебным биологическим ВИДОМ

В экологических системах, виды вынуждены адаптироваться к среде, в которой обитают им враждебные и эволюционирующие виды [467]. Поскольку волки и рыси охотятся на зайцев, зайцам приходится приспосабливаться, чтобы бежать быстрее, использовать маневры уклонения и путать свои следы. В свою очередь, хищные виды совершенствуют методы охоты. Такая адаптация организма, сталкивающегося с противоборствующим и постоянно эволюционирующим видом, называется «эффектом Красной Королевы» (рисунок 6.1). Настоящая глава содержит анализ и компьютерное моделирование эволюционного поведения патогена, ускользающего либо от адаптирующейся иммунной системы, либо от своего вторичного паразита. Рассмотрены несколько биологических уровней: клетка, хозяин, и популяция хозяев.



Рисунок 7.1: Эффект Красной Королевы. Эффект носит название Красной Королевы из романа Льюиса Кэрролла «Алиса в Зазеркалье». Как сказала Алисе Красная Королева: «Вот видите, нужно бежать изо всех сил, чтобы оставаться на одном месте». Точно так же, эволюция вируса гриппа на популяционном уровне обусловлена иммунным ответом у людей, выздоровевших от естественной инфекции или вакцинированных, количество которых постепенно накапливается в популяции. Чтобы избежать исчезновения, вирус должен постоянно мутировать, чтобы генетически дистанцироваться от этого иммунного ответа. На основе иллюстрации Джона Тенниела из «Алиса в Зазеркалье» (“Through the Looking Glass, and what Alice Found There”) Льюиса Кэрролла, 1871 г.

7.1 Эволюция респираторного вируса в популяции под давлением гуморального иммунитета

Чтобы избежать иммунного распознавания хозяевами, ранее подвергшимися заражению, вирусы накапливают мутации в иммунологически значимых областях генома [259]. Например,

глобальная персистенция вируса гриппа из-за повторного заражения ранее инфицированных людей вызывается быстрой эволюцией областей связывания антител в белке гемагглютинаина, служащем вирусным антирецептором [259]. В ответ, иммунная система следует по пятам за эволюцией вируса, производя новые лимфоциты, которые распознают эти новые антигенные варианты. Разработка противовирусного лечения и эффективных профилактических мер зависит от предсказания эволюции вируса в масштабах популяции.

До работы [25] рассмотренной ниже, общая (аналитическая) связь между параметрами популяционного уровня, иммунологией и эволюционным поведением вируса была неясна. Аналитический подход [25] объединил стандартный эпидемиологический подход «восприимчивый-инфицированный-выздоровевший» [266, 468] с иммунобиологией [355] и теорией стохастической многолокусной эволюции (*глава 3*). Анализ [25] показал, что вирус непрерывно ускользает от коллективной иммунной памяти, в виде одинокой волны бегущей в адаптивном ландшафте, созданном этой иммунной памятью и движущем вместе с волной. В этой работе, средний коэффициент отбора был выражен, в общем виде, через число репродукции вируса на уровне популяции и «генетическое расстояние перекрестного иммунитета», определенное как количество мутаций в эпитопе вируса, которое приводит к вероятности передачи вируса, равной половине значения для наивного хозяина. Затем проблема сведена к многолокусной теории эволюции описанной в *главе 2*. Предсказания сопоставлены с данными для вируса гриппа А, что позволяет оценить два неизвестных параметра модели и сравнить эти оценки с результатами полученными в экспериментальных животных. Применение данной модели к эволюции эпитопов нейтрализующих антител SARS-CoV-2 обсуждается в обзоре [1], который входит в список работ диссертации.

7.1.1 Модель передачи гриппа в популяции

Эпидемиологическая модель с иммунной памятью, структурированная по штамму

Главная идея [25] состояла в том, чтобы скрестить модели вирусной эволюции с иммунологией и эпидемиологическими моделями, изучающими распространения патогенов в популяции. Вирусы генетически изменяются, чтобы избежать иммунной памяти в ранее инфицированных людях [469]. Предполагается, что все взрослые люди либо инфицированы сейчас, либо были инфицированы вирусом ранее. Все люди классифицируются по последовательности РНК антитело-связывающей области вируса, расположенной в гене гемагглютинаина вируса гриппа. Инфицирующий штамм помечен антигенной координатой x , определяемой как генетическое расстояние, с точки зрения аминокислотных изменений, от исходного штамма 1918 года. После устранения инфекции из организма, человек сохраняет иммунологическую память, которая

обеспечивает полную защиту от того же варианта вируса и частичную вероятностную защиту от генетически близких вариантов. Ниже рассматривается одномерное пространство x , представляющее собой ствол филогенетического дерева. Предполагается, что иммунная система каждого выздоровевшего помнит лишь о самом последнем заражении [266, 267], каковое приближение обосновано в исходной работе [[25], приложение]. Многомерные варианты этой эпидемиологической модели структурированной по штаммам были также рассмотрены [[25], приложение]. Поскольку основное внимание уделяется долгосрочной эволюции, факторы, вызывающие сезонные колебания, пренебрегаются.

Модель также использует тот факт, что эпитопы нейтрализующих антител у всех людей расположены одинаково. Стоимость мутаций предполагается незначительной, в результате длительной эволюции. Это предположение справедливо только для вирусов, вызывающих респираторные инфекции на уровне населения, такие как вирусы обычной простуды, гриппа и, возможно, SARS-CoV-2, а также для эволюции ВИЧ внутри пациентов (только нейтрализующие антитела).

Пусть $i(x, t)dx$ — доля популяции, состоящая из лиц, инфицированных в настоящее время вариантами в интервале $[x, x + dx]$, а $r(x, t)dx$ — доля лиц, ранее инфицированных вариантами $[x, x + dx]$, а затем выздоровевших. В отсутствие мутаций, динамика распределений $i(x, t)$ и $r(x, t)$ описывается обычными дифференциальными уравнениями вида

$$\begin{aligned} \frac{dr(x, t)}{dt} &= -r(x)R_0 \int_x^\infty dy K(x - y)i(y, t) + i(x, t) \\ \frac{di(x, t)}{dt} &= i(x, t) \left[R_0 \int_{-\infty}^x dy K(y - x)r(y, t) - 1 \right] + (\text{mutation}) \end{aligned} \quad (7.1)$$

Каждый человек либо инфицирован, либо выздоровел, как указано в

$$\int_{-\infty}^\infty dx [r(x, t) + i(x, t)] = 1 \quad (7.2)$$

Время в уравнениях (7.1) измеряется в единицах среднего времени между последовательными передачами, что близко к среднему времени выздоровления, t_{rec} (таблица 7.1). Включение мутаций описано ниже.

Эпидемиологические процессы здесь следующие. Во-первых, люди, выздоровевшие от штамма x , могут быть инфицированы штаммом y с вероятностью, пропорциональной функции перекрестного иммунитета $K(x - y)$, которая зависит от генетического (более точно, антигенного) расстояния между штаммами $u = x - y$, так что

$$K(u) > 0, u < 0; K(u) \equiv 0, u > 0; K(-\infty) = 1$$

Предполагается, что люди, выздоровевшие от варианта x , могут быть инфицированы только вариантами, предшествующими x , что определяется соотношением $y > x$. Это приближение заметно упрощает анализ, но незначительно влияет на результаты ([25], приложение). Во-вторых, инфицированные люди выздоравливают с частотой 1. Максимальная скорость передачи равна базовому числу воспроизводства в наивной популяции R_0 .

Включение мутации и случайного генетического дрейфа

Пока что, рассматривалась динамика уже существующих вариантов. Фактически, антигенная эволюция обусловлена появлением новых вариантов вируса. Вариант x иногда подвергается случайной мутации вперед, $x \rightarrow x + \Delta x$, что помогает ему снизить энергию связывания антител и, следовательно, распознавание. Новый штамм гриппа с новой антигенной координатой $x + \Delta x$ может, с некоторой вероятностью, передаться другому человеку.

Соответствующими параметрами модели являются средняя частота мутаций U_b на геном за инфекционный период (таблица 7.1) и распределение эффекта мутации Δx между локусами. Модель предполагает плотность вероятности вида

$$\rho(\Delta x) = \frac{e^{-(\Delta x)^\beta}}{\Gamma\left(1 + \frac{1}{\beta}\right)}, \quad \Delta x > 0 \quad (7.3)$$

где β — фиксированный параметр. Ниже рассматриваются случаи $\beta = 1$ и $\beta = 2$. Подробное обсуждение случая $\beta = 1$ дано в § 3.4. Параметры модели приведены в таблице 7.1.

Таблица 7.1: Входные параметры и предсказанные величины модели, уравнения (7.1-7.3)

Обозначение	Название	Единица измерения	Штамм H3N2	Штамм H1N1
<i>Входные параметры модели</i>				
R_0	Базовое число репродукции	1^b	1.8^a	1.46^a
t_{rec}	Время излечения	день	5^a	5^a
U_b	Частота мутаций на геном	$1/t_{rec}$ 1/год	$5 \cdot 10^{-4}$ 0.036 ^c	$8 \cdot 10^{-4}$ 0.058 ^c
$a = 1/ K'(0) $	Расстояние перекрестного иммунитета	аминокислота	15^c	14^c
$K(u)$	Функция перекрестного иммунитета	1	$\frac{ u }{a+ u }$ или $e^{- u /a}, u < 0$	то же
N	Численность населения	1	10^8	10^8
β	Параметр распределения фитнес-эффекта мутации	1	2	2
<i>Предсказания</i>				
σ	Средний коэффициент отбора	1	0.048^d	0.028^d

$\frac{365N_i}{t_{rec}N}$	Ежегодная заболеваемость	1/год	0.07 ^d	0.04 ^d
c	Средняя частота замен	1/ t_{rec} год	0.036 2.6 ^a	0.031 2.26 ^a
T_{MRCA2}	Среднее время до ближайшего предка пары	год	3.03 ^a	4.59 ^a

^a Известно из опубликованных данных о штаммах гриппа А Н3N2 и Н1N1 [262, 268, 470, 471]

^b 1 означает «безразмерный».

^c Входной параметр модели подогнанный к опубликованными данным по c и T_{MRCA2}

^d Прогнозируемое значение для лучшей подгонки параметров.

Эпидемиологическая и эволюционная динамика в этой модели анализируется в два этапа следующим образом:

Шаг 1. Ищется решение в виде, согласно которому, в диапазоне параметров $a \gg 1$, индивидуальная плотность инфицированных $i(x, t)$ представляет собой одиночный пик по x , узкой по сравнению с плотностью выздоровевших $r(x, t)$. Этот *анзац* используется для вывода $r(x, t)$.

Шаг 2. Для получения определенной формы распределения зараженных особей $i(x, t)$, применяется стандартная теория бесполой эволюции (§§ 3.1 или 3.2).

7.1.2 Двухкомпонентная бегущая волна

Ниже изложены основные результаты, подробные выводы приведены в исходной работе [25]. Во-первых, вначале можно пренебречь мутацией и искать решение уравнения (7.1) в форме бегущей волны

$$r(x, t) = r(x - ct)$$

$$i(x, t) = i(x - ct)$$

где $x - ct \equiv u$ — координата антигена в подвижной системе отсчета, c — средняя скорость замен. Так как система обладает трансляционной инвариантностью, максимум $i(u)$ можно выбирать где угодно, например в нуле: $[di/du]_{u=0} = 0$.

Плотность выздоровевших

Подставляя приведенный выше анзац в (7.1) и решая полученное ОДУ, для плотности зараженных и выздоровевших имеем

$$i(u) = A c f(u)$$

$$r(u) \approx \begin{cases} A \exp(-AR_0 \int_u^0 dv K(v)) & u < 0 \\ 0 & u > 0 \end{cases} \quad (7.4)$$

где константа A определяется из (7.2), а $f(u)$ считается дельта-функцией, т.е., узким пиком неизвестной формы, с единичным интегралом и шириной значительно меньшей ширины $r(u)$. Скорость бегущей волны c и форма $i(u)$ будут рассмотрены ниже в § 7.1.3.

При больших значениях R_0 (численно работает уже при весьма небольших R_0), $K(v)$ в подынтегральном выражении выше можно аппроксимировать линейным разложением при $v = 0$. После интегрирования, восстановленная индивидуальная плотность становится полугауссовой

$$r(u) \approx \begin{cases} \frac{2R_0}{\pi a} e^{-\left[\frac{R_0 u}{a\sqrt{\pi}}\right]^2} & u < 0 \\ 0 & u > 0 \end{cases} \quad (7.5)$$

и $A = 2R_0/(\pi a)$. Отсюда доля инфицированных в населении определяется выражением

$$\frac{N_{inf}}{N} \equiv \int_{-\infty}^{\infty} du i(u) = Ac = \frac{2R_0 c}{\pi a} \ll 1 \quad (7.6)$$

а годовая заболеваемость

$$\text{Annual incidence} = \frac{2R_0 c}{\pi a} \cdot \frac{365}{t_{rec}} \quad (7.7)$$

что можно сравнить с данными наблюдений.

Аналитическое решение (7.4) и (7.5) получено из предположения, что пик зараженных $i(u)$ значительно уже максимума выздоровевших $r(u)$. Это предположение и формула (7.5) были проверены моделированием Монте-Карло, основанным на уравнениях (7.1) со стохастическим членом случайных мутаций, при типичных значениях параметров. Моделирование подтверждает анализ и демонстрирует бегущую волну с двумя компонентами (рисунок 7.2), где пик плотности зараженных $i(u)$ гораздо уже пика выздоровевших $r(u)$. Усредненная по времени плотность выздоровевших $r(u)$ согласуется с аналитическим результатом (7.5). Как и положено, она имеет резкую ступеньку на фронте волны, на пике $i(u)$, и длинный хвост позади, при $u < 0$. Ступенька обусловлена выздоровлением инфицированных особей. Убывание длинного хвоста вызвано повторным заражением выздоровевших особей, генетически (антигенно) удаленных от зараженного пика.

Движущийся адаптивный ландшафт

Распределение инфицированных $i(u)$ а также скорость эволюции находятся из независимого аргумента путем соединения данной модели SIR со стандартной теорией бегущей волны [11, 16]. Отличие от стандартной теории в том, что адаптивный ландшафт здесь не фиксирован, а движется вместе с волной. Кроме того, адаптивный ландшафт не задан извне, а должен быть получен из «первых принципов», т.е. эпидемиологической модели (7.1). Распределение иммунной памяти в генетическом пространстве $r(u)$ создает адаптивный ландшафт, который и толкает бегущую волну.

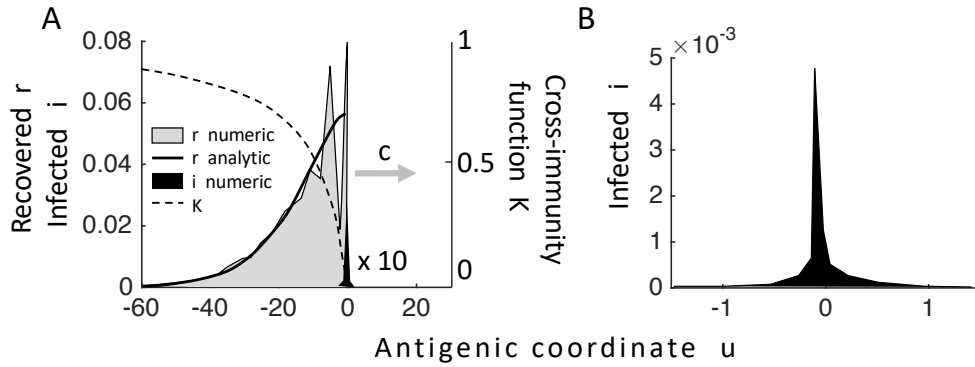


Рисунок 7.2: Эпидемиологическая модель предсказывает устойчивую волну бегущую вдоль оси приспособленности. (А) Частота выздоровевших и инфицированных в популяции в системе отсчета, движущейся вместе с волной. Здесь ось абсцисс представляет собой антигенную координату в этой системе отсчета, $u = x - ct$. Сплошная линия показывает аналитический предсказание для $r(u)$, формула (7.5). Небольшой черный пик — это распределение инфицированных особей $i(u)$. Серая область показывает результат стохастического моделирования эпидемиологической модели по уравнениям (7.1)-(7.2) плюс случайные мутации (7.3). Штриховая линия — функция перекрестного иммунитета $K(u)$ (таблица 7.1). (В) Инфицированная индивидуальная плотность $i(u)$. Параметры: $R_0 = 2$, $a = 9$, $U_b = 5.8 \cdot 10^{-6}$ на эпитоп на передачу, $N = 10^8$. По работе [25].

Приспособленность вируса в популяции хозяина (число репродукции) определяется как среднее число вторичных инфекций на одного инфицированного человека [472-475]. Число воспроизводства меньше, чем его значение в наивной популяции, R_0 , потому что оно уменьшено иммунной памятью популяции. Вирус размножается, только если число репродукции больше единицы. Логарифм приспособленности $w(x, t)$ может быть определен как экспоненциальная скорость роста плотности инфицированных $i(x, t)$ на период передачи вируса. Используя второе уравнение в (7.1), получается

$$w(x, t) = \frac{\partial \log i(x, t)}{\partial t} = R_0 \int_{-\infty}^x dy K(y - x)r(y, t) - 1$$

Очевидно, что адаптивный ландшафт перемещается во времени: поскольку $r(y, t)$ представляет собой бегущую волну, а K зависит только от разницы $y - x$, функция $w(x, t)$ является бегущей волной с той же скоростью. В движущейся системе отсчета, привязанной к волне, для путешествующего адаптивного ландшафта получается

$$w(u) = R_0 \int_{-\infty}^u dv K(v - u)r(v) - 1 \quad (7.8)$$

Эта величина положительна перед пиком зараженной плотности, $u > 0$, и отрицательна за пиком, $u < 0$. Результат для $w(u)$, полученный из численного решения уравнений (7.4) и (7.8), показан на рисунке 7.3. Асимптотические случаи таковы:

$$w(u) \approx \begin{cases} R_0 - 1 & u \gg a \\ \sigma & |u| \ll a \\ -1 & u < 0, \quad |u| \gg a \end{cases} \quad (7.9)$$

где введено новое обозначение

$$\sigma \equiv -R_0 \int_{-\infty}^0 du r(u) \frac{dK(u)}{du} \quad (7.10)$$

представляющее собой наклон адаптивного ландшафта в нуле. Равенство $w(0) = 0$ означает, что скорость роста инфицированных равна нулю на пике, как и должно быть; $w(u)$ имеет тот же знак, что и u . Для больших абсолютных значений u , таких, что $|u| \gg a$, адаптивный ландшафт $w(u)$ выходит на плато, формула (7.9). В области $|u| \ll a$, где расположен пик заболеваемости, адаптивный ландшафт можно аппроксимировать линейным приближением с положительным наклоном σ .

Если средняя мутация изменяет антигенную координату на единицу, то σ представляет собой среднее изменение логарифма приспособленности w из-за события мутации, т. е. средний коэффициент отбора. Для достаточно большого R_0 , из формул (7.5) и (7.10), σ можно аппроксимировать рядом Тейлора по $1/R_0$

$$\sigma(a, R_0) = \frac{1}{a} \left[R_0 - \xi_1 + \frac{\xi_2}{R_0} + O\left(\frac{1}{R_0^2}\right) \right] \quad (7.11)$$

где $a \equiv 1/|K'(0)|$, а второе и третье слагаемые являются, формально, малыми поправками к первому слагаемому, если R_0 велико. Таким образом, среднее давление отбора обратно пропорционально расстоянию перекрестного иммунитета a и увеличивается с базовым числом репродукции R_0 . Его значение предполагается малым, $\sigma \sim \frac{1}{a} \ll 1$.

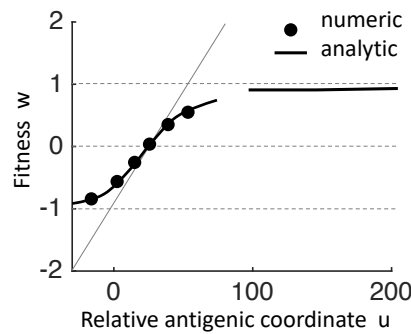


Рисунок 7.3: Адаптивный ландшафт для путешествий и его линейная аппроксимация вблизи инфицированного пика. Сплошная кривая: аналитический результат уравнения. (7.8). Черные кружки: моделирование Монте-Карло на основе (7.1)-(7.3). Тонкая линия: линейная аппроксимация со средним коэффициентом отбора $\sigma = 0.066$ из формулы (7.9). Параметры как на рисунке 7.2: $R_0 = 2, a = 9, U_b = 5.8 \cdot 10^{-6}, N = 10^8$. По работе [25].

Численные коэффициенты ξ_1 и ξ_2 в формуле (7.11) зависят от вида функции перекрестного иммунитета $K(u)$ (табл. 7.1). Для медленно затухающей функции перекрестного иммунитета в таблице 7.1 они равны $\xi_1 = 2$ and $\xi_2 = 3\pi/\sqrt{2}$. Для экспоненциально убывающей функции перекрестного иммунитета $K(u) = 1 - \exp(-|u|/a), u < 0$, они равны $\xi_1 = 1$ and $\xi_2 = \pi/2\sqrt{2}$. Адаптивный ландшафт $w(u)$, рассчитанный по смоделированной плотности выздоровевших (7.8), согласуется с аналитическим результатом (рисунок 7.3).

7.1.3 Связь с теорией эволюции

Антигенное разнообразие и скорость эволюции

Выше скорость вирусной эволюции s оставалась неопределенной, и следующая задача состояла в том, чтобы его получить. После того, как средний коэффициент отбора σ был получен в формуле (7.11), эпидемиологическая проблема свелась к эволюции бесполой популяции со многими эволюционирующими локусами. Для таких систем, скорость эволюции может быть выражена через размер популяции, коэффициент отбора и скорость мутаций [7, 11, 12, 16, 72, 73] (глава 3).

В общем случае, коэффициент отбора $s = \sigma \Delta x$ как-то распределен по локусам. Предположим, что антигенное расстояние Δx выбирается из распределения (7.3) с большим параметром β . Дисперсия антигенной координаты $Var[x] = \langle (\Delta x)^2 \rangle$ и скорость адаптации v имеют вид [16]

$$Var[x] = \frac{2 \log(N_{\text{inf}} \sigma)}{\log\left(\frac{\beta \sigma}{U_b}\right)} \quad (7.12)$$

$$v = \sigma^2 Var[x] \quad (7.13)$$

В дополнение к скорости адаптации, определяемой как среднее изменение приспособленности в единицу времени, другой мерой скорости эволюции является средняя частота замен c (в главах 2 и 3, обозначенная V)

$$c = \frac{v}{s^*} = \frac{\sigma^2}{s^*} Var[x] \quad (7.14)$$

$$s^* \equiv \sigma \left(\frac{\sqrt{2}}{\beta} \log \frac{\sigma}{U_b} \right)^{\frac{1}{\beta-1}} \quad (7.15)$$

где $s^* \gg \sigma$ — наиболее вероятный коэффициент отбора фиксированной мутации [16]. Выражения (7.12) и (7.15) приближительны, так как логарифмические члены внутри больших логарифмов пренебрегаются.

Чтобы связать эти результаты с рассматриваемой проблемой, средний коэффициент отбора σ подставляется из формулы (7.10), а размер зараженной популяции N_{inf} — из (7.6). В результате, оба показателя скорости эволюции c и v выражаются через расстояние перекрестного иммунитета a и эпидемиологические параметры (таблица 7.1). В пределе очень большого β , когда плотность распределения коэффициента отбора (7.3) быстро затухает, формулы (7.12)–(7.15) совпадают с результатами модели с постоянным коэффициентом отбора σ [11].

Аналитический результат для скорости волны (7.14) (рисунок 7.4, ломаные линии) был протестирован численно методом Монте-Карло в широком диапазоне значений N и U_b (рисунок 7.4, символы). Были использованы два численных метода: (i) полное моделирование исходной модели, определяемой (7.1) включая случайные мутации с распределением (7.3), но не включая случайный генетический дрейф, и (ii) моделирование на основе алгоритма Морана [368] с фиксированным размером популяции, коэффициентом отбора (7.10), случайными мутациями и случайным дрейфом. В обоих методах, эффект приспособленности случайно возникающей мутации был получен из случайного распределения (7.3).

Оба численных метода моделирования предсказали сходные скорости эволюции c (рисунок 7.4). Результаты моделирования подтвердили аналитическими результатами и воспроизвели медленный рост скорости эволюции с N и U_b , за исключением самых малых значений. Сокращения IS, CI, MM на рисунке 7.4 указывают на режим эволюции по отношению к числу одновременно полиморфных локусов. В небольших популяциях, фиксация аллелей происходит по одиночке (IS) или по парам с клональной интерференцией меж ними (CI) [16, 65, 81]. В больших популяциях, большое количество локусов меняется одновременно (MM) [7, 11, 12, 16, 72]. Результаты (7.12)-(7.15) получены для режима MM, что и объясняет расхождение моделирования и анализа при самых малых U_b и N . Крутизна распределения эффекта мутации β на скорость эволюции в (7.3) влияет слабо.

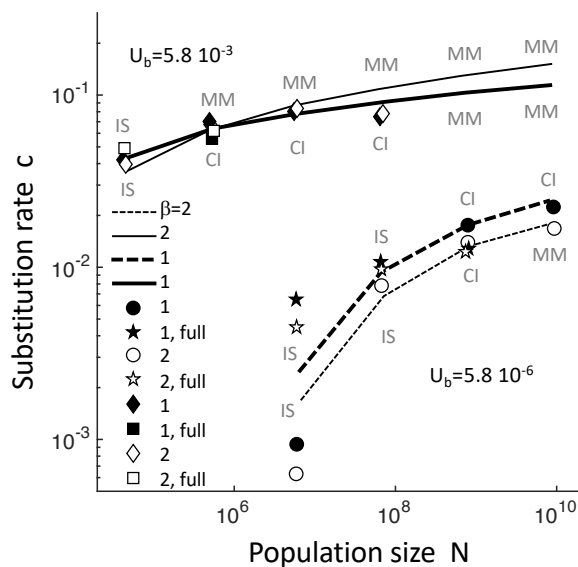


Рисунок 7.4: Стохастическое моделирование подтверждает аналитические результаты для скорости эволюции. Четыре пунктирные линии представляют собой аналитические результаты для скорости волны рассчитанные по формулам (7.14) и (7.15) при двух предельных значениях частоты мутаций U_b для РНК-вирусов и двух значениях параметра β распределения коэффициента отбора (7.3). Символы показывают результаты, полученные двумя методами стохастического моделирования: полное стохастическое моделирование эпидемиологической модели (7.1), и упрощенное моделирование типа Морана с размером популяции N_{inf} и коэффициентом отбора $\sigma = 0.066$ вычисленными аналитически из эпидемиологической модели. Серые буквы обозначают режим эволюции: изолированная фиксация (IS), множественные мутации (MM), парная клональная интерференция (CI). Фиксированные параметры: $R_0 = 2$, $a = 9$, U_b и β показаны. По работе [25].

Приведенный выше анализ и моделирование показывают, что средний коэффициент отбора σ обратно пропорционален a , формула (7.11), что скорость антигенного ускользания s также обратно пропорциональна a и очень медленно (логарифмически) увеличивается с размером популяции N и частотой мутаций U_b , формула (7.14), и что скорость адаптации v обратно пропорциональна a^2 , формула (7.13). Годовая заболеваемость (7.7) также меняется как $1/a^2$.

Время до самого последнего общего предка

Другой важной наблюдаемой величиной является время до появления самого последнего общего предка двух сосуществующих вирусов. Этот параметр был получен аналитически для различных моделей бесполой популяций [89, 90, 257, 390]. Их результаты похожи и можно представить в виде

$$T_{MRCA2} = z \sqrt{\frac{2 \log(N\sigma)}{v}} \quad (7.16)$$

где численный коэффициент $z \sim 1$ зависит от деталей модели, таких как форма распределения коэффициента отбора. В том случае, когда он фиксирован, $z = 1.5$. Для распределения Гаусса (7.3) при $\beta = 2, z = 3$. Поскольку случай Гаусса более реалистичен чем фиксированный коэффициент отбора, а для других форм распределения выражение для T_{MRCA2} было недоступно, то для сравнения с данными по гриппу, было выбрано значение $\beta = 2$.

7.1.4 Сравнение с данными по гриппу А

Последним этапом было сравнение теоретических результатов с имеющимися данными по двум штаммам гриппа, А Н3Н2 и А Н1Н1. Входные параметры модели и предсказываемые параметры находятся в таблице 7.1. Размер популяции N , число репродукции в наивной популяции R_0 , а также период передачи вируса (близкий к времени выздоровления) t_{rec} находятся по данным наблюдений [262, 268, 470, 471]. Однако параметры a и U_b трудно измерить, потому что они включают биологические взаимодействия на нескольких биологических уровнях: клетка, хозяин и популяция хозяев. С другой стороны, две величины, предсказанные моделью, T_{MRCA2} и скорость замен s , известны. Следовательно, неизвестные параметры модели a и U_b , были подогнаны, чтобы соответствовать известным значениям предсказанных величин T_{MRCA2} и s (рисунок 7.6А).

Годовая заболеваемость гриппом А (Н3Н2) колеблется в интервале 1-15% населения. Средняя скорость эволюции и T_{MRCA2} составляют $s = 0.036$ аминокислотных

замен/геном/передача с периодом передачи 5 дней и $T_{MRCA2} = 3.0$ года, соответственно [262, 471]. Число населения большой страны составляет порядка $N = 10^8 - 10^9$. Число репродукции в наивной популяции $R_0 = 1.8$ аппроксимировано его средним значением, измеренным для самых мощных пандемий, 1918-го и 1968-го годов. Такая крупная пандемия обычно вызывается рекомбинацией вирусных хромосом или другими формами антигенного сдвига, вызывая плохое иммунное распознавание, что приближается к наивной популяции.

Общая длина интересующей нас области — $L = 120$ несинонимичных нуклеотидов, которые являются первым и вторым нуклеотидами в 60 кодонах пяти областей белка гемагглютинина, связываемых нейтрализующими антителами [268, 270]. Частота мутаций могла бы быть оценена по частоте синонимических замен, которая составляет $5.8 \cdot 10^{-5}$ на нуклеотид на передачу [268]. Однако не все аминокислоты одинаково важны для связывания антител, и не все мутации дешевы для вируса. Следовательно, ожидается, что эффективная длина эпитопа будет ниже 120. Таким образом, определить U_b из эксперимента трудно, и его величина обычно оценивается весьма приблизительно [262, 267]. Следовательно, параметры a и U_b подогнаны так, чтобы соответствовать предсказаниям данной модели для c и T_{MRCA2} для двух штаммов гриппа, H2N3 и H1N1. Подогнанные значения U_b и a одинаковы для двух штаммов, хотя штамм H2N3 эволюционирует быстрее и имеет более короткое время T_{MRCA2} , чем штамм H1N1, из-за большего инфективности и числа воспроизводства R_0 , что приводит к большему давлению естественного отбора σ .

Полученная оценка $a = 14-15$ мутаций подтверждается независимыми данными по гриппу в лошадях [476]. Прогнозируемая годовая заболеваемость (4–7)% также находится в пределах экспериментально наблюдаемого диапазона, 1–15%, и оценок численного моделирования другими авторами [267, 268, 270]. Результат подгонки $U_b = 3 \cdot 10^{-4}$ в 3.3 раза больше, чем оценка в работе [267].

Приведенные выше результаты [25] объясняют обратную корреляцию, наблюдаемую между T_{MRCA2} и частотой замен c в штаммах H2N3, H1N1 и вирусе гриппа В [262]. Причина корреляции заключается в том, что предсказанная частота замен c линейно пропорциональна коэффициенту отбора σ , а T_{MRCA2} обратно пропорционален пропорциональна σ . Широкая вариация σ и, следовательно, c и T_{MRCA2} между штаммами вызвана, в основном, широкой вариацией числа репродукции R_0 по формуле (7.10). Остальные параметры или почти постоянны, или слабо влияют на межштаммную вариацию. Зависимость c и T_{MRCA2} от параметров U_b и N логарифмически медленная, а подогнанный параметр a мало меняется между штаммами [25].

7.1.5 Многомерное пространство и старая иммунная память

Приведенный выше анализ основан на следующих приближениях:

- (i) Сигмоидальный адаптивный ландшафт аппроксимируется своим линейным разложением
- (ii) Заражение вперед по антигенной (генетической) координате не допускается.
- (iii) Учитывается иммунная память только о последнем штамме [266, 267].
- (iv) Антигенное (генетическое) пространство одномерно.

Как показано в оригинальной работе ([25], приложение), эти приближения мало влияют на предсказания. Хотя антигенное пространство области связывания антител многомерно. Наличие большего количества дополнительных измерений не сильно меняет результаты. Похожая модель была моделирована численно в пространстве с топологией дерева ([25], рисунок S6), что дало квазиодномерную бегущую волну, состоящую из длинного стебля фиксированных мутаций и коротких ветвей с проходными вариантами вируса, как и наблюдаемая филогенетика гриппа А [259, 267, 268, 270]. Автоматическое формирование бегущей волны в виде «змейки» наблюдалось и при моделировании в двумерном генетическом пространстве с одной или двумя антигенными координатами ([25], рисунок C5). Причина, по которой антигенное пространство может быть сведено к дереву, заключается в том известном факте, что возвращение пути случайного блуждания в начало координат в m -мерном пространстве имеет малую вероятность порядка $1/m \ll 1$. По той же причине, в 1D топологии вирусного пути, старые ячейки памяти пренебрежимо малы ([25], приложение).

7.1.6 Заключение

Рассмотрена модель эволюции вируса в населении, предполагающая пренебрежимо малую стоимость мутаций и фиксированное расположение эпитопов в геноме. Эти предположения соответствуют некоторым респираторным вирусам, таким как вирус гриппа, обычная простуда, коронавирусы. Иммунная память популяции создает адаптивный ландшафт, который выражается через параметры перекрестного иммунитета и базовое число передачи вируса в наивном населении. В результате, получается бегущая волна в пространстве антигенных вариантов с двумя компонентами населения: выздоровевшие и инфицированные. Эти два компонента волны имеют разные размер и ширину (рисунок 7.2). Инфицированная подгруппа невелика по размеру и куда менее разнообразна генетически (по вирусу).

Скорость эволюции вируса, ежегодная заболеваемость и среднее время до недавнего предка выражаются через параметры модели $N, U_b, R_0, K(x - y)$ (табл. 7.1). Коэффициент отбора и коэффициент замен пропорциональны $f(R_0)/a$, где $f(R_0)$ — монотонно возрастающая

функция с $f(0) = 0$. Все наблюдаемые слабо зависят от размера популяции N и частоты мутаций U_b , что является универсальным свойством мультилокусных моделей [3]. Два типа моделирования Монте-Карло подтверждают эти предсказания. На основе эпидемиологических и геномных данных по вирусу гриппа А, эти теоретические результаты обеспечивают оценку важных параметров a и U_b , которые согласуются с экспериментальными наблюдениями в лошадях. Кроме того, модель объясняет обратную пропорциональность между временем до ближайшего предка вируса и частотой замен, наблюдаемую в разных штаммах гриппа.

Эти аналитические результаты и оценки также согласуются с предыдущим численным моделированием, основанным на аналогичной модели [267]. Позже, они были подтверждены аналитически двумя группами исследователей [271, 272]. Хотя эти исследователи использовали более грубое приближение для коллективного иммунитета, предполагая, что все население имеют общую иммунную память, как если бы это были бактерии [468], они получили выражения для эффективного коэффициента отбора σ , подобные (7.11), но с другой функцией $f(R_0)$. Один из артефактов «бактериального» приближения к иммунитету человека — огромное преувеличение роли старых клеток памяти, которые остались от ранних инфекций [271]. На самом деле, старая память никакой роли не играет [25], поэтому старые вакцины, как всем известно, не работают.

Подводя итог, сочетание стандартного эпидемиологического подхода с теорией мультилокусной эволюции (глава 3) и правильной иммунологией позволяет связать воедино эпидемиологические и иммунологические параметры с наблюдаемыми параметрами эволюции гриппа.

7.2 Эволюция ВИЧ в хозяине под давлением клеточного иммунитета

В контроле репликации ВИЧ внутри хозяина, основным фактором являются CD8 Т клетки, цитотоксические лимфоциты (CTL), которые убивают инфицированные клетки и снижают репликацию вируса с помощью секреции цитокинов. Вирус избегает полной вычистки и устанавливает персистирующую хроническую инфекцию за счет комбинации двух методов: путем уничтожения хелперных CD4 Т-клеток и, таким образом, снижения чувствительности иммунного ответа [174, 175, 224, 273, 274], а также за счет появления мутаций в антигенно-важных областях, эпитопах, распространяясь по инфицированным клеткам. В результате, вирус становится частично устойчивым к иммунному ответу. Именно этот процесс эволюционного

ускользания ВИЧ, вместе с компенсирующими мутациями, запускает эволюцию ВИЧ в пациентах (§ 2.1) и в конце приводит к СПИД (§ 7.1).

Знание факторов, контролирующих порядок мутированных локусов, полезно для прогнозирования консервативных эпитопов и разработки эффективной стратегии вакцинации, которая до сих пор отсутствует именно из-за быстрой эволюции этого вируса. С этой целью, в работе [18] разработана математическая модель вирусной динамики, управляемой давлением отбора множества клонов цитотоксических лимфоцитов (CTL), которые сами изменяются во времени в ответ на стимуляцию антигеном (эпитопом, то есть пептидом вируса связывающим рецептор Т клетки, TCR). Результаты [18], рассмотренные ниже, показывают, что процесс антигенного ускользания регулируется компромиссом между ценой и выгодой ускользающих мутаций. Потеря распознавания вируса иммунной системой из-за мутаций только частична, и этот факт важен для предсказания порядка фиксирования мутаций ускользания и, в конечном счете, для предсказания консервативных эпитопов, на которые могут быть нацелены вакцины. Предсказания модели [18] использованы для объяснения корреляции между потерей распознавания и стоимостью мутации наблюдаемых в пациентах в гене полимеразы [295], а также для оценки диапазона потерь распознавания в трех экспериментальных исследованиях [296-298]. Частичная природа ускользания также объясняет различные временные последовательности мутаций, наблюдаемые у пациентов, такие как смена доминантного эпитопа с течением времени и невложенная последовательность мутаций («чехарда») [282, 356, 477].

7.2.1 Модель динамики ВИЧ в присутствии множественных эпитопов

Модель включает динамику нескольких вирусных штаммов в присутствии нескольких клонов CTL, как показано на диаграмме (рисунок 7.7А). Она формализуется системой ОДУ

$$\frac{dT}{dt} = \lambda - d_T T - \beta T \sum_i f_i I_i \quad (7.17)$$

$$\frac{dI_i}{dt} = \beta T f_i I_i - d_i I_i - \kappa I_i \sum_j r_{ij} E_j \quad (7.18)$$

$$\frac{dE_j}{dt} = \sigma + c E_j \sum_i \frac{(r_{ij} I_i)}{h_j + \sum_i r_{ij} I_i} - d_E E_j \quad (7.19)$$

где $f_i < 1$ и $r_{ij} < 1$ — относительная скорость репликации штамма i и его относительное узнавание клоном CTL с номером j по сравнению с исходно переданной последовательностью, соответственно.

Модель отталкивается от динамических моделей введенных в §§ 7.1 и 7.2 и включает в себя следующие процессы. Клетки-мишени, число которых обозначено как T , пополняются с

линейной скоростью λ клеток в день, умирают (или покидают высокоинфицируемую фазу) со скоростью d_T и инфицируются со скоростью, пропорциональной вирусной приспособленности и количеству активно инфицированных клеток I_i в системе. Каждый геном содержит n эпитопов (участков распознаваемых одним клоном CTL), каждый из которых соединяет m аминокислот (локусов), что соответствует 2^{mn} возможных штаммов. Здесь, как и везде в диссертации, предполагается присутствие только двух аллелей с *близкой приспособленностью* на локус, т.е. двух вариантов аминокислоты из 10 возможных, химически различных вариантов; всего разных аминокислот 20. Изредка в вирусной популяции наблюдаются 3 варианта аминокислоты из-за разных мутаций в одном кодоне, каковым редким событием пренебрегается.

Инфицированная клетка нумеруется интегрированным провирусом i , содержащим n эпитопов: $g^i = [e_1^i, e_2^i, \dots, e_n^i]$, где эпитоп с номером j — бинарная последовательность $e_j^i = [a_{j1}^i, a_{j2}^i, \dots, a_{jm}^i]$, и $a_{jk}^i = 0, 1$ указывает на наличие или отсутствие мутации в эпитопе j генома i на позиции внутри эпитопа $k = 1, 2, \dots, m$.

Мутация в эпитопе j на позиции k уменьшает логарифм скорости репликации на коэффициент отбора («стоимость мутации») s_{jk} , что учитывается в уравнении (7.18) фактором

$$f_i \equiv \exp\left(-\sum_j s_{jk} a_{jk}^i\right) \quad (7.20)$$

предполагающим отсутствие эпистаза (*глава 5*). По определению, штамм со всеми $s_{jk} = 0$ является переданным начальным штаммом с приспособленностью 1. Формула (7.20) предполагает, что мутация имеет логарифмически аддитивный эффект. Другими словами, он игнорирует эпистаз, обсуждавшийся в *главе 2*.

Предполагается, что клон активированных CTL, число которых обозначено E_j , пополняется со скоростью σ , размножается со скоростью, пропорциональной количеству инфицированных клеток, экспрессирующих родственный эпитоп j , «авидности клона» $1/h_j$, но не выше предельной скорости пролиферации c , уравнение (7.19). CTL умирают со скоростью d_E и убивают инфицированные клетки с эффективностью k . Таким образом, на основе существующих данных, модель предполагает, что зараженные клетки гибнут как от CTL так и вирусной цитотоксичности ([18], методы). Хотя модель не включает подавления производства вируса цитокинами, которое на самом деле существует, это упрощение не меняет выводов работы [18]. Мутация в эпитопе j в основании k снижает логарифм узнавания соответствующим клоном CTL на относительную величину α_{jk} . Это учитывается следующим выражением для степени узнавания генома i клоном j

$$r_{ij} \equiv \exp\left(-\sum_k \alpha_{jk} a_{jk}^i\right) \quad (7.21)$$

Далее, для простоты предполагалось, что avidности разных CTL клонов равны, $h_j \equiv h$. Потери приспособленности и узнавания ниже обозначаются $\Delta r_{ij} \equiv 1 - r_{ij}$ и $\Delta f_i \equiv 1 - f_i$.

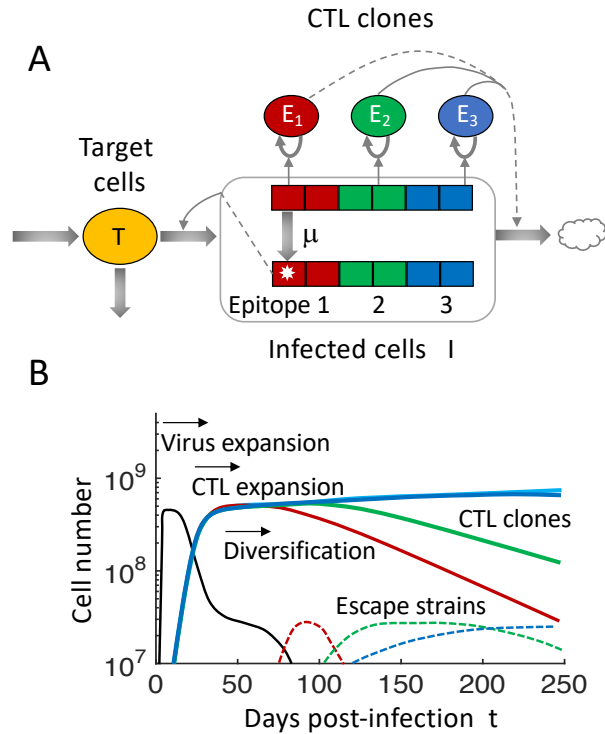


Рисунок 7.5: Компьютерная модель взаимодействия между ВИЧ и несколькими клонами CTL. (А) Модель (7.17)-(7.21) включает три взаимодействующих клеточных компартмента: клетки-мишени T , инфицированные клетки I и множественные клоны CTL, E . Вирусные геномы содержат несколько эпитопов, которые могут мутировать, частично блокируя распознавание CTL. Ускользящая мутация обозначается звездочкой. Каждый клон CTL распознает один вирусный эпитоп и стимулируется к делению со скоростью, пропорциональной количеству инфицированных клеток с распознаваемыми эпитопами. Широкие серые стрелки: поток клеток из одного компартмента в другой. Тонкие стрелки: зависимость потока от третьего компартмента. Пунктирные линии представляют ослабление силы взаимодействия. (В) Три фазы эволюции ВИЧ предказанные моделью. Передаваемый штамм вируса инициирует инфекцию (черная кривая). В ответ на растущее число инфицированных клеток, активируются несколько клонов CTL (сплошные цветные линии), и система достигает устойчивого состояния. Наконец, штаммы вируса с ускользящими мутациями (штриховые цветные линии) замещают передающийся штамм. В ответ на пониженные сигналы активации, некоторые клоны CTL угасают. Ускользящие штаммы окрашены как последний клон CTL, от которого штамм ускользнул. Параметры: количество эпитопов, $n = 6$; количество локусов на эпитоп, $m = 1$. Эпитопы 1-3 имеют параметры, разрешающие ускользание $\Delta r_i = [0.1, 0.2, 0.3]$, $\Delta f = 0.01$, эпитопы 4-6 имеют параметры, запрещающие ускользание, $\Delta f_i = \Delta r_i = 0.1$. Остальные параметры в таблице 7.2. По работе [18].

К данной модели, добавляются случайные мутации с вероятностью $\mu = 3 \cdot 10^{-5}$ на основании на поколение инфицированных клеток (временной +период $1/d_I$) [56]. Поскольку общее количество всех типов клеток в острой и хронической фазе очень велико, $> 10^8$, динамика моделируется детерминистически, с низким порогом в одну зараженную клетку, ниже которого штамм считается вымершим. Параметры модели и расчетный диапазон приведены в таблице 7.2.

Упрощенная модель для изучения порядка ускользящих мутаций

В дополнение к основной модели, удобно ввести упрощенную модель-компаньон, которая явно не рассматривает динамику, но легко предсказывает траекторию (порядок) фиксации мутаций ускользания на фазовой диаграмме $(\Delta f, \Delta r)$. Значения стоимости приспособленности Δf и потери распознавания Δr выбираются случайным образом из равномерного распределения $[0,1]$ для $n = 10$ эпитопов, с $m = 10$ локусами на эпитоп. Локусы ранжированы в порядке разницы $\Delta r - n\Delta f$. Когда CTL клон уменьшится в ответ на ускользание локуса в эпитопе, иммунное давление на все другие локусы в этом эпитопе тоже снижается. Следовательно, сокращение числа CTL, формально описываемый уравнением (7.19), можно приблизительно представить следующим образом: Δr уменьшается для всех оснований в эпитопе как

$$\Delta r(n + 1) = \Delta r(n) \exp \left[-d'_E \sum_i \Delta r_i(0) \right]$$

для всех оснований i в эпитопе, которые уже ускользнули. Здесь параметр d'_E определяется как скорость сокращения на мутацию, в отличие от d_E в основной модели, которая определяется за сутки.

Таблица 7.2. Параметры модели

Параметр	Значение	Определение	Ссылка
n	1 – 8	Количество эпитопов, распознаваемых в течение первых 100 дней	[282, 283, 356, 477]
m	2 – 9	Количество оснований на эпитоп, важных для распознавания	[282, 283, 356, 477]
d_T	1.0/день	Частота, с которой активированные клетки-мишени покидают высокоинфекционную фазу	[375, 435, 444]
λ/d_T	$5 \cdot 10^8$ клеток	Уровень активированных целевых клеток	
d_I	1.0/день	Индукционная вирусом частота гибели инфицированных клеток	[446]
$\kappa \sum_j E_j^{ss}$	4.0/день	CTL-индуцированная частота гибели инфицированных клеток	[211, 375, 478]
β	$1.1 \cdot 10^{-8} (\text{d} \cdot \text{cell})^{-1}$	Базовая эффективность заражения клетки-мишени	[211, 375, 478]
s	$[0 - \infty]$	Стоимость мутации	[277, 291, 292, 296-298]
α	$[0 - \infty]$	Снижение распознавания	[277, 291, 292, 296-298]
Δf_i	$[0 - 1]$	Относительное снижение скорости репликации	Свободный
Δr_{ij}	$[0 - 1]$	Относительное уменьшение распознавания	Свободный
κ	10^{-9} /день/клетка	Эффективность уничтожения	[452]
σ/d_E	10^3 клеток	Исходная популяция CTL	[211]
c	1.0/день	Максимальная скорость роста CTL	[211]
d_E	0.1/день	Смертность эффекторных CTL	[211, 375, 478]
h_j	$2.5 \cdot 10^8$ клеток	Количество распознанных клеток для полумаксимальной пролиферации (обратная avidность).	[231]

7.2.2 Моделирование динамики ускользания антигена

Фазы ВИЧ-инфекции

Модель (7.17)-(7.21) была решена численно в пакете MATLAB™. Начальные условия следующие: клетки-мишени T находятся на стационарном уровне неинфицированного пациента, имеется небольшое начальное количество CTL-клеток, и вводится небольшое количество вируса. Результаты решения предсказывают три фазы ВИЧ-инфекции (рисунок 7.5B):

Фаза 1: Передаваемый штамм ВИЧ размножается и истощает клетки-мишени.

Фаза 2: CTL его распознающие размножаются и уменьшают количество инфицированных клеток до тех пор, пока не установится устойчивое состояние (хроническая инфекция).

Фаза 3: Появляются и размножаются ускользающие мутанты вируса. Часть клонов CTL сжимается, но общее количество вируса остается примерно прежним [18].

Динамика вируса и CTL во время ускользания зависит от количества активных эпитопов n и потери узнавания на мутацию Δr (таблица 7.2). В простейшем случае, когда имеется только один клон CTL, находящийся в устойчивом состоянии, а нарастающая ускользающая мутация полностью отменяет распознавание, $\Delta r = 1$, клон сжимается, и мутантная вирусная нагрузка растет до тех пор, пока клетки-мишени не истощатся настолько, чтобы остановить этот рост.

Однако, если потеря распознавания из-за мутации является частичной, $\Delta r < 1$, вирусная нагрузка увеличивается только временно, и клон CTL расширяется, пока не достигнет нового устойчивого состояния. Когда несколько клонов CTL с одинаковой авидностью контролируют популяцию вируса (рисунок 7.5B) и в эпитопе появляется ускользающая мутация, родственные эпитопу клоны CTL сокращаются, поскольку они размножаются с меньшей скоростью, чем другие клоны CTL, но погибают все с одинаковой частотой d_E , уравнение (7.35). За счет других клонов CTL, популяция инфицированных клеток остается под контролем после мутации ускользания. Предсказанная трехфазная динамика наблюдалась у ВИЧ-инфицированных, включая снижение клонов CTL родственными мутированными эпитопам [282, 283, 356, 477].

Детерминанты скорости ускользания мутантного штамма

Присутствие CTL создает положительный отбор ускользающих мутаций (фаза 3 на рисунке 7.5B). Мутантные клоны в разных эпитопах растут с разной скоростью, из-за различий в потере приспособленности и узнавания. Экспоненциальные темпы роста определяют, какой мутантный штамм будет наиболее приспособленным и, таким образом, будет доминировать в популяции вируса. Делается предположение, что начальное состояние до мутации является стационарным, в котором правые части уравнений (7.17) — (7.19) равны нулю. Далее, в эпитопе 1 происходит ускользающая мутация, вызывающая потерю приспособленности Δf_i и потерю распознавания

Δr_i . Как следует из (7.18), (7.20) и (7.21), мутантный штамм первоначально растет как $I_i(t) = I_i(0) \exp(\epsilon_i t)$, где ϵ_i определяется выражением

$$\epsilon_i = \frac{1}{I_i} \frac{dI_i}{dt} = kE_{tot} \left(\frac{\Delta r_i}{n_1} - \Delta f_i \right) \quad (7.22)$$

а $1/n_1$ — доля популяции CTL, распознающих эпитоп 1. Таким образом, баланс между потерей узнавания и стоимостью приспособленности определяет, будет ли расти мутантный штамм, $\epsilon_i > 0$. Следовательно, в чтобы оценить стоимость мутации из скорости ее роста, как это обычно делается с помощью однолокусной модели (§ 2.1), необходимо измерить не только ϵ_i , но и потерю распознавания. По данным в пациентах, частота ускользания ϵ_i резко замедляется в течение первых 100 дней после заражения [282, 283, 286]. Предсказания модели согласуются с этими выводами. Наблюдаемое уменьшение ϵ_i с течением времени может быть предсказано только из изменения $\Delta r/n$, если предположить, что Δf невелико. Когда новые CTL клоны растут, наблюдаемое число атакуемых им эпитопов n увеличивается с временем.

Траектория ускользающих мутаций в плоскости затраты-выгода

Во время хронической инфекции, ВИЧ активирует 10-100 поддающихся обнаружению CTL-клонов против различных эпитопов [278], в которых зафиксировано 5-30 ускользающих мутаций [282, 356]. Чтобы исследовать типичную траекторию ускользающих мутаций по эпитопам и их локусам в плоскости «затрат и выгод» ($\Delta f, \Delta r$), была использована упрощенная модель, описанная выше (§ 7.2.1.1). Параметры Δr и Δf генерировались случайным образом и равномерно в интервале от 0 до 1 для геномов с $n = 10$ эпитопами и $m = 10$ локусами на эпитоп (рисунок 7.6). Локусы нумеровались в порядке убывания частоты ускользания ϵ_i , формула (7.22).

После генерации большого количества случайных наборов, выяснилось, что значения Δr и Δf локуса каждого ранга обратно коррелируют. Типичная траектория в плоскости «затрат и выгод» проходит от высоких выгод (Δr) и низких затрат (Δf) к низким потерям и высоким затратам (рисунок 7.6А). Скорость ускользания на эпитоп уменьшается с количеством совершенных ускользаний (рисунок 7.6В). В соответствии с наблюдениями [283, 356], эпитопы ускользают на двух или более локусов.

Как уже упоминалось, ускользающая мутация вызывает сжатие родственного клона CTL. Этот сжатие снижает потерю распознавания для последующих мутаций в этом эпитопе, потому что иммунное давление на эпитоп ослабляется. Следовательно, уменьшение числа CTL имеет важные последствия для траектории в плоскости «затрат и выгод». В отсутствие уменьшения, траектория остается прямой до тех пор, затраты не превысят выгоды (рисунок 7.6А). Когда CTL клоны уменьшаются, средняя траектория изгибается по направлению к оси абсцисс (рисунок 7.6С). Из-за обеднения CTL, общее количество ускользающих мутаций становится меньше (с 55

на рисунке 7.6А, оно падает до 20 на рисунке 7.6В). Также уменьшается число ускользящих мутаций на эпитоп (рисунок 7.6В, D).

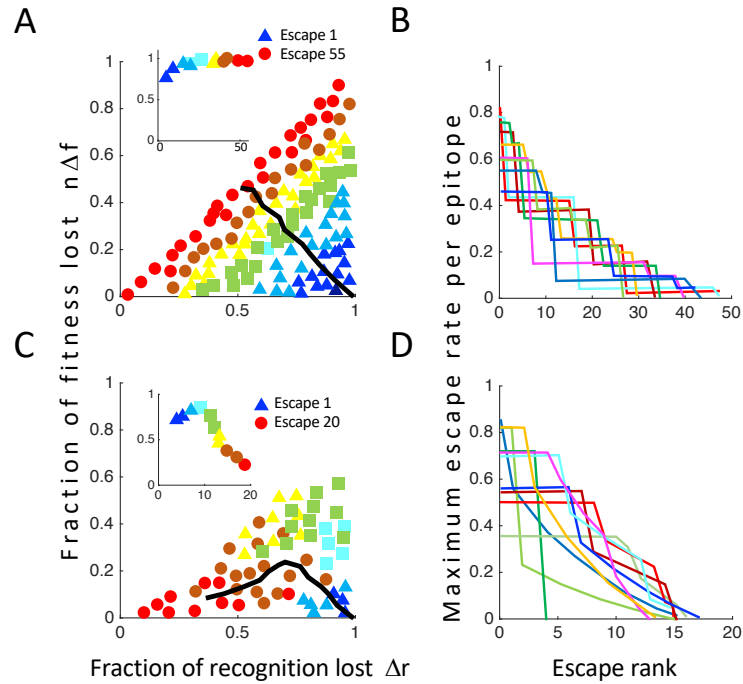


Рисунок 7.6: Траектория ухода в плоскости затрат-выгод с течением времени изгибается из-за распада СТЛ. (А) Для каждого ускользящего локуса, показана относительная стоимость приспособленности, Δf , умноженная на число эпитопов, $n = 10$, и относительная потеря распознавания, Δr , формула (7.22) и табл. 7.2. Цвета показывают ранг ускользящих мутаций, от ранних ускользящих мутаций (синий) до поздних ускользящих мутаций (красный). Врезка: Подгонка наклона траектории для каждого ранга. (В) Максимальный наклон $\Delta r_i/n_1 - \Delta f_i$ для всех 10 эпитопов в типичном прогоне. Δf и Δr генерируются случайным образом для 100 оснований ($m = 10$ локусов на эпитоп). Последовательность ускользящих локусов (черная линия) показана без уменьшения числа СТЛ (А, В) или с уменьшением числа СТЛ (С, D) для 1000 запусков моделирования. Уменьшение числа СТЛ моделируется через зависимость $\Delta r(n+1) = \Delta r(n)\exp[-d'_E \sum_i \Delta r_i(0)]$, суммируя по всем i основаниям в эпитопе, которые ускользнули, с $d'_E = 1/m = 0.1$. По работе [18].

7.2.3 Объяснена корреляция между стоимостью побега и выгодой в гене *Pol*

Результаты показанные на рисунке 7.6 позволяют интерпретировать тот факт [295], что стоимость ускользящих мутаций положительно коррелирует с потерей иммунного распознавания вируса в пациентах (рисунок 7.7). Действительно, дорогостоящие мутации не будут расти в популяции, если они также не приносят еще большей пользы вирусу (7.22). Слабая сила наблюдаемой корреляции (наклон = -0.12) объясняется большим количеством эпитопов-мишеней в хронической фазе, $n \sim 10$ (большинство последовательностей вируса в работе [295] были взяты из хронической фазы). Уменьшение числа СТЛ, вызванный ускользанием еще больше снижает корреляцию. Это также объясняет почему ускользящие мутации наблюдаемые во время острой фазы ВИЧ-инфекции имеют гораздо больший наклон корреляции, чем во время хронической фазы ([18], рисунок 3В).

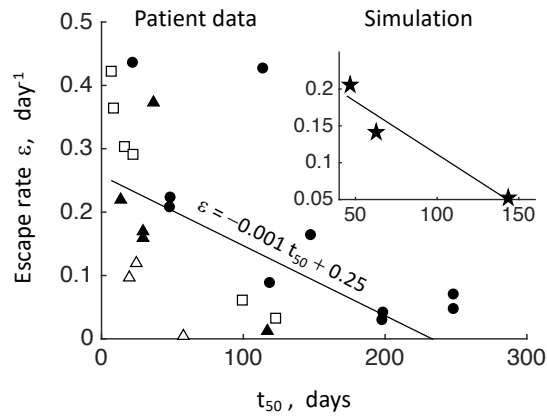


Рисунок 7.7: Скорость ускользания ϵ и время полу-ускользания t_{50} имеют отрицательную корреляцию в двух экспериментальных исследованиях. Доля мутантного эпитопа, как функция времени t , подгоняется кривой $f(t) = 1/[1 + \exp[-\epsilon(t - t_{50})]]$, § 2.1, что позволяет оценить параметры ϵ и t_{50} . Символы показывают данные одного пациента, исследованного в [282] (закрашенные кружки), и пациентов, исследованных в [356], с идентификаторами SH40 (закрашенные треугольники), SH58 (открытые треугольники) и SH77 (открытые квадраты). Вставка: пример моделирования, показывающий корреляцию между коэффициентом ускользания ϵ и временем распространения мутации на 50% популяции инфицированных клеток, обозначенным t_{50} . Параметры ϵ и t_{50} найдены для трех ускользающих мутаций, показанных на рисунке 7.5B, которые происходят в первые 200 дней после заражения. По работе [18].

Важным нюансом, который следует учитывать при сравнении предсказаний с экспериментом [295], является метод измерений. Потеря узнавания CTL, Δr , включает изменения в процессинге антигена, его презентацию на клеточной поверхности, и связывание в комплексе HLA-эпитоп-TCR (ниже «связывание HLA»), в то время как авторы [295] используют только потерю связывания HLA, ΔB . Чтобы получить перевод единиц, необходимо сравнить данные из публикаций, в которых были измерены оба параметра [296-298]. Результатом является прямая корреляция между Δr и ΔB , выраженная в виде линейной зависимости $\Delta B = 0.78 \Delta r - 0.004$ (рисунок 7.8).

7.2.4 Три порядка антигенного ускользания в эпитопе с двумя локусами

В ответ на давление CTL, с течением времени, в эпитопах могут возникать различные комбинации (гаплотипы) мутаций. Авторы [356] заметили, что порядок появления ускользающих гаплотипов изменяется между эпитопами. Можно классифицировать эти порядки ускользания на «простой», «вложенный» и «чехарда» и проиллюстрировать каждый с помощью эпитопа с двумя локусами и четырьмя возможными гаплотипами. Ускользающая мутация на локусе обозначается аллелем 1, в противном случае, она равна 0. «Простой» порядок (наблюдаемый нечасто) представлен одним ускользающим гаплотипом, $00 \rightarrow 10$. «Вложенный» порядок добавляет новую мутацию к единственному мутанту, $00 \rightarrow 10 \rightarrow 11$. Этот порядок ожидается, если оба

локуса находятся под постоянным положительным давлением отбора, $\epsilon > 0$, на протяжении всей инфекции. «Чехарда» вызвана переключением доминантной мутации: 00 \rightarrow 10 \rightarrow 01.

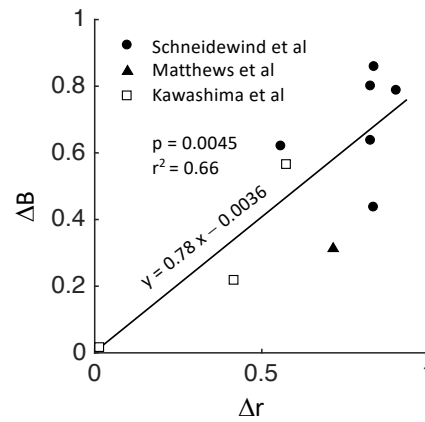


Рисунок 7.8: Оценка взаимосвязи между Δr и ΔB по трем опубликованным экспериментам [296-298]. По работе [18].

Эффект “чехарды” можно рассматривать, как вариант клональной интерференции, когда первая ускользящая мутация испытывает более слабое давление отбора, чем вторая мутация, возникающая позже и вытесняющая первый клон [65] (*глава 3*). Разница в том, что, в настоящей модели, вторая мутация становится более выгодной со временем, из-за зависящего от времени характера давления отбора, вызванного иммунным ответом.

Геном, состоящий из большого числа $n \gg 1$ простых двухлокусных эпитопов, $m = 2$, можно использовать для иллюстрации более общего случая. Из (7.22), частоты ускользания, для каждого из трех мутантных гаплотипов эпитопа j , определяются выражениями

$$\begin{aligned} \epsilon_{10} &= kE_{tot} \left(\frac{\Delta r_1}{n_j(t)} - \Delta f_1 \right) \\ \epsilon_{01} &= kE_{tot} \left(\frac{\Delta r_2}{n_j(t)} - \Delta f_2 \right) \\ \epsilon_{11} &= kE_{tot} \left(\frac{\Delta r_1 + \Delta r_2 - \Delta r_1 \Delta r_2}{n_j(t)} - (\Delta f_2 + \Delta f_1 - \Delta f_1 \Delta f_2) \right) \end{aligned} \quad (7.23)$$

где $1/n_j(t)$ — доля общей популяции СТЛ, E_{tot} , занятая клоном E_j . Без ограничения общности, можно предположить, что начальная скорость ускользания гаплотипа 10 выше, чем у 01: $\epsilon_{10} > \epsilon_{01}$.

Как только самый быстрый ускользящим гаплотипом станет достаточно большим, родственник СТЛ-клон E_j начнет уменьшаться. Общая вирусная нагрузка останется прежней, потому что другие клоны СТЛ будут контролировать вирус. По мере снижения доли клона $1/n_j(t)$ и его давления отбора, стоимость Δf_i в (7.23) станет более важной, и произойдет смена лучшего гаплотипа. В долгосрочной перспективе, уменьшение клона СТЛ приведет к тому, что

дикий тип 00 снова станет доминировать в популяции, но наличие компенсаторных мутаций, не рассмотренных в данной работе, стабилизирует последний ускользнувший гаплотип (§ 2.1).

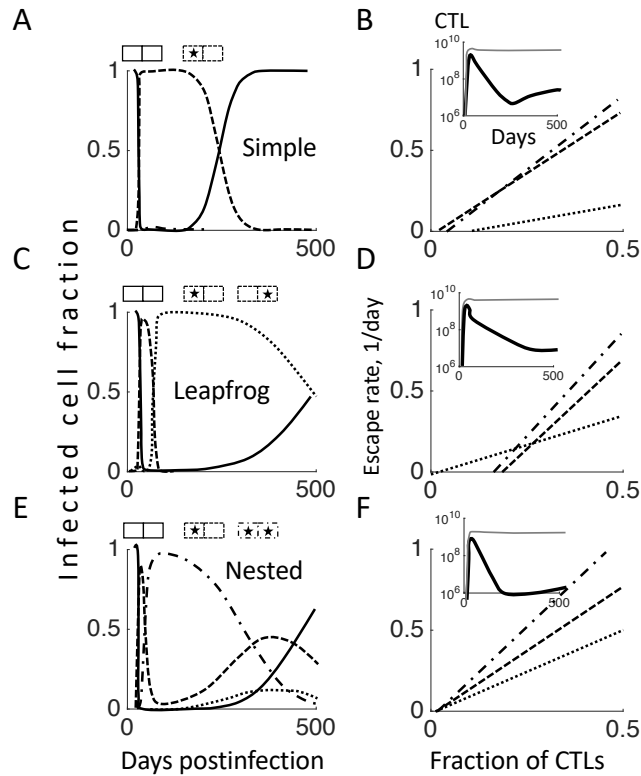


Рисунок 7.9: Давление динамического отбора со стороны CTL вызывает три возможных порядка ускользания внутри эпитопа с двумя локусами. Ускользание согласно модели на рисунке 7.5А, уравнения (7.17) - (7.19). Доля популяции, содержащая каждый из четырех гаплотипов, показана для порядков «простой» (А), «чехарда» (С) и «вложенный» (Е). (В, D, F) Зависимость частоты ускользания каждого гаплотипа (7.23) от доли родственных CTL. Вставка: размер клона CTL (толстая кривая) и общее количество CTL (тонкая). Параметры: (А,В) $\Delta r_{1k} = [0.4, 0.1]$, $\Delta f_{1k} = [0.01, 0.01]$, (С, D) $\Delta r_{1k} = [0.6, 0.2]$, $\Delta f_{1k} = [0.1, 0.003]$, (Е, F) $\Delta r_{1k} = [0.4, 0.25]$, $\Delta f_{1k} = [0.003, 0.003]$. Везде $n = 2, m = 2$. Остальные параметры в таблице 7.2. По работе [18].

Разные порядки доминантных гаплотипов на рисунке 7.9 следующие:

«Простой» (рисунок 7.9А) возникает, если второе основание не может сбежать, потому что $\epsilon_{01} < 0$ в формуле (7.23), или если ϵ_{01} положительно, но мало, так что мутант 11 не успевает вырасти до наблюдаемого уровня, прежде чем дикий тип 00 станет снова самым приспособленным. Этот случай чтобы параметры двух одиночных мутантов сильно отличались.

Вложенный порядок (рисунок 7.9D) имеет место, когда оба одиночных мутанта 01, 10 сохраняют положительную скорость роста в течение длительного периода времени, и двойной мутантный гаплотип имеет самую высокую скорость ускользания из четырех. Вложенный порядок требует, чтобы параметры двух одиночных мутантов были схожими.

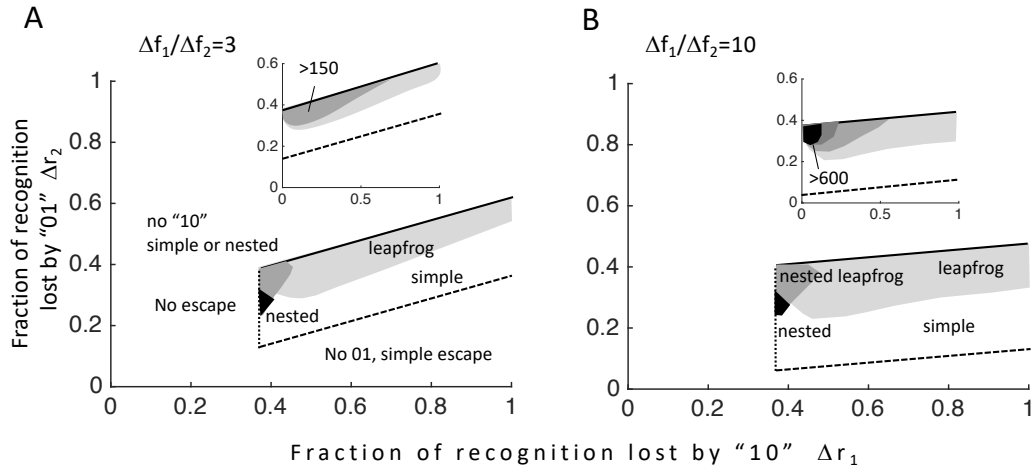


Рисунок 7.10: Порядок появления ускользящих вариантов в одном эпитопе содержит информацию о потере узнавания и приспособленности на мутацию. Оси абсцисс и ординат: параметры потери распознавания на первом и втором локусах, Δr_1 и Δr_2 , соответственно. Области, требующие $\Delta f_1 < 0$, не допускаются (пунктирная линия). Заштрихованные области между этими тремя линиями соответствуют областям пространства параметров, из которых выходят оба узла. Показаны области: «чехарда» $10 \rightarrow 01$ (рисунок 7.9B), «вложенная» $10 \rightarrow 11$ (рисунок 7.9D), «вложенная чехарда» $10 \rightarrow 01$. Вставка: показан период времени, в течение которого гаплотип 01 доминирует в ускользящем эпитопе. Фиксированные параметры: $\epsilon_{10} = 0.5$, $n = 3$, которые соответствуют фазе ранней ВИЧ инфекции [см. [18], рисунок S3, для параметров хронической инфекции]. Второй локус более дешевый для мутации, чем первый: (A) $\Delta f_1/\Delta f_2 = 3$, (B) $\Delta f_1/\Delta f_2 = 10$. Остальные параметры в таблице 7.2. По работе [18].

«Чехарда» (рисунок 7.9C) возникает, если из-за уменьшения STL, гаплотип 01 получает преимущество над изначально доминирующим 10 до того, как тот проигрывает дикому типу 00. Чтобы это произошло, локус 1 должен иметь как большую потерю приспособленности, так и большую потерю распознавания, чем локус 2.

$$\Delta r_1 > \Delta r_2, \quad \Delta f_1 > \Delta f_2$$

Соответствующий клон STL распадается быстрее в начальный период времени, когда доминирует вариант 10, а затем замедляется (вставка на рисунок 7.9D). Причиной является относительно высокий Δr_1 .

Похожее численное моделирование было выполнено для различных наборов $\Delta r_1, \Delta r_2, \Delta f_1, \Delta f_2$ (рисунок 7.10). Номер эпитопа n был фиксирован. Значение ϵ_{10} и отношение $\Delta f_1/\Delta f_2$ были заданы при значениях, характерных либо для острой, либо для хронической инфекции, а также для быстрого или медленного убывания STL (рисунок 7.10 и рисунок S3 в исходной работе [18], соответственно). Диапазоны значений ϵ , наблюдаемые у ВИЧ-инфицированных в разные фазы инфекции, показаны на рисунке 7.7.

При больших ϵ_{10} , «чехарда» наблюдается при больших Δr_1 и в широком диапазоне Δr_2 ([18], рисунок S3). В ряде случаев, гаплотип 11 наблюдается как короткое промежуточное звено между гаплотипами 10 и 01 («вложенная чехарда» на рисунке 7.10). Период времени, в течение которого данный гаплотип доминирует в популяции, зависит от значений Δf_k и Δr_k . Чем менее затронуто гаплотип, тем дольше он доминирует в популяции (время t_{01} на врезке рисунка 7.10).

Родственные СТЛ убывают быстрее всего, когда вариант 11 является самым большим (врезка на рисунке 7.9F).

7.2.5 Заключение

Неполная потеря узнавания СТЛ влияет на скорость и порядок мутаций не меньше, чем стоимость мутации для вируса. Данная модель также объясняет положительную корреляцию между потерей узнавания и стоимостью мутации, наблюдаемую в ВИЧ [295]. Результаты подчеркивают необходимость прямых измерений потерь узнавания различных ускользающих мутаций. В работе по респираторным вирусам в § 7.1, стоимостью мутации пренебрегали, а потеря узнавания описывалась функцией перекрестного иммунитета $K(u)$.) Подход фазовой траектории может помочь с прогнозированием консервативных эпитопов для использования в вакцинах. Приведенный выше анализ имеет несколько упрощений, уместность которых обсуждается в оригинальной работе [18].

7.3 Эволюционная стабильность ВИЧ в хозяине в присутствии дефектных интерференционных частиц

Если геном вируса подвергнется обширным делециям, в нем будут отсутствовать элементы, необходимые для его репликации. Для репликации, он должен быть дополнен вирусом дикого типа, заражающим ту же клетку. Такие условно реплицирующиеся варианты вируса называются «дефектными интерферирующими частицами» (DIP) [299, 300]. Реплицируясь за счет дикого типа, DIP представляют собой вторичных паразитов, препятствующих его репликации, истощая его продукты. DIP много раз предлагался в качестве терапии различных заболеваний, включая ВИЧ. Основная проблема с этой, казалось бы, простой идеей — сомнительная эволюционная стабильность такого взаимодействия. Эволюция резистентности к DIP кажется особенно вероятной для ВИЧ, поскольку естественные DIP для ВИЧ, в отличие от многих других вирусов, не обнаружены.

В работе обсуждаемой в данном разделе [17] исследуется возможность появления вариантов ускользания ВИЧ. Цель состоит в том, чтобы предсказать параметры конструкции DIP, необходимые для стабильного подавления ВИЧ, от которого ВИЧ не мог бы ускользнуть генетически. Разработана простая, но достаточно реалистичная математическая модель динамики вируса для двух связанных биологических уровней, клетки и хозяина.

Модель динамики ВИЧ и гипотетической DIP была построена на двух биоскалах: одноклеточной шкале, описывающей образование частиц ВИЧ и DIP, и шкале хозяина, где предыдущие модели [180, 191] обобщены, чтобы включить наличие нескольких копий DIP на клетку. Коэффициент отбора мутации на биоуровне хозяина выведен аналитически из одноклеточной модели. Его знак проверен, чтобы выяснить, уменьшает ли эволюция ВИЧ кражу продуктов, т.е., может ли ВИЧ ускользнуть от DIP.

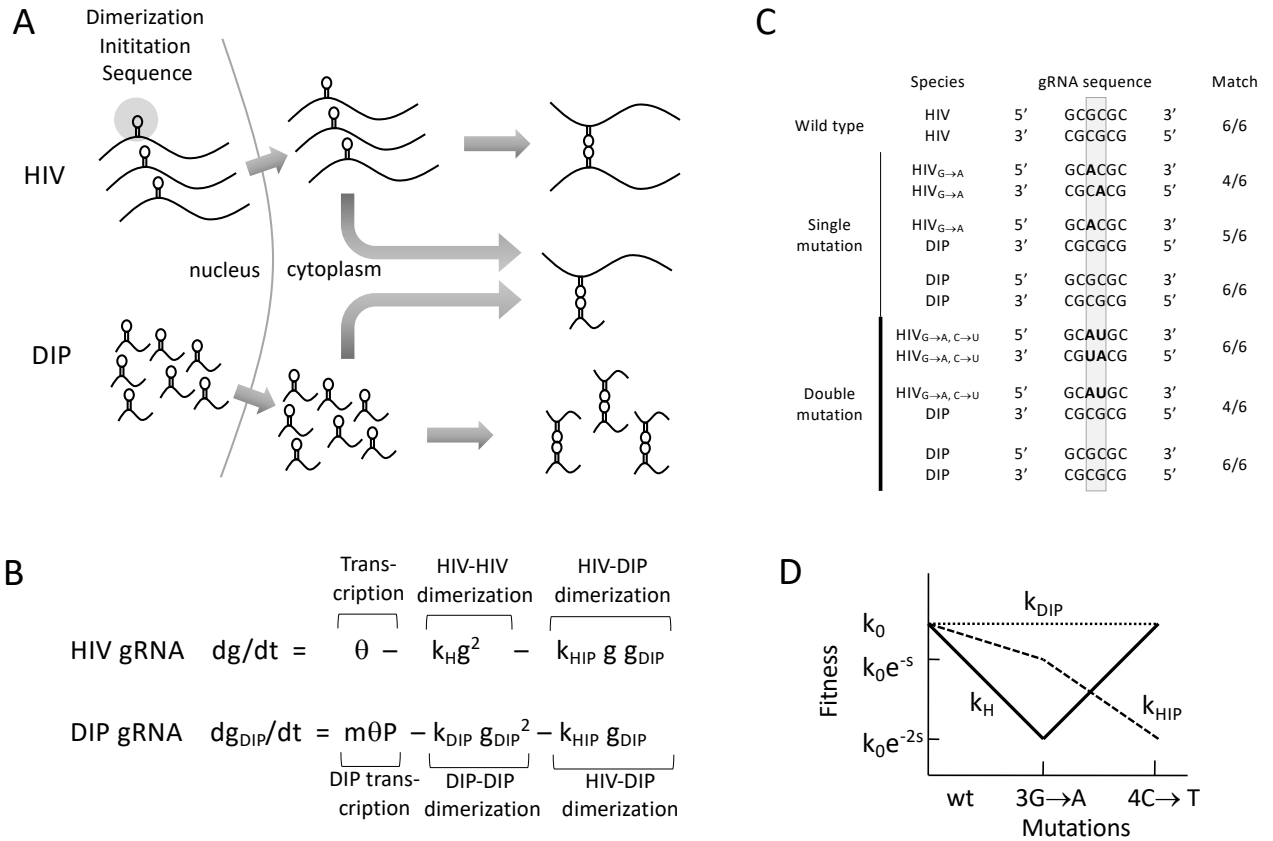


Рисунок 7.11: Интерференция DIP путем «кражи генома» эволюционно нестабильна из-за расходящейся эволюции последовательностей инициации димеризации в ВИЧ и DIP. А) Механизм интерференции. Мономеры геномной РНК у ВИЧ и DIP образуют три типа димерных комплексов (ВИЧ-ВИЧ, ВИЧ-DIP и DIP-DIP), за счет так называемой «петли поцелуя» между последовательностями инициации димеризации, которые включают палиндромную последовательность с консенсусной последовательностью GCGCGC. Благодаря более высокой скорости транскрипции и многочисленным копиям провируса, гРНК у DIP более распространены чем у ВИЧ, поэтому гетеродимеры ВИЧ-DIP сильно превосходят монодимеры гРНК ВИЧ числом. В) Упрощенная модель, описывающая число мономеров гРНК для ВИЧ и DIP в инфицированной клетке, $g(t)$ и $g_{DIP}(t)$ соответственно; θ представляет собой составной параметр, пропорциональный линейной скорости продукции гРНК, а P представляет собой асимметрию экспрессии генов между ВИЧ и DIP. Параметры k_H , k_{IP} , и k_{HIP} представляют собой коэффициенты димеризации для ВИЧ-ВИЧ, DIP-DIP и ВИЧ-DIP соответственно. С) Две мутации в петле поцелуев приводят к дивергентной эволюции ВИЧ и DIP. Верхний ряд: точное совпадение для любой пары гРНК (ВИЧ-ВИЧ, ВИЧ-DIP и DIP-DIP). Средние ряды: если в ВИЧ возникает одна мутация (зеленый прямоугольник), гомодимер ВИЧ-ВИЧ имеет два несовпадения, гетеродимер ВИЧ-DIP имеет одно несовпадение, а гомодимеры DIP-DIP не имеют ни одного несоответствия. Нижние ряды: вторая компенсаторная мутация возникает при ВИЧ, и гетеродимеризация невыгодна. Д) Эволюционная «приспособленность» гомодимеров и гетеродимеров, оцененная по (В) и (С). Хотя одна мутация способствует гетеродимеризации, вторая мутация способствует гомодимеризации ВИЧ. По работе [17].

7.3.1 ВИЧ ускользает от DIP крадущего геномы

В принципе, DIP может подавлять ВИЧ, связывая его геномную РНК (гРНК) («кража генома»). Lentivirusные геномы упакованы в вирионы в виде димеров РНК. Димеризация инициируется сигналом инициации димеризации (DIS), расположенным в стволовой петле 1 (SL1) генома ВИЧ. Он имеет консенсусную последовательность GCGCGC [479], палиндром из шести нуклеотидов.

Было предположено, что в DIP по конструкции отсутствует белок-трансактиватор, и, следовательно, он может экспрессироваться только в клетках, коинфицированных ВИЧ. Минимальная модель, включающая только недимеризованные и димеризованные геномы, может предсказать эволюцию димеризации генома между DIP и ВИЧ (рисунок 7.11А, В). Эта модель (и модель в § 7.3.2 ниже) учитывает следующие вирусологические факты:

(i) Экспериментальные данные демонстрируют, что неполные РНК, но сохраняющие консенсусный палиндром DIS, могут связывать геномы ВИЧ [479, 480]. Вирионы, содержащие такие гетеродимеры РНК, не могут реплицироваться и служат только в качестве супрессора репликации ВИЧ.

(ii) Большинство ВИЧ-инфицированных клеток содержат только один интегрированный провирус ВИЧ [15, 481] из-за короткого времени жизни инфицированных клеток [170] и подавления вирусного рецептора CD4 белком Nef экспрессируемым ВИЧ, что служит для высвобождения вирионов из клетки. Интеграция DIP, однако, не ограничивается одной копией. Множественные копии интегрированного провируса DIP увеличивают продукцию геномной РНК DIP в m раз, где $m = 1, 2, 3, \dots$, — число копий, которое случайным образом варьируется между клетками и среднее значение которого пропорционально концентрации вируса в организме.

(iii) Более короткие РНК-геномы DIP без сплайсинга (вырезки и сшивки) [316, 482] транскрибируются быстрее, чем геномы ВИЧ, поэтому геномы РНК DIP экспрессируются быстрее, чем геномы РНК ВИЧ [316, 482-484]. Из-за переизбытка DIP, большая часть геномных РНК ВИЧ крадет DIP для производства нежизнеспособных вирионов (рисунок 7.11А), оставляя меньшее количество для использования для образования димеров РНК ВИЧ-ВИЧ.

Для анализа эволюционной стабильности, можно рассмотреть одиночную мутацию в DIS ВИЧ типа GCGCGC \rightarrow GCGAGC. Мономеры С и А некомплементарны и, следовательно, эта мутация снижает вероятность гетеродимеризации DIP-ВИЧ. Она же приводит к двум несоответствиям в гомодимеризации ВИЧ-ВИЧ, что оказывает более сильное влияние на вероятность димеризации (рисунок 7.11С). Однако компенсаторная симметричная мутация восстанавливает новый идеальный палиндром в ВИЧ и делает димеризацию DIP-ВИЧ менее благоприятной, чем ВИЧ-ВИЧ. Таким образом, предполагая, что коэффициент димеризации

является функцией количества несовпадений, можно предсказать дивергентную эволюцию последовательности DIS между ВИЧ и DIP (рисунок 7.11D).

Конечно, этот простой аргумент не принимает во внимание снижение приспособленности палиндрома DIS, отличающегося от консенсуса [479] GCGCGC, и рассматривает только два поколения вируса. Однако более общая модель, учитывающая эти факторы, также предсказывает прогрессивное снижение гетеродимеризации по сравнению с гомодимеризацией [17].

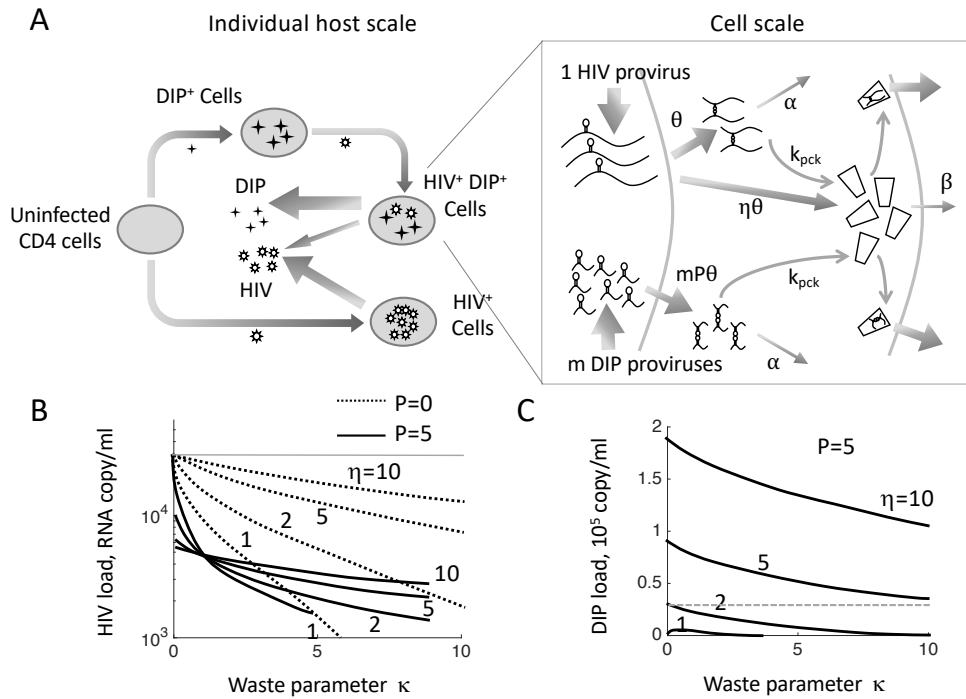


Рисунок 7.12. DIP, которые крадут капсид, стабильно подавляют стационарное количество ВИЧ в широком диапазоне параметров. (A) Схема модели, включающей две уровня биологической организации. Уровень индивидуального хозяина представляет собой простейшую модель репликации ВИЧ, расширенную за счет включения DIP, (7.25)-(7.30). Неинфицированные клетки могут быть инфицированы либо ВИЧ, либо DIP. DIP+ клетки могут быть суперинфицированы ВИЧ. Модель на уровне одной клетки описывается уравнениями (7.24). Суперинфицированная клетка имеет один встроенный провирус ВИЧ и m копий провируса DIP. РНК ВИЧ транслируется в белки, образующие пустые капсиды. DIP не экспрессирует белки. Широкие стрелки обозначают многоступенчатые процессы. Часть стабильных димерных геномов и полных капсидов теряется. Остальные геномы, ВИЧ или DIP, упаковываются в капсиды и высвобождаются в виде инфекционных частиц. (B) Стационарная нагрузка ВИЧ и (C) стационарная нагрузка DIP при различных значениях двух параметров одиночных клеток: «параметр отходов» κ и отношение образования капсида к геному η (таблица 7.3). Горизонтальная серая линия на (B) показывает вирусную нагрузку ВИЧ в отсутствие отходов и DIP ($\kappa = P = 0$), которая считается средней нагрузкой у нелеченых людей ($3 \cdot 10^4$ копий РНК/мл крови). Параметры в (B) и (C): отношение экспрессии DIP к ВИЧ $P = 5$ и базовое число репродукции $R_0 = 10$ (таблица 7.3). Снижение нагрузки ВИЧ в присутствии отходов ($\kappa > 0$, красные линии) по сравнению со стационарным уровнем ВИЧ без лечения (серая линия) частично связано с потерей продуктов ВИЧ (черные пунктирные линии рассчитаны при $P = 0$) и частично с DIP, который конкурирует с ВИЧ за доступные клетки-мишени и крадет капсид ВИЧ в суперинфицированных клетках. Первый эффект более важен при $\eta \sim 1$, а второй—при больших $\eta \gg 1$ [[17], рисунок S2]. По работе [17].

7.3.2 Альтернативный механизм DIP-интерференции: кража белка

Зная, что «кража генома» нестабильна, следующей попыткой будет интерференция путем конкуренции за белки. DIP зависят от белков дикого типа, которые упаковываются и покидают инфицированную клетку, таким образом конкурируя за белки с ВИЧ. Для наглядности, ниже обсуждается кража капсидов, хотя анализ является достаточно общим, чтобы его можно было применить к краже и других белков, таких как оболочка, протеаза или обратная транскриптаза. В зависимости от конструкции, DIP также может красть сразу несколько белков. Единственная разница заключается в точке цикла репликации вируса, когда будут нужны украденные белки.

Ключом к успеху является объединение двух биоуровней: одноклеточной модели, описывающей конкуренцию за капсиды, и динамики вируса в организме хозяина. Ниже обсуждаются как биологические уровни, так и их интеграция. Коэффициент отбора для резистентных мутантов ВИЧ выводится явным образом из уравнений модели. Численные расчеты проводились в программе MATLAB™ (версия R2011a).

Одноклеточная модель с захватом капсида

Чтобы сделать анализ прозрачным, одноклеточная модель (рисунок 7.12А, справа) включает только динамику димеризованных геномов РНК ВИЧ, количество которых в клетке обозначено G , способных к капсидированию капсидов C и димеризованных геномов DIP G_{DIP} . Параметры модели определены в таблице 7.3. Число вирионов DIP и ВИЧ на клетку, предсказанное моделью, используется в качестве входных данных для модели индивидуального пациента, описанной ниже в § 7.3.2.2. Уравнения модели имеют вид

$$\begin{aligned}\frac{dG}{dt} &= \theta - k_{\text{пск}}GC - \alpha G \\ \frac{dC}{dt} &= \eta\theta - k_{\text{пск}}(G + G_{DIP})C - \beta C \\ \frac{dG_{DIP}}{dt} &= mP\theta - k_{\text{пск}}G_{DIP}C - \alpha G_{DIP}\end{aligned}\tag{7.56}$$

Процессы здесь следующие: образование и распад геномов ВИЧ с линейной скоростью θ , упаковка этих геномов в капсиды с коэффициентом $k_{\text{пск}}$, который считается одинаковым для ВИЧ и DIP, образование капсидов с линейной скоростью $\eta\theta$, и конкуренция за инкапсидирование между DIP и ВИЧ. Все геномы являются гомодимерами, либо ВИЧ-ВИЧ, либо DIP-DIP; гетерозиготные геномы не учитываются (см. выше). Один провирус DIP экспрессирует больше копий РНК, чем провирус ВИЧ [316, 482-484], и отношение обозначено P . Клетки с суперинфекцией классифицируют на основе количества копий DIP, обозначенного m , так что отношение между экспрессией DIP и ВИЧ в такой клетке равно mP . Как мы увидим ниже, несколько копий DIP на клетку очень важны для результатов. Геномы и капсиды деградируют с

частотой α и β соответственно. Из-за разделения шкал времени между процессами в клетке (от минут до часов) и хосте (от дней до недель) предполагается, что все три переменные состояния, G , G_{DIP} и C , быстро достигают устойчивого состояния.

Таблица 7.3: Параметры модели и переменные состояния для модели кражи капсида одной клетки, рисунок 7.12А и уравнения (7.24).

<i>Переменные состояния</i>				
Обозначение	Определение	Единицы		
G	Концентрация полноразмерных димеризованных геномных мРНК ВИЧ	1/мкл		
G_{DIP}	Концентрация полноразмерных димеризованных геномных мРНК DIP	1/мкл		
C	Концентрация способных к капсидированию капсидов в цитоплазме	1/мкл		
m	Число копий провируса DIP (MOI) на клетку	1		
<i>Параметры модели</i>				
Обозначение	Определение	Единицы	Значение	Ссылка
θ	Скорость накопления геномов ВИЧ	1/мкл/день	$\left. \begin{array}{l} \\ \\ \\ \end{array} \right\} \kappa = \frac{\alpha\beta}{\theta k_{psc}}$	Свободный пар.
k_{psc}	Константа упаковки	мкл/день		
α	Частота потери геномов	1/день		
β	Частота потери капсидов	1/день		
η	Отношение накопления капсидов к геномам	1	1.2 – 5	[375, 480]
P	Асимметрия экспрессии между DIP и ВИЧ	1	8 – 10	[316]

Модель динамики ВИЧ и DIP на уровне хозяина

Стационарные числа ВИЧ и DIP в клетке используются в качестве входных данных для модели вирусной динамики в организме хозяина (рисунок 7.12А, слева). По сравнению с простейшей моделью динамики вируса [180, 182, 183], модель включает DIP, коинфекцию клеток DIP и ВИЧ, а также тот факт, что клетки с коинфекцией продуцируют меньше ВИЧ, чем клетки, инфицированные только ВИЧ. Модель на рисунке 7.12А, слева, описывается такими уравнениями:

$$\frac{dT}{dt} = b - (d + kV + kV_{DIP})T \quad (7.25)$$

$$\frac{dI}{dt} = kVT - \delta I \quad (7.26)$$

$$\frac{dT_{DIP\ m}}{dt} = kV_{DIP}T_{DIP\ m-1} - (d + kV + kV_{DIP})T_{DIP\ m}, \quad m = 1, 2, 3, \dots \quad (7.27)$$

$$\frac{dI_{D\ m}}{dt} = kVT_{DIP\ m} - \delta I_{D\ m}, \quad m = 1, 2, 3, \dots \quad (7.28)$$

$$\frac{dV}{dt} = n\delta I + n\delta \sum_{m=1}^{\infty} \psi_m I_{D\ m} - cV \quad (7.29)$$

$$\frac{dV_{DIP}}{dt} = n\delta \sum_{m=1}^{\infty} \rho_m \psi_m I_{Dm} - cV_{DIP} \quad (7.30)$$

Процессы, включенные в уравнения (7.25) — (7.30) таковы. Неинфицированные клетки, допускающие репликацию вируса T пополняются из линейного источника [180] с механизмом, обсуждаемым в работах [184, 185]. Этот компартмент истощается из-за их естественной гибели, инфицирования частицами ВИЧ или инфицирования DIP. Зараженные клетки I производят n вирионов и умирают со средней скоростью $\delta \sim 1/\text{день}$ в (7.26). Клетка также может быть инфицирована m копиями провируса DIP, T_{DIPm} в (7.27). Эти клетки не экспрессируют белки ВИЧ и погибают с той же скоростью, что и неинфицированные клетки. При коинфекции ВИЧ, такая клетка становится коинфицированной, I_{Dm} в (7.28). Эти ВИЧ+DIP+ клетки умирают так же быстро, как ВИЧ+DIP- клетки, I , после образования $\psi_m n$ частиц ВИЧ и $\rho_m \psi_m n$ частиц DIP. Наконец, свободные вирусные частицы V и V_{DIP} удаляются с частотой c , которая является самой высокой частотой в модели, уравнения (7.29) и (7.30).

Таблица 7.4: Переменные состояния и параметры для модели индивидуального хозяина, уравнения (7.25)-(7.30). Две нижние строки: G , G_{DIP} и C представляют собой стационарные значения переменных состояния внутриклеточной модели.

<i>Переменные состояния</i>				
Обозначение	Определение	Единица измерения		
T	Неинфицированные CD4+ Т-клетки, перmissive для репликации вируса	клетка/мкл		
I	CD4+ Т-клетки, инфицированные только ВИЧ	клетка/мкл		
T_{DIPm}	CD4+ Т-клетки, инфицированные m копиями провируса DIP, но не инфицированные ВИЧ, $T_{DIP0} = T$	клетка/мкл		
I_{Dm}	Клетки инфицированные ВИЧ и m копиями провируса DIP	клетка/мкл		
V	Вирусная нагрузка ВИЧ	копия РНК /мл		
V_{DIP}	Вирусная нагрузка DIP	копия РНК /мл		
<i>Параметры модели</i>				
Обозначение	Определение	Единица измерения	Значение	Ссылка
b	Линейная скорость продукции неинфицированных клеток	клетка/мл/день	$R_0 = \frac{bkn}{cd} \sim 10$	[438]
d	Смертность неинфицированных клеток	1/день		
k	Фактор инфекционности	мл/день/копия РНК		
n	Размер всплеска ВИЧ из одной инфицированной клетки	РНК копия/клетка		
c	Частота потери вирионов	1/день		
δ	Смертность инфицированных ВИЧ клеток	1/day	1.0/day	[234]

$n\psi_m$	Число вирионов ВИЧ из клетки коинфицированной ВИЧ и m копиями провируса DIP	1	$\frac{k_{\text{pck}}GC}{\delta}$	Модель на уровне клетки
$n\rho_m\psi_m$	Число вирионов DIP из того же	1	$\frac{k_{\text{pck}}G_{\text{DIP}}C}{\delta}$	

Параметры модели собраны в табл. 7.4. Дополнительные параметры ρ_m, ψ_m связанные с DIP представляют собой выходные данные модели с одной ячейкой, уравнения (7.24). Они могут быть выражены в терминах параметров одноклеточной модели и количества геномов и капсидов, в стационарном состоянии

$$\begin{aligned} n &= \frac{k_{\text{pck}}[GC]_{P=0}}{\delta} \\ \psi_m n &= \frac{k_{\text{pck}}GC}{\delta} \\ \rho_m \psi_m n &= \frac{k_{\text{pck}}G_{\text{DIP}}C}{\delta} \end{aligned} \quad (7.31)$$

Используя стационарные значения G, C, G_{DIP} найденные из (7.24), параметры ψ_m и ρ_m можно выразить через параметры уровня клетки η, mP и k , что было сделано в исходной работе [17]. Было показано, что n пропорционально θ/δ и зависит от η и k .

Поскольку основное внимание работы было на хронической ВИЧ-инфекции, динамическая система на уровне хозяина, заданная уравнениями (7.25)-(7.31), также в стационарном состоянии. Стационарные переменные на уровне хоста получены аналитически [17]. Поскольку выражения довольно большие и в неявном виде, они здесь не приводятся, а результаты показаны графически (рисунок 7.12В и С).

Параметры модели

Таблицы 7.3 и 7.4 содержат параметры модели на обоих биологических уровнях. К счастью, результаты выражаются через четырех безразмерных параметров, изменив единицы измерения:

(i) Параметр отходов

$$k = \frac{\alpha\beta}{\theta k_{\text{pck}}}$$

(ii) Отношение способных к капсидированию капсидов к димеризованным геномам ВИЧ, продуцируемым в единицу времени η , называемое ниже отношением продукции капсида к геному.

(iii) Базовый коэффициент воспроизводства, R_0

(iv) Отношение экспрессии генома DIP к ВИЧ, P . Лентивирусные DIP с $P = 8 - 10$ были сконструированы [316]. Ниже основное внимание уделяется консервативному значению $P = 5$, но интервал $P = 2-30$ также был изучен [17].

Подавление ВИЧ динамически стабильно

Стационарные значения нагрузки ВИЧ и DIP у отдельного пациента, нормированные на нагрузку ВИЧ при отсутствии DIP и отходов ($P = \kappa = 0$), могут быть выражены аналитически [17] через R_0, ψ_m, ρ_m и, в конечном итоге, через четыре безразмерных параметра κ, η, P, R_0 , где три первых относятся к одноклеточной модели, а R_0 к уровню хозяина. Это громоздкое аналитическое решение, не приводимое здесь но иллюстрированное на рисунке 7.12В графически, демонстрирует, что присутствие DIP подавляет ВИЧ в широком диапазоне параметров. Степень подавления ВИЧ и нагрузка DIP увеличиваются с асимметрией экспрессии DIP и ВИЧ, $P > 1$ ([17], рисунок 2d, e).

Условие динамической устойчивости DIP в популяции задается условием, которое при малом параметре отходов $\kappa \ll 1$ имеет вид [17]

$$\eta > \eta_c = \frac{1 + P}{P} \frac{R_0}{R_0 - 1}$$

Иными словами, чтобы DIP был стабильным, ВИЧ должен генерировать дополнительные капсиды по сравнению со своим геномом, чтобы DIP мог на них паразитировать. Если число капсидов ограничено, DIP не будет устойчиво размножаться. Этот порог на η значительно понижается (примерно в R_0 раз), если в модель включено деление клеток зараженных DIP но не ВИЧ (§ 7.4).

Снижение нагрузки ВИЧ по сравнению с ее значением при $\kappa = 0$, $R_0 = 10$ (рисунок 7.12В) лишь частично обусловлено наличием DIP. Остальная часть снижения связана с увеличением отходов, κ , что уменьшает число вирионов ВИЧ на клетку n . В качестве контроля, нагрузка ВИЧ при отсутствии DIP (т.е. при $P = 0$) показана на рисунке 7.12В (пунктирные линии).

Подавление ВИЧ вызвано множественностью DIP-инфекции

Одной из причин сильного подавления ВИЧ является множественное заражение клеток DIP. Средняя множественность DIP-инфекции, обозначаемая $E[m]$, довольно велика даже при скромных значениях η и P (рисунок 7.17А). Следовательно, результаты существенно зависят от множественной интеграции DIP в геном клетки. Количество копий DIP, m , варьируется среди коинфицированных клеток. Используя модель индивидуального пациента (7.25)-(7.30), средняя множественность DIP-инфекции $E[m]$ была получена аналитически [17] как функция параметра

отходов κ , асимметрии экспрессии P и отношения чисел капсидов и геномов η . Средняя множественность интегрированных геномов DIP, $E[m]$, увеличивается с P и η (рисунок 7.13А). Таким образом, отношение DIP к геномной мРНК ВИЧ внутри клетки определяется P и η двумя путями: напрямую, через молекулярную архитектуру DIP, и косвенно, через амплификацию этого фактора через $E[m]$. Для проверки вклада множественности DIP в интерференцию с ВИЧ, был рассмотрен вариант модели, где DIP ограничивался одной копией на клетку ($m = 1$ или $m = 0$). Затем была пересчитана вирусная нагрузка ВИЧ (рисунок 7.13В). Этот «контроль» продемонстрировал, что при $E[m] \leq 1$ снижение вирусной нагрузки ВИЧ маленькое.

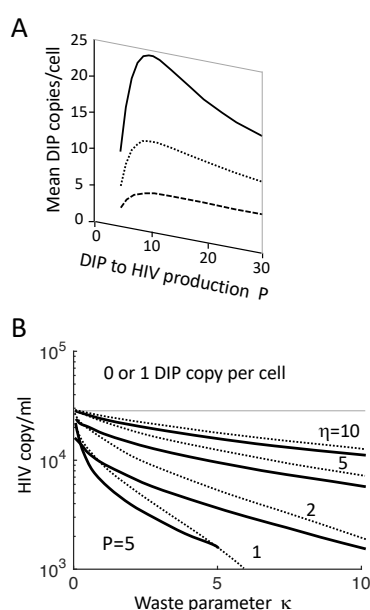


Рисунок 7.13: Высокая средняя множественность заражения DIP повышает динамическую стабильность DIP и подавление ВИЧ. (А) Среднее количество копий DIP на клетку, $E[m]$, как функция асимметрии экспрессии P при отсутствии отходов, $\kappa = 0$, и фиксированных значениях $\eta = 2$ (штриховая кривая), 5 (пунктир), и 10 (сплошная). (В) Сплошная кривая: Стационарная нагрузка ВИЧ в зависимости от параметра отходов κ , когда множественность DIP-инфекции m искусственно ограничена числом 1. Пунктирная кривая: То же самое, в отсутствие DIP ($P = 0$, рисунок 7.12В). Параметры $P = 5$, $R_0 = 10$ такие же как на рисунке 7.12. По работе [17].

Таким образом, множественность интегрированных геномов DIP, m , имеет решающее значение для DIP-опосредованного подавления ВИЧ. С другой стороны, множественные инфекции DIP усиливают даже умеренную асимметрию экспрессии P и создают относительно высокие нагрузки DIP, даже когда η чуть больше 1 (рисунки 7.12С и 7.13). Этот вывод справедлив в широком диапазоне значений $P > 1$ и η ([17], рисунок 3D,E) при выполнении условия динамической устойчивости.

7.3.3 Подавление ВИЧ с помощью DIP является эволюционно стабильным

Следующий вопрос заключается в том, может ли ВИЧ ускользнуть от DIP, изменяя свой геном. В принципе, ВИЧ мог бы мутировать, чтобы увеличить отходы капсида, параметр k . Например, снижение эффективности упаковки k_{psc} позволило бы деградировать больше капсидов, чтобы осталось меньше для DIP. Таким образом, увеличение отходов капсида могло бы принести пользу ВИЧ из-за снижения интерференции со стороны DIP. Однако эта мутация также имеет свою цену, потому что увеличение отходов продуктов снизит выход вириона. Нагрузка ВИЧ, построенная как функция k , показывает преимущества и затраты как два компонента подавления ВИЧ: один из-за снижения числа частиц ВИЧ на клетку, а другой из-за интерференции DIP (рисунок 7.12B). С увеличением k , одна компонента становится больше, а другая меньше. Априори не ясно, какой эффект возьмет верх.

Чтобы определить, находится ли ВИЧ под давлением отбора для увеличения или уменьшения отходов, коэффициент отбора ∂s , определяемый как малое увеличение приспособленности мутантного штамма ВИЧ из-за малого изменения параметра отходов k , получен в исходной работе аналитически [17]. Положительные значения ∂s означали бы, что такая мутация выгодна ВИЧ и будет расти в числе. Ответ имеет общий вид

$$\partial s = \left[\frac{\partial n_{av}}{n_{av}} \right]_{E[m]=const} \quad (7.32)$$

$$n_{av} = n \frac{1 + \sum_{m=1}^{\infty} \psi_m E[m]^m}{1 + \sum_{m=1}^{\infty} E[m]^m}$$

где среднее число потомков вируса на среднюю клетку, n_{av} , представляет собой мальтузианскую приспособленность ВИЧ к присутствию DIP в устойчивом состоянии, $E[m]$ представляет собой среднюю копию DIP на клетку, обсуждавшуюся выше, а ψ_m и n выражены аналитически через η, mP, k [17]. Мутация, вызывающая изменение параметра отходов k , изменяет параметры ψ_m и n , которые зависят от k . Поскольку влияние мутации на k , обозначаемое ∂k , неизвестно и может варьироваться в зависимости от основания, результирующий коэффициент отбора выражается в нормализованной форме $\partial s / (\partial k / k)$.

Стоит подчеркнуть, что коэффициент отбора здесь получен из молекулярной модели. Большинство исследований используют коэффициент отбора в качестве входного параметра и определяют его путем подгонки данных [6, 15, 77, 154, 286, 295, 485-487]. Кроме того, даже полезный аллель, вероятно, вымрет из-за комбинации эффектов случайного дрейфа и сцепления [33]. Чтобы закрепиться в популяции, мутация должна произойти в штамме с высокой

приспособленностью [16, 76]. В настоящей работе, этой сложностью пренебрегли, поскольку основное внимание уделялось общему направлению эволюции, а не ее точной скорости.

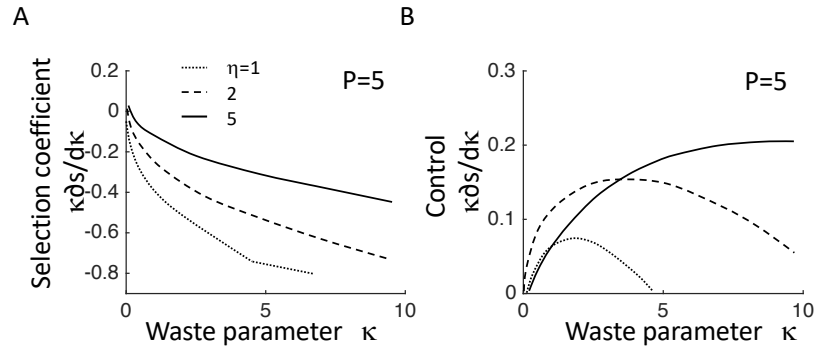


Рисунок 7.14: ВИЧ не может ускользнуть от DIP, увеличивая растрату ресурсов. (А) Нормализованный коэффициент отбора $\partial s/(\partial \kappa/\kappa)$ как функция κ , для различных отношений скорости продукции капсидов и геномов η , формула (7.32). В) Отрицательный контроль для $\partial s/(\partial \kappa/\kappa)$ в клетках с двойной инфекцией ВИЧ+DIP+, когда пренебрегают изменениями числа ВИЧ частиц на клетку из-за увеличения отходов капсида. Параметры $R_0 = 10, P = 5$. По работе [17].

Окончательный результат (рисунок 7.14А) показывает, что мутанты ВИЧ с повышенным отходом капсида невыгодны, поскольку $\partial s/(\partial \kappa/\kappa) < 0$ для диапазона η и κ , и это несмотря на то, что DIP интерференция, как и ожидалось, уменьшается с увеличением κ (рисунок 7.12В для малых η). Из-за убывания числа ВИЧ вирионов n , стоимость мутации в только-ВИЧ-инфицированной клетке без DIP выше, чем ее выгода в коинфицированной от уменьшения интерференции. Чтобы сделать это совершенно очевидным, расчет был повотрен, фиксируя n в уравнении (7.32). Как и ожидалось, в этом случае $\partial s/(\partial \kappa/\kappa) > 0$, и ВИЧ эволюционирует в сторону большего количества отходов капсида и меньшего вмешательства DIP (рисунок 7.14В). Направление эволюционного отбора устойчиво в широком диапазоне $P > 2$ ([17], рисунок 3D,E).

На интуитивном уровне, большая часть ВИЧ, в этой модели, производится в одиночно инфицированных клеток, потому что, в куда более многочисленных коинфицированных, они подавлены DIP. Следовательно, именно количество вирионов ВИЧ в одиночно инфицированных клеток, а не интерференция DIP в клетках с двойной инфекцией, определяет направление эволюции ВИЧ в организме хозяина.

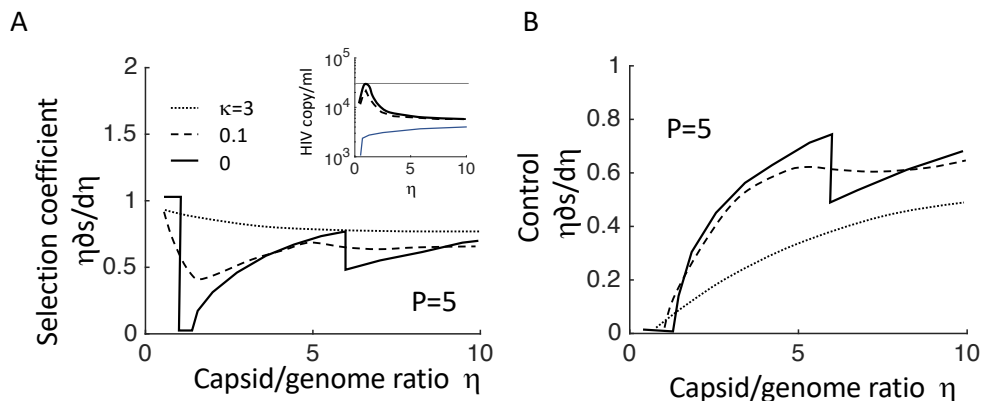


Рисунок 7.15: ВИЧ не может ускользнуть от DIP, уменьшая производство капсидов. (А) Нормированный эффективный коэффициент отбора $\partial s / (\partial \eta / \eta)$, как функция η , для трех значений параметра отходов κ . Вставка: ВИЧ-нагрузка как функция η при трех значениях параметра отходов κ . Положительные значения $\partial s / (\partial \eta / \eta)$ подразумевают, что мутация выбрана, и ВИЧ эволюционирует в сторону увеличения η . (В) Отрицательный контроль при фиксированном числе вирионов ВИЧ n . Параметры как на рисунке 7.12. По работе [17].

Другой возможный путь ускользания ВИЧ от интерференции DIP заключается в уменьшении избытка капсидного материала, т. е. уменьшении параметра η . Точно такой же анализ на основе формулы (7.32) показывает [17], что ВИЧ всегда развивается в сторону высоких значений η . Мутанты ВИЧ, которые производят больше капсидов, более приспособленные, в широком диапазоне η и κ (рисунок 7.15А). Абсолютное значение коэффициента отбора зависит от η и наличия DIP. При $\eta < 1$, когда DIP отсутствует, но капсидов геномам не хватает, ВИЧ выигрывает от увеличения η за счет увеличения потомства вирионов. При $\eta_c < \eta < P + 1$, когда присутствует DIP, также полезно увеличение продукции капсида из-за присутствия DIP, который крадет большинство капсидов в клетках с коинфекцией.

Интересно, что в узком интервале η , таком, что $1 < \eta < \eta_c$, в пределе малых отходов $\kappa \rightarrow 0$ и ниже порога динамической устойчивости DIP коэффициент отбора равен нулю (рисунок 7.15А). Интуитивно понятно, что у ВИЧ уже больше капсидов, чем ему нужно, и ему не нужно конкурировать с DIP, которого нет. Только конечная скорость потери продукта ($\kappa > 0$) приводит к небольшому положительному коэффициенту отбора.

В качестве контроля процедуру повторили с фиксированным числом вирионов n . Значение коэффициента отбора, опять таки, положительно везде, где DIP динамически устойчив (рисунок 7.15В). Обратите внимание на разрыв при $\kappa = 0$ и $\eta = 1 + P$, вызванный тем, что DIP перестает конкурировать с ВИЧ за капсиды при $\eta > 1 + P$, когда капсидов достаточно для них обоих. Коэффициент отбора слабо зависит от P ([17], рисунок 4).

7.3.4 Заключение

Показано [17] что интерференция путем ко-димеризации между геномами DIP и ВИЧ является эволюционно нестабильной. В подтверждение данного результата, различные подтипы ВИЧ демонстрируют разные последовательности DIS глава[488]. Этот результат разнится с работой [489], из-за гипотетических предположений в [489], расходящихся с реальной вирусологией ВИЧ [36].

С другой стороны, интерференция путем кражи белков является эволюционно стабильной при условии, что она стабильна динамически. Основные выводы этого анализа, эволюционная нестабильность кражи генома и эволюционная стабильность кражи белка в широком диапазоне параметров, соответствуют здравому смыслу, согласно которому паразит (DIP), который вредит

своему хозяину (ВИЧ) без пользы для себя, не будет эволюционно стабилен в долгосрочной перспективе. И наоборот, в сценарии с кражей капсида ожидается, что DIP, который истощает избыток капсидов ВИЧ, но их использует, будет персистировать.

Модель имеет следующие ограничения и допущения.

(i) Модель не учитывает эпистаз (*глава 5*).

(ii) Модель не объясняет, почему, в отличие от более простых гамма-ретровирусов, ВИЧ и другие лентивирусы не проявляют спонтанного образования DIP [312]. (Из-за качественных различий в стратегии репликации, настоящая модель не может быть применена к гамма-ретровирусам.)

(iii) В модели предполагается, что все события происходят в случайные моменты времени, без учета временных задержек, которые потребовали бы структурирования по времени [490]. Авторы [332] использовали такую динамику и предсказали хаотические колебания [490]. Искусственные лентивирусные DIP [484, 491, 492] не обнаруживают такого поведения, и настоящая модель ее не предсказывает. Возрастная структура динамики с временными задержками вряд ли изменит основные выводы. В заключение, модель эволюции ВИЧ без эпистаза, на уровне хозяина, показывает [17], что ВИЧ не может ускользнуть от DIP.

7.4 Эволюционный конфликт ВИЧ между уровнями хозяина и популяции в присутствии дефектных интерферирующих частиц

Как показано в § 7.3, DIP-резистентный мутант ВИЧ находится под отрицательным отбором в хозяине. Однако если он все же появится, например, в результате редкой большой флуктуации, он мог бы затем распространиться в населении обработанной DIPом. Действительно, из-за более высокой вирусной нагрузки, он имеет более высокую вероятность передачи [225], чем дикий тип, и, в конечном итоге, мог бы его вытеснить. В данном разделе, эта возможность анализируется с акцентом на демографические группы с высоким риском инфицирования ВИЧ [338-340]. Предыдущее эпидемиологическое моделирование показывает, что DIP могут автоматически концентрироваться и снижать распространенность ВИЧ в группах высокого риска [334]. Следовательно, эти группы высокого риска представляют собой ключевую группу населения, в которой могут возникать устойчивые к DIP мутанты. Работа [21] расширяет модель [17] до уровня населения, чтобы понять, могут ли устойчивые к DIP мутанты ВИЧ распространяться в группах высокого риска.

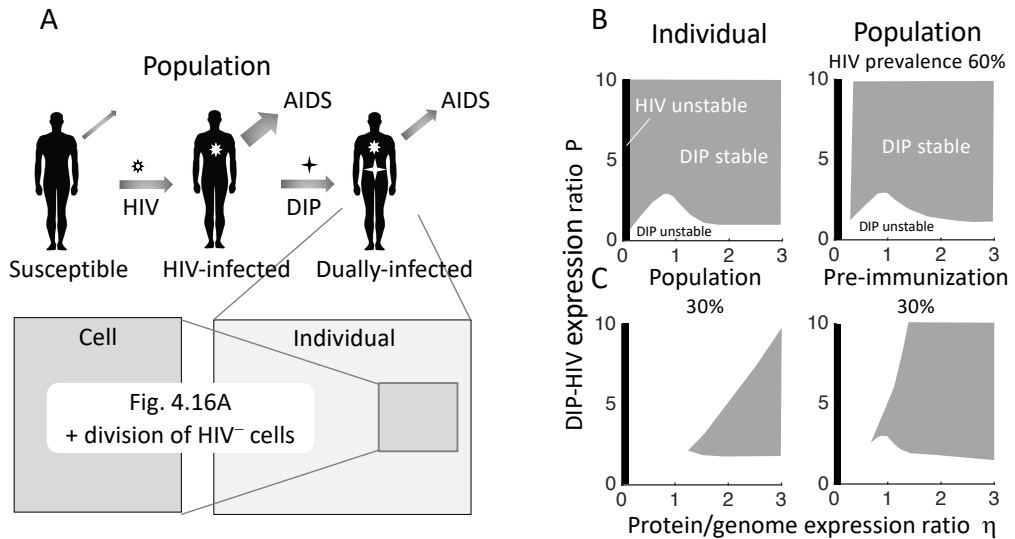


Рисунок 7.16: DIP и ВИЧ динамически костабильны на трех биологических уровнях. (А) Схема многомасштабной математической модели, которая отслеживает уровни ВИЧ и DIP на трех биологическом уровнях: отдельных клеток, индивидуального хозяина, и населения. Модель клетки показана на рисунке 7.12А в § 7.3. Выходные данные клеточной модели (число частиц ВИЧ и DIP) используются в качестве входных данных для модели на уровне хозяина. Модель на уровне хозяина показана на рисунке 7.12А, плюс деление CD4 Т-клеток, как неинфицированных, так и инфицированных только DIP. Выходные данные модели на уровне хозяина (вирусная нагрузка ВИЧ и DIP) используются в качестве входных данных для модели на уровне населения [уравнения (7.33) и таблица 7.5]. Частоты смертности и коэффициент передачи вируса рассчитываются на основе нагрузки ВИЧ и DIP у носителей по работе [225]. (В) Динамическая стабильность DIP на уровне индивидуального хозяина (слева) и эпидемиологическом уровне (справа), определенная количественно для группы населения высокого риска заражения ВИЧ. Процент начальной распространенности ВИЧ до внедрения DIP, в этом примере, равен 60%. (С) Динамическая стабильность DIP в группе высокого риска (30% начальной распространенности ВИЧ) в двух случаях: предварительной DIP-иммунизации (справа) и когда DIP могут заражать только ВИЧ-инфицированных (слева). DIP, созданные с $P > 3$, стабильны, за исключением очень низких значений η (область исчезновения ВИЧ); по мере увеличения начальной распространенности ВИЧ, границы DIP-стабильности в населении приближаются к границам DIP-стабильности внутри хозяев. По работе [21].

7.4.1 Модель динамики ВИЧ и DIP на трех биологических уровнях

Модель состоит из три уровней биологической организации (рисунок 7.16А). Каждый уровень представлен системой обыкновенных дифференциальных уравнений, описывающих репликацию и передачу ВИЧ и DIP. Уравнения пренебрегают стохастическими эффектами, потому что количество инфицированных клеток в организме и инфицированных людей в населении исчисляется миллионами. Целью является не описание процесса эволюции с учетом сложного взаимодействия детерминистических и стохастических эффектов [3] как в главах 2-4, а лишь определение его общего направления путем расчета коэффициентов отбора устойчивых к DIP мутантов, что можно выполнить и в детерминистическом приближении (§§ 7.2 и 7.3).

Таблица 7.5. Безразмерные параметры на уровне населения

Обозначение	Определение	Единица	Значение	Ссылка
R_0^{pop}	Базовое число репродукции в населении	1 ^a		[225]
μ	Относительная скорость передачи ВИЧ из ВИЧ+DIP+ людей по сравнению с ВИЧ+DIP- людьми	1		[225]
ϕ	Относительная скорость передачи DIP из ВИЧ+DIP+ людей по сравнению с передачей ВИЧ из ВИЧ+ DIP- людей	1		[225]
B	Относительное сокращение продолжительности жизни из-за неподавленной ВИЧ-инфекции	1	3.5	[225]
τ	Относительное снижение смертности от ВИЧ вследствие суперинфекции DIP	1		[493]
δ_I	Смертность ВИЧ+DIP- людей	1/год	0.1	[225]

Модели индивидуального хозяина и одной клетки была введена в § 7.3. В этом разделе, модель уровня хозяина немного изменена, чтобы учесть деление неинфицированных и только DIP-инфицированных клеток. Модель на уровне населения основана на стандартном эпидемиологическом подходе «восприимчивые-зараженные» [494], включая DIP, как описано в [334], но только в одной демографической группе, группе высокого риска. Все соответствующие параметры составной модели приведены в таблицах 7.5 и 7.6.

Одна клетка

Динамика DIP-интерференции в одной клетке описывается уравнениями (7.24) в § 7.3. Ключевыми параметрами являются отношение продукции РНК-геномов DIP к ВИЧ, P , и отношение продукции капсида ВИЧ к продукции генома ВИЧ, η (таблица 7.6).

Таблица 7.6. Безразмерные параметры на уровне клетки и хозяина

Обозначение	Определение	Значение	Ссылка
<i>Параметры уровня клетки</i>			
κ	Параметр отходов капсида и генома	0.01	Таблица 7.3
P	Отношение скорости транскрипции DIP к таковой для ВИЧ	> 1	Таблица 7.3
η	Capsid-to-genome production rate ratio for HIV	> 0	Таблица 7.3
<i>Параметры уровня хозяина</i>			
$R_{00} = [R_{0 \text{ host}}]_{\eta=1, \kappa=0}$	Базовое число репродукции в хозяине	10	Таблица 7.4 [438]
δ/d	Сокращение продолжительности жизни клеток из-за ВИЧ-инфекции	10	[180]
c/d	Отношение частоты вычищения вируса к смертности инфицированных клеток	10	[346]

h_0	Среднее число делений клетки-мишени за время ее жизни	3.3	[180]
-------	---	-----	-------

Перечислены составные безразмерные параметры, относящиеся к окончательным результатам на уровне популяции. Определения κ и R_{00} в терминах исходных параметров модели даны в таблицах 7.3 и 7.4 в § 7.3. Все параметры были введены в § 7.3, кроме h_0 , который является новым параметром.

Уровень хозяина

В отдельном хозяине, DIP конкурирует с ВИЧ за доступные клетки-мишени CD4 T. Клетки, коинфицированные как ВИЧ, так и DIP, производят значительно меньше вирионов по сравнению с клетками, инфицированными только ВИЧ, но вместо этого производят большое количество вирионов DIP, использующих белки экспрессируемые ВИЧ. В результате, вирусная нагрузка ВИЧ у лиц, получивших лечение DIP, куда ниже, чем у лиц, инфицированных только ВИЧ. Соответствующая вирусная нагрузка DIP, наоборот, может быть очень высокой из-за высокой множественности DIP-инфекции в клетке (§ 7.3).

DIP и ВИЧ асимметричны в трех отношениях:

(i) Образование вирусных частиц DIP полностью зависит от присутствия ВИЧ в той же клетке.

(ii) В ВИЧ-инфицированных клетках, экспрессия белка Vpr геном ВИЧ останавливает деление инфицированных клеток [495, 496], а экспрессия белка Nef подавляет суперинфекцию [496, 497]. Напротив, интегрированный ген DIP ничего не делает, потому что он не экспрессируется. Следовательно, DIP может делиться вместе с клеткой.

(iii) ВИЧ-инфицированные клетки живут один день, а DIP-инфицированные клетки столько же, сколько и неинфицированные клетки-мишени, то есть несколько месяцев [170, 346]. В результате, несколько копий провируса DIP могут накопиться в геноме клетки до того, как ВИЧ эту клетку заразит (§ 7.3).

Уравнения динамики уровня хозяина приведенные в исходной работе [21] имеют вид тот же вид что в § 7.3, (7.25)-(7.30), но с двумя дополнительными членами: неинфицированные клетки и DIP-инфицированные клетки могут делиться со скоростью hd , где d — скорость их естественной гибели, а h — безразмерное отношение частоты делений клетки к ее смертности. Скорость деления гомеостатична, т.е. зависит от общего количества всех клеток и уменьшается, когда оно приближается к нормальному уровню неинфицированного хозяина [[21], текст S1, уравнения (17), (18), (51), (52)]. Параметры модели на уровне хозяина приведены в таблице 7.6.

Эпидемиологический уровень

Теперь добавим к этим двум биоуровням, клетки и хозяина, третий (рисунок 7.16А, вверху). Данная эпидемиологическая модель описывает хорошо перемешанную популяцию людей с высоким риском заражения с тремя компартментами

$$\begin{aligned}\frac{1}{\delta_I} \frac{dS}{dt} &= \frac{1}{B} (1 - S) - R_0^{\text{pop}} (S I + \mu S I_D) \\ \frac{1}{\delta_I} \frac{dI}{dt} &= R_0^{\text{pop}} (S I + \mu S I_D - \phi I I_D) - I \\ \frac{1}{\delta_I} \frac{dI_D}{dt} &= R_0^{\text{pop}} \phi I I_D - \tau I_D\end{aligned}\tag{7.33}$$

Здесь три переменные — числа незараженных восприимчивых лиц $S(t)$, ВИЧ-инфицированных $I(t)$, и с коинфекцией ВИЧ и DIP $I_D(t)$, нормированные к числу восприимчивой популяции S в отсутствие ВИЧ и DIP. Те же уравнения в абсолютных единицах даны в исходной работе [21].

Уравнения (7.33) включают следующие эпидемиологические процессы. Восприимчивые люди S входят в группу высокого риска с постоянной линейной скоростью и покидают ее с фиксированной частотой. Они могут быть инфицированы ВИЧ лицами с инфекцией или коинфекцией (с разной вероятностью) и перейти в компартмент I . В этом компартменте, они могут быть коинфицированы DIP людьми с коинфекцией и сами перейти в этот компартмент, I_D . ВИЧ-инфекция прогрессирует в СПИД, что описывается как удаление из популяции. Суперинфекция DIP замедляет прогрессирование СПИДа и снижает передачу ВИЧ (за счет снижения вирусной нагрузки ВИЧ). DIP реплицируются только в коинфицированных. Предполагается, что DIP заражают только уже инфицированных ВИЧ пациентов. Если это предположение снять, DIP имеют более широкую область динамической устойчивости (рисунок 7.16В, С).

Параметры модели определены следующим образом: δ_I — средняя скорость прогрессирования СПИДа у ВИЧ-инфицированных, где δ_I/B — частота исключения из группы высокого риска; R_0^{pop} — базовый показатель репродукции ВИЧ в популяции; μ и ϕ — отношения вероятности передачи ВИЧ и DIP от коинфицированных к вероятности передачи ВИЧ от однократно инфицированных; τ — снижение скорости прогрессирования до СПИДа из-за подавления ВИЧ с помощью DIP. Оценки значений параметров и литературные источники приведены в таблице 7.5.

Три биологических уровня (клетка, хозяин и популяция хозяина) связываются друг с другом, предполагая стационарное состояние на двух нижних уровнях и используя стационарные значения переменных состояния в качестве входных данных для модели более высокого уровня. Это предположение оправдано, поскольку процессы на низких уровнях протекают намного быстрее, чем на высоких (минуты-часы, дни-месяцы и годы-декады, соответственно). Начиная с входных параметров η, κ , и P , модель уровня клетки предсказывает количество вирионов, продуцируемых инфицированной клеткой за ее жизнь, для ВИЧ и DIP (таблица 7.3). Принимая это в качестве исходных данных, модель на уровне хозяина предсказывает стационарное

количество вируса для ВИЧ и DIP (таблица 7.4). Этот результат, в свою очередь, используется для расчета эпидемиологических параметров в уравнениях (7.33), а именно, времени прогрессирования к СПИД и вероятностей передачи ВИЧ и DIP, используя результаты эпидемиологических исследований [225, 493]. Результатом верхнего уровня является доля ВИЧ-инфицированных и ВИЧ-DIP-коинфицированных лиц в популяции, рассчитанная как функция η и P , которая может служить для оценки терапевтической ценности DIP как самораспространяющегося агента. Надежность прогнозов по отношению к вариациям модели обсуждается в § 7.4.5.

7.4.2 Условия сосуществования ВИЧ и DIP в группах высокого риска

Уровень хозяина

Улучшенная модель качественно воспроизводит результаты § 7.3. ВИЧ и DIP достигают динамически стабильного стационарного состояния, и ВИЧ стабильно подавляется. Включение деления DIP+ВИЧ-клеток в модель хозяина [[21], текст S1, раздел B] расширяет область стабильности DIP по сравнению со старым условием стабильности, которое требует $\eta > 1$ (§ 7.3). Если число делений клетки за среднее время ее жизни h больше, чем $1 + R_{00}/P$, где $R_{00} = 10$ — число репродукции вируса на уровне хозяина при $\eta = 1, \kappa = 0$ (табл. 7.6), DIP динамически стабилен уже при $\eta > 0.1$ [[21], текст S1, формула (50)]. При меньших значениях η , нестабилен сам ВИЧ, так что эта область значений биологически бессмысленна (рисунок 7.16B,C, вертикальная черная полоса). В то же время, для стабильности требуется достаточно большая асимметрия экспрессии провируса DIP по отношению к ВИЧ, P . При малых значениях P , ВИЧ может размножаться в организме хозяина, а DIP — нет (рисунок 7.16B, слева). Максимальное значение P , при котором DIP нестабилен, составляет $P_{crit} \approx 3$, что соответствует $\eta \approx 1$. Условие $P > P_{crit}$ представляет собой важное ограничение на конструкцию DIP. Ниже рассматривается область $P > P_{crit}$.

Популяция хозяев

Для популяции высокого риска, точные границы области стабильности зависят от распространенности ВИЧ до того, как в нее была введена DIP (рисунок 7.16B, справа), и от того, могут ли DIP преинфицировать людей до того, как они были инфицированы ВИЧ (рисунок 7.16C) [334].

Если DIP не может заражать людей до инфекции ВИЧ, субпопуляция, доступная для DIP-инфекции, состоит только из ВИЧ-инфицированных. Поскольку распространенность ВИЧ до DIP зависит от базового числа репродукции в популяции, R_0^{pop} [191], эффективное число

репродукции DIP, обозначаемое R_{eff} , также от него зависит. В этом случае, условие распространения и устойчивости DIP имеет вид

$$R_{\text{eff}} = R_0^{\text{pop}} - \frac{B\tau}{\phi} > 1 \quad (7.34)$$

(параметры определены в таблице 7.5). Когда распространенность ВИЧ увеличивается из-за увеличения R_0^{pop} , второй член в (7.34) становится малой поправкой, а область стабильности DIP расширяется и приближается к области стабильности на уровне хозяина. Например, если 60% популяции высокого риска являются ВИЧ-инфицированными, стабильность DIP у индивидуума обычно подразумевает его стабильность в популяции (рисунок 7.20B).

Условие устойчивости (7.34) предполагает, что DIP может суперинфицировать только лиц, ранее инфицированных ВИЧ. Если же DIP могут преинфицировать индивидуумов и оставаться латентными до последующего инфицирования ВИЧ [334], что подтверждается исследованиями латентности ВИЧ [19] (*глава 6*), то распространенность ВИЧ гораздо меньше влияет на область стабильности DIP (рисунок 7.16C). С преинфекцией или без, но всякий раз, когда DIP стабильно сосуществует с ВИЧ, заболеваемость и распространенность ВИЧ-инфекции в популяции будут сильно подавлены ([21], рисунок S2). Эволюционная устойчивость этого подавления анализируется ниже.

7.4.3 Противоположные давления отбора на разных уровнях

Чтобы проанализировать эволюцию мутантов ВИЧ, устойчивых к DIP, делается мысленный эксперимент: мутант ВИЧ вводится в популяцию хозяев, инфицированных диким типом ВИЧ и DIP, и конкуренция мутанта с диким типом анализируется. Например, эти два штамма ВИЧ могут различаться значением параметра η . Параметр P , как и в § 7.3, предполагается фиксированным. Как и в анализе динамической стабильности, используется разделение масштабов времени: процессы на уровне одного зараженного человека происходят намного быстрее, чем на уровне населения.

На каждом биологическом уровне, могут распространяться мутанты ВИЧ двух типов, устойчивые к DIP: те, которые увеличивают количество ВИЧ в инфицированном хозяине, и те, которые увеличивают распространенность ВИЧ в населении. Поскольку ВИЧ и DIP динамически связаны на всех биоуровнях, заранее не очевидно, распространится ли резистентный мутант в индивидууме или популяции. В § 7.3, коэффициент отбора мутантов был получен для отдельного хозяина. Ниже давление отбора вычисляется также для уровня популяции и сравнивается между двумя биоуровнями.

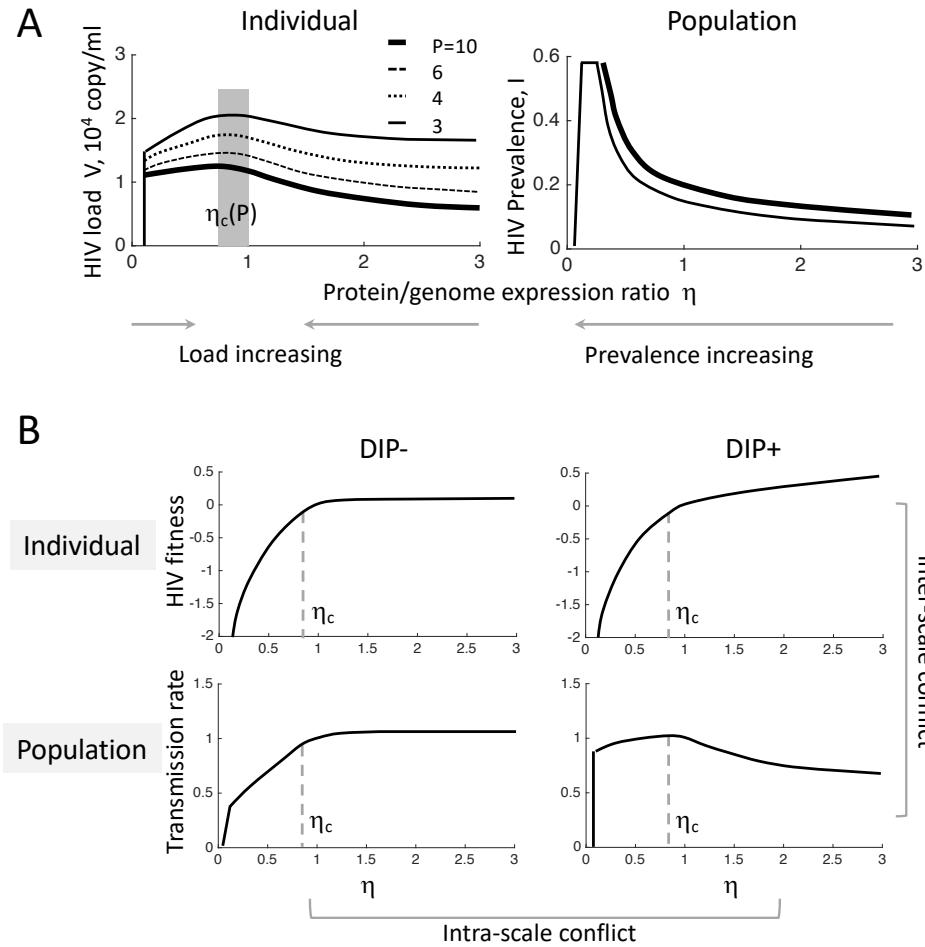


Рисунок 7.17: Устойчивые к DIP мутанты ВИЧ сталкиваются с противоречивым давлением отбора. (А) Стационарная ВИЧ-нагрузка в организме хозяина, нормализованная к своему максимальному значению (слева), и стационарная распространенность ВИЧ в популяции высокого риска (справа), в зависимости от параметров P и η . При $P > 3$, нагрузка ВИЧ максимальна при $\eta = \eta_{max} \approx 1$ (серый прямоугольник). Для $P < 3$, см. [[21], рисунок S2]. (В) Различный адаптивный ландшафт по параметру η в хозяине (наверху) и популяции хозяев (внизу) для DIP- клеток или DIP- индивидуумов (слева) и DIP+ клеток или DIP+ индивидуумов (справа). Индивидуальная приспособленность на уровне хозяина рассчитана на основе скорости расширения или сокращения мутантов ВИЧ с коэффициентом отбора (7.32). Скорость передачи вируса в популяции рассчитана на основе стационарной вирусной нагрузки [21]. И приспособленность, и скорость передачи нормализованы к их значениям при η_{max} и $P = 6$, а график репрезентативен для интервала $P > 3$ ([21], рисунок S3). Таким образом, присутствие DIP+ людей и DIP- людей оказывает противоположное давление отбора на значение η в популяции. По работе [21].

В работе [21], рассмотрены мутации меняющие асимметрию экспрессии гена между DIP и ВИЧ, параметр η . ВИЧ-нагрузка у коинфицированного ВИЧ+DIP+ человека достигает максимума при значении $\eta = \eta_{max}$, несколько меньшем $\eta = 1$ (рисунок 7.17А, слева). Мутанты ВИЧ с $\eta = \eta_{max}$ являются лучшими кандидатами на устойчивость к DIP, потому что любая мутация η в сторону η_{max} увеличивает нагрузку ВИЧ у хозяина с двойной инфекцией и, следовательно, вероятность передачи ВИЧ другому лицу. С другой стороны мутант, который снижает η , увеличивает распространенность ВИЧ в популяции имеющей DIP (рисунок 7.17А, справа). Оба эти мутанта, мутант, увеличивающий вирусную нагрузку, и мутант, увеличивающий распространенность, могли бы привести к снижению подавления ВИЧ.

Более того, эти мутанты могли бы привести систему к полному сопротивлению DIP на каждом биоуровне, т.е., в ту область η , где DIP нестабилен (рисунки 7.16В-С и 7.17А). Если $P > 3$, полная резистентность хозяина к лечению DIP могла бы возникнуть только при очень низких значениях $\eta \sim 0.1$, близких к порогу исчезновения ВИЧ. Однако на уровне популяции, резистентность могла бы проявляться, в принципе, и в более широком диапазоне.

Целью является расчет пользы или вреда этих мутантов для ВИЧ с точки зрения естественного отбора. Выше в § 7.3 описано, как можно определить адаптивный ландшафт на уровне хозяина по скорости расширения или сокращения мутантного штамма, введенного в небольшом количестве в стационарного хозяина, коинфицированного ВИЧ и DIP. Приспособленность варианта ВИЧ зависит от выхода вируса из отдельных клеток и распределения множественности DIP среди клеток инфицированного человека. Более высокое η всегда соответствует большему количеству вирусного потомства на клетку, независимо от множественности DIP. Следовательно, мутанты ВИЧ с более высокими значениями η всегда выгодны в хозяине, независимо от наличия или существовавшего ранее уровня DIP (рисунок 7.17В, верхний ряд; [21], рисунок S3). Этот результат представляет собой пример «трагедии общего достояния» [498]: мутация в ВИЧ, усиливающая выработку капсида, полезна для самого мутанта, но она также усиливает эксплуатацию всех остальных вариантов ВИЧ в хозяине DIPом. (Уступить рэкету значит обречь весь квартал на рэкет, отсюда необходимость кооперации у людей, которой нет у вирусов.)

В присутствии DIP, давление отбора зависит от биологического уровня. Число репродукции ВИЧ на уровне популяции R_0^{pop} определяется скоростью его передачи и, следовательно, вирусной нагрузкой ВИЧ в хозяине [225]. Коэффициенты передачи ВИЧ и DIP, представленные в уравнениях (7.33) и таблице 7.5, рассчитываются на основе вирусной нагрузки. Далее, выражая вирусную нагрузку через параметры в клетке, как описано в § 7.3, скорость передачи вариантов ВИЧ выражается через η и P [21]. Важно, что лица, инфицированные только ВИЧ, и лица, инфицированные как ВИЧ, так и DIP, имеют разные скорости передачи ВИЧ (рисунок 7.17В, нижний ряд). У инфицированных с DIP, скорость передачи и вирусная нагрузка имеют максимум при $\eta = \eta_{max}$, напротив, у инфицированных без DIP, передача ВИЧ всегда увеличивается с η .

Это противоречивое давление отбора толкает эволюцию в разных направлениях по оси η . В целом модель предсказывает два эволюционных конфликта: межуровневый конфликт между адаптивными ландшафтами на уровне хозяина и популяции и внутриуровневый конфликт между передачей ВИЧ от лиц с DIP и без (рисунок 7.17В).

7.4.4 Эволюционные конфликты останавливают появление DIP-устойчивых мутантов ВИЧ

Следующая задача — определить, какие мутанты ВИЧ могли бы распространяться среди населения в условиях этих конфликтов. С этой целью, модель популяционного уровня обобщается на два штамма ВИЧ: исходный штамм в стационарном состоянии и потенциально DIP-резистивный штамм (рисунок 7.18А). Когда два штамма заражают одного и того же хозяина, более приспособленный штамм быстро фиксируется (вертикальные стрелки на рисунок 7.18А). Такая коинфекция часто наблюдается в группе высокого риска [6]. Конкуренционное исключение внутри хозяина (§ 2.1) происходит очень быстро по сравнению с динамикой распространенности вируса в популяции. В каждый момент времени, лиц с транзиторной коинфекцией двумя штаммами ВИЧ немного, и ими можно пренебречь.

Для того чтобы мутантный ВИЧ распространился в популяции, польза от мутации для вируса должна перевешивать стоимость. Предположим, что мутантный вариант ВИЧ с отношением продукции капсида к геному $\eta_{mut} = 1$, который является наиболее обнадеживающим устойчивым мутантом, конкурирует с исходным (диким) штаммом ВИЧ с $\eta_{wt} = 2$. Такой мутант увеличивает вирусную нагрузку ВИЧ в человеке с коинфекцией и, следовательно, передается от него лучше, чем от дикого типа (рисунок 7.17В, нижний ряд). В то же время, мутант с $\eta_{mut} = 1$ проигрывает дикому типу при передаче от хозяев без DIP, а также будет менее приспособленным внутри любого хозяина, как с DIP так и без DIP (рисунок 7.17В). В модель с двумя штаммами позволяет рассчитать выгоды и стоимость мутации и в итоге демонстрирует, что мутант не может распространиться, если DIP сконструированы с $P \sim 3$ или выше [21] (рисунок 7.18В).

Увеличение расчетного значения P увеличивает преимущество передачи мутанта ВИЧ в хозяевах с DIP, но не меняет его в хозяевах без DIP. Таким образом, значение P контролирует величину внутриуровневого конфликта. Межуровневый конфликт контролируется включением и выключением, в математической модели, коинфекции пациентов разными вирусами. Межуровневый конфликт отсутствовал бы, если бы коинфекции не было, но она часто наблюдается у пациентов с высоким риском заражения и оказывает сильное потенциальное влияние на эволюцию ВИЧ [6] (§ 2.1). Когда асимметрия экспрессии P снижена или коинфекция отсутствует, мутант распространяется лучше (рисунок 7.18В). Итак, для предотвращения распространения DIP-устойчивых мутантов в популяции, необходимы как высокий уровень $P > 3$, так и коинфекция.

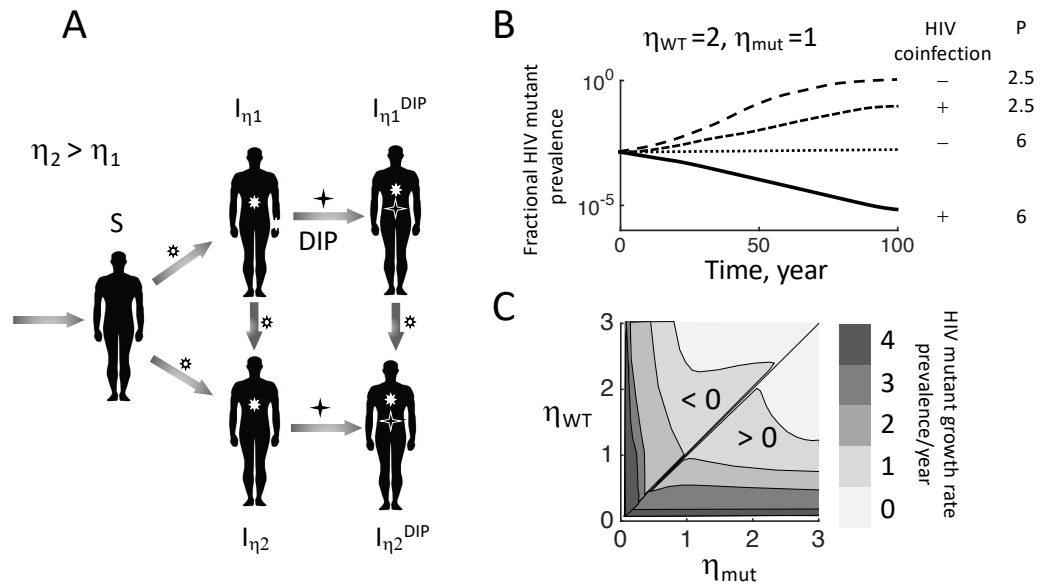


Рисунок 7.18: Из-за межуровневого и внутриуровневого эволюционного конфликта, DIP-резистивные мутанты ВИЧ не могут размножаться в популяции хозяев. (А) Схема эволюционной модели с конфликтом. Два мутанта ВИЧ с разными значениями η конкурируют за передачу в популяции, получающей DIP. Когда два штамма ВИЧ заражают одного и того же человека, штамм с более высоким значением η замещает другой штамм благодаря своей более высокой приспособленности (рисунок 7.17В). (В) Распространенность мутанта ВИЧ со сниженным η в зависимости от времени ($\eta_{wt} = 2, \eta_{mut} = 1$) с начальной долей мутанта 0.1% и двумя значениями P (показано). Показаны два случая, с коинфекцией штамма ВИЧ и без нее. Коинфицирование хозяина двумя штаммами ВИЧ приводит к отбору штаммов ВИЧ с повышенными значениями η , несмотря на цену увеличения η в скорости передачи у DIP+ВИЧ+ индивидуумов (межуровневый конфликт, рисунок 7.17В). Противоположные давления отбора на η со стороны присутствия особей DIP+ и DIP- (внутриуровневый конфликт, рисунок 7.17В) более значительны для $P = 6$, чем для $P = 2.5$, где это очень слабый эффект. (С) Начальная скорость расширения или сокращения мутантного штамма ВИЧ с $\eta = \eta_{mut}$ при введении в популяцию, инфицированную как DIP, так и ВИЧ дикого типа, $\eta = \eta_{wt}$. Каждая комбинация мутантного ВИЧ и ВИЧ дикого типа представлена точкой (η_{mut}, η_{wt}) на плоскости. В каждой точке затенение карты показывает максимальное собственное значение якобиана для мутантного штамма по всем значениям $P > 3$. Таким образом, могут расти только мутанты с $\eta_{mut} > \eta_{wt}$, а ВИЧ не может эволюционировать вдали от DIP путем лишения его белка, если $P > 3$. По работе [21].

Для определения более общего условия на параметры двух штаммов η_{mut} и η_{wt} , рассчитана начальная скорость размножения мутантов ВИЧ и найдено ее наибольшее значение по P в интервале $P > 3$ (рисунок 7.18С). Обобщая результат, получается что мутант ВИЧ с $\eta_{mut} < \eta_{wt}$ вообще никогда не распространится (рисунок 7.18С), если $\eta_{wt} > \eta_{max}$ где $\eta_{max} \approx 1$ из (7.17). Такие мутанты, увеличивающие вирусную нагрузку, могут размножаться в обоих биологических уровнях только тогда, когда $\eta_{wt} < \eta_{max}$ и все давления отбора совпадают ([21], рисунок S5). В этом интервале, увеличение η приводит к увеличению нагрузки ВИЧ и скорости передачи. Однако этот эффект присутствует и без DIP и является улучшением самого ВИЧ, а не выработкой резистентности, и, следовательно, не представляет для нас интереса.

7.4.5 Приближения

Приближения данного анализа были проверены в исходной работе [21], включая форму гомеостатической функции выключения пополнения клеток, маленький параметр отходов $k \ll 1$, кража только одного белка, и мономорфизм двух штаммов ВИЧ. В частности:

(i) Редкая мутация. В модели, мутанты ВИЧ внедряются в пациентов только в результате коинфекции, игнорируя образование новых мутантов ВИЧ в организме хозяина. Это предположение является представляет собой наилучший сценарий развития сопротивления к DIP в масштабе популяции.

(ii) Быстрый отбор в хозяине. При совместном заражении хозяина двумя штаммами, штамм с более высокой приспособленностью мгновенно заменяет штамм с более низкой приспособленностью. Если значения η очень близки, это предположение может быть неверным. Однако даже при разнице с коэффициентом отбора 5% , временной масштаб конкуренции штаммов не превышает 1 года (§ 2.1), что намного короче временных масштабов эволюции пандемий.

(iii) Путь ускользания. Поскольку минимальный дизайн DIP не кодирует белки, он не может восстановить η до высоких значений, поэтому мутация, влияющая на η , является наиболее многообещающим путем ускользания. Другой возможный параметр ускользания— это параметр отходов, k , который, в свою очередь, зависит от коэффициента упаковки k_{psc} и скорости образования гРНК, θ (таблица 7.3), оба из которых зависят от генома ВИЧ. В § 7.3, было найдено, что отбор на уровне хозяина только уменьшает k , и что, изменив его, ВИЧ не может избежать DIP. Этот факт, очевидно, остается на уровне населения. ВИЧ может также попытаться ускользнуть, изменив скорость производства геномов θ , которая напрямую влияет на число вирусных потомков в клетке n и, следовательно, на R_{00} (таблицы 7.5 и 7.8). Например, мутации могут модифицировать белок Tat, ответственный за трансактивацию генов ВИЧ и DIP. Однако, изменение θ одинаково влияет на продукцию геномов DIP и ВИЧ, что делает его худшим кандидатом, чем асимметрично контролируемый η [21].

(iv) Отбор, зависящий от частоты аллеля. Вообще говоря, естественный отбор, воздействующие на ВИЧ на уровне населения, не может быть описан ландшафтом приспособленности. Другими словами, невозможно выразить скорость экспансии мутанта ВИЧ через разницу в приспособленности w между двумя штаммами, $s = w(\eta_2) - w(\eta_1)$. Действительно, из-за коинфекции хозяина и мгновенного конкурентного исключения в хозяине, скорость экспансии мутанта имеет член в виде ступеньки, который не зависит от величины $\eta_2 - \eta_1$, а только от его знака (рисунок 7.18А). Но даже и без коинфекции хозяина, s зависит от частоты DIP+ особей в популяциях, что является динамическим параметром, вызывающим

долговременные колебания ([21], рисунок S6). Таким образом, естественный отбор, на популяционном уровне, еще и зависит от частоты аллели. К счастью, обеими этими усложнениями можно смело пренебречь, если адаптивный ландшафт рассчитывается для небольших изменений η , как это и сделано в работе [21].

(v) Отсутствие эпистаза в модели (*глава 5*) остается наиболее важным ограничением анализа. Несмотря на многообещающие сдвиги с разработкой терапевтического DIP [30, 317-319, 321], прогресс идет медленно, и причиной может быть эпистаз, который позволяет ускользание вируса от DIP через долины приспособленности. Этот вопрос требует дополнительных исследований, на основе применения методов описанных в *главе 4*.

7.4.6 Заключение

Распространение в населении мутантов ВИЧ резистивных к DIP, предотвращается за счет снижения темпа роста в хозяине, а также вероятности передачи из людей не коинфицированных DIP. Любой штамм ВИЧ, который является динамически и эволюционно стабильным в хозяине, сохраняет это свойство и в масштабе популяции. Общим следствием эволюционных конфликтов между биологическими уровнями является возможность эволюционной стабильности ВИЧ в присутствии DIP на любом биологическом уровне, коль скоро DIP экспрессируется в коинфицированных клетках, по крайней мере, в 3 раза больше, чем ВИЧ. Таким образом, DIP мог бы быть эволюционно стабильным и самораспространяющимся антивирусным лечением, по крайней мере, в отсутствие эпистаза (*глава 5*). Моделирование эволюционной стабильности, учитывающее эпистаз, будет проведено в будущем.

Заключение

В результате работ, вошедших в данную диссертацию, ускорено соединение генетики популяций и вирусологии, с использованием методов статистической физики. Разработан общий аналитический подход к многолокусной эволюции, как в половых, так и бесполовых популяциях. Разработаны компьютерные и аналитические модели, проливающие свет на главные свойства ВИЧ инфекции. Создана математическая теория эволюции вирусов в присутствии иммунного ответа и интерферирующих частиц. Разработан высоконадежный метод определения адаптивного ландшафта из геномных данных. Многие важные результаты, опубликованные до 2019 года, подтверждены другими группами.

Эти результаты также оказали значительное влияние на биомедицинские приложения. Доказав предыдущее предположение [2], что эффективный размер популяции ВИЧ достаточно велик, чтобы содержать любые одиночные мутации, $10^6 - 10^8$ зараженных клеток, эта работа объяснила быстрое появление лекарственной устойчивости на ингибиторы, действующие на один локус. В конце концов, внедрение тройных коктейлей ингибиторов решило эту проблему, так как тройные мутации слишком редки, чтобы существовать в зараженном человеке.

Теория эволюции иммунологически важных областей генома респираторных вирусов способствует разработке эффективных вакцин. Работа над ролью латентности ВИЧ окончательно показала, что искоренение вируса из организма не является достижимой целью, и вместо этого необходимо искать функциональную терапию. Работа над дефектными интерференционными частицами пока еще не привела к супрессивной терапии, но уже стимулировала разработку живых вакцин.

В настоящее время, дальнейшие исследования И.М.Рузина проводятся со студентами Санкт-Петербургского Политехнического Университета. Работа идет по следующим направлениям:

- Разработка пакета программного обеспечения для измерения эпистатического взаимодействия в любой адаптирующейся популяции, от вирусов до животных и человека.
- Применение нового и уже опубликованного программного обеспечения для измерения фитнес-ландшафта SARS-CoV-2.
- Использование измеренного адаптивного ландшафта в качестве исходных данных для моделей эволюции SARS-CoV-2; предсказание появления вызывающих озабоченность вариантов, в вероятностном смысле.
- Включение адаптивного ландшафта SARS-CoV-2 в предыдущую модель эволюции областей связывания антител, разработанную для гриппа.

- Использование результатов, чтобы оптимизировать выбор обновления вакцины и предложить тип вакцины (В-клеточная вакцина, Т-клеточная вакцина и т. д.).
- Эволюция резистивности ВИЧ к DIP с учетом эффектов эпистаза

Выводы

1. Эффективный размер популяции у среднего нелеченного ВИЧ-инфицированного составляет не менее миллиона активно-инфицированных клеток, поэтому естественный отбор в хозяине сильнее стохастических факторов, и невыгодные мутации придающие устойчивость к монотерапии существуют в популяции вируса уже до терапии.
2. Высокое разнообразие и скорость эволюции ВИЧ в хронической инфекции у нелеченных пациентов обусловлены мутациями ускользания в иммунологически-важных областях и мутациями, возникающими для их компенсации. Порядок мутаций ускользания во времени определяется их положением в плоскости «стоимость-выгода». Средний коэффициент отбора компенсирующей мутации посередине хронической инфекции несколько меньше одного процента.
3. В бесполовой популяции, эффекты генетического сцепления вызывают потерю полезных аллелей и замедляют адаптацию во много раз, пропорционально длине генома. Существует оптимальная вероятность мутации, при которой скорость адаптации оптимальна между накоплением выгодных и невыгодных аллелей. Это предсказание проверено экспериментально на вирусе полиомиелита у мышей.
4. Частично компенсируя эффекты генетического сцепления, даже редкая рекомбинация позволяет ускорить адаптацию на много порядков. В хронической нелеченной ВИЧ-инфекции, вероятность рекомбинации примерно 1%. Наличие общей родословной гомологичных локусов снижает эффективность рекомбинации. Даже очень частая рекомбинация не делает эволюцию далеко расположенных локусов независимой.
5. Низкая надежность существующих методов обнаружения эпистаза в образцах ДНК вызвана эффектом генетического сцепления. Усреднение по небольшому числу популяций и использование дополнительных локусов позволяет исключить этот эффект и достичь идеальной надежности обнаружения, по крайней мере, для геномов меньше сотни локусов.
6. Причиной СПИДа, парадоксально, является адаптация ВИЧ к индивидууму. Эволюционное преимущество латентного состояния ВИЧ заключается в усилении передачи вируса между людьми.
7. Давление отбора создаваемое иммунной памятью населения на респираторный вирус, выражается, в общем виде, через только два параметра: базовое число передачи в наивном населении и перекрестную реактивность антител.
8. В отсутствие эпистаза, вирус не может развить устойчивость к своей дефектной интерференционной частице ни на уровне хозяина, ни на уровне популяции.

Указатель терминологии

- С**
- CD4 Т-клеток, 31
- Д**
- D79Н, 191
DFE, 150
- А**
- адаптация, 41
адаптивный ландшафт, 24
активация клеток, 182
алгоритм Райта-Фишера, 62
алгоритм сломанной палочки, 62
аллель, 18
аллельная ассоциация, 122
аллостерический, 22
антигенное ускользание, 31
антигенная координата, 201
анти-ретровирусная терапия, 9
- Б**
- базовое число репродукции, 172
- В**
- вероятности латентности, 173
вероятность передачи вируса, 175
вероятность фиксации, 89
вероятность установления *restab*, 176
вероятность рекомбинации, 61
ветвящийся процесс, 177
вирус иммунодефицита обезьян, 28
- вируса иммунодефицита обезьян, 29
вложенная ПЦР-амплификация, 37
внутренняя плотность распределения, 151
время выборки, 43
вставка, 37
выравнивание, 65
высокоточный мутант G64S, 191
- Г**
- гаплотип, 16
гаплотип, 56
генератор псевдослучайных чисел, 54
генетики популяций, 9
генетическая линия, 89
генетический полиморфизм, 39
генетический дрейф, 16
генетического сцепления, 17
генетическое разнообразие, 15
генетическое расстояние, 38
гетерозиготность, 38
главный комплекс гистосовместимости
 класса I, 36
годовая заболеваемость, 205
- Д**
- двух-локусная модель, 56
двухкомпонентная модель, 179
делеция, 37
дерево типа Больтхаузена-Шнитмана, 137
дерегуляция гомеостаза, 169
детерминистическое уравнение, 41
дефектная интерферирующая частица, 33

дикий тип, 41
 динамика межгеномных корреляций, 113
 дифференциальное уравнение в частных
 производных, 77
 долина приспособленности, 22
 доля мутантов, 42
 доля локусов с одним и тем же предком, 107

Е

естественный отбор, 14

З

замена, 39

И

иммунное распознавание, 30
 инбридинг, 105
 ингибиторы протеазы, 36
 инфекционная доза, 180
 инфекцию слизистой оболочки, 173
 оценка коэффициентов отбора, 157

К

квазиравновесие сцепления, 23
 квазиравновесие между эпистазом и
 беспорядком, 120
 класс приспособленности, 76
 кластер невыгодных аллелей, 125
 клона, 20
 клональной интерференции, 16
 ключевое ограничение метода QLE, 149
 компенсаторная мутация, 146
 консенсусная последовательность, 37
 коррелятор, 126
 коэффициент отбора, 19
 коэффициент отбора s , 42

коэффициент частичного ускользания, 32
 край высокой приспособленности, 78

Л

латентное, 27
 лидер волны, 110
 линия потомков аллеля, 89
 локус, 9

М

максимум по вероятности мутации, 193
 Марковская цепочка, 44
 межгеномные корреляции, 105
 миноритарный аллель, 38
 многолокусной эволюции, 19
 модель структурированная по штаммам, 202
 модель с одним локусом, 52
 мутантный, 41
 мутация, 14

Н

наиболее приспособленный класс, 80
 наименее распространенный гаплотип, 58
 накопление мутантов, 42
 начальное условие, 42
 нейроминидаза вируса гриппа, 140
 нейтральном пределе, 56
 не прямое взаимодействие, 142
 несинонимичный, 15
 низкоточный мутант H273R, 191

О

общий предок, 137
 ОДУ, 41
 оптимальной вероятности латентности, 172
 отбор на разнообразие, 15

отрицательный отбор, 14

очищающий отбор, 14

П

пандемический штамм, 145

параметр затухания клона, 110

первичная мутация, 146

передача от хозяина к хозяину, 29

передний фронт волны, 101

плотность распределения коэффициентов отбора, 150

плотность событий слияния генетических линий, 107

полимераза, 17

полиморфизм, 15

полиморфное основание, 39

полиморфный локус, 39

полиномиальное распределение, 62

полиовирус, 30

положительный отбор, 15

полудетерминистический подход, 75

полупериод адаптации, 45

популяция хозяев, 34

приспособленность, 14

провирусной ДНК, 37

прогрессирование к СПИДу, 161

процент ложных пар, 136

процент обнаружения эпистатических пар, 136

путешествующий адаптивный ландшафт, 206

Р

размер популяции, 51

распределение полезных мутаций по эффекту на приспособленность, 25

распределение приспособленности предков, 112

расстояния перекрестного иммунитета, 207

регуляторная цепь, 28

рекомбинация, 17, 59

рекомбинационный перекрест, 63

бегущая волна, 77

С

свиной грипп, 24

синдром приобретенного иммунодефицита, 25

синонимичные, 15

системную инфекцию, 173

скорость эволюции, 8

среднее времени до последнего общего предка, 116

стационарное состояние, 43

стоп-кодон, 37

стохастический порог, 55

стохастическими эффектами, 16

структура генетических линий, 109

Т

топология сети, 128

точечная мутация, 37

точка полной компенсации, 127

трех-локусная корреляция, 142

У

уединенная волна, 65

узкое место передачи, 36

универсальный след эпистаза, 120

уравнение Колмогорова, 54

устойчивость к противовирусным
препаратам, 8

Ф

филогенез, 21

фундаментальная теорема Фишера, 78

функция перекрестного иммунитета, 202

функция генерации рекомбинантов, 99

Х

храповик Мюллера, 18

храповик Мюллера, 81

Ц

цена мутации, 32

цепочка передачи, 44

цитотоксические лимфоциты, 213

Ч

частота аллелей, 38

частота замен, 102

число потомков, 14

Э

эволюция ВИЧ, 36

эволюционная стабильность, 33

эксперимент по конкуренции роста, 55

энтропия, 122

эпистаз, 21, 119

эпистатическая сеть, 122

эпитоп, 30

эффект генетического фона, 16

эффективный коэффициент отбора, 87

эффективный размер популяции, 16

Библиография

1. Rouzine I. M., Rozhnova G. Evolutionary implications of SARS-CoV-2 vaccination for the future design of vaccination strategies // *Commun Med (Lond)*. – 2023. – V. 3, № 1. – P. 86.
2. Coffin J. M. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy // *Science*. – 1995. – V. 267. – P. 483-488.
3. *Mathematical Modeling of Evolution. Volume 1. One-Locus and Multi-Locus Theory and Recombination*. De Gruyter Series in Mathematics and Life Sciences. / Rouzine I. M. – Berlin: De Gruyter, 2020. De Gruyter Series in Mathematics and Life Sciences.
4. *Mathematical Modeling of Evolution. Volume 2. Fitness Landscape, Red Queen, Evolutionary Enigmas, and Applications to Virology*. De Gruyter Series in Mathematics and Life Sciences. / Rouzine I. M. – Berlin: De Gruyter, 2023. De Gruyter Series in Mathematics and Life Sciences.
5. Rouzine I. M., Coffin J. M. Linkage disequilibrium test implies a large effective population number for HIV in vivo // *Proc Natl Acad Sci U S A*. – 1999. – V. 96, № 19. – P. 10758-63.
6. Rouzine I. M., Coffin J. M. Search for the mechanism of genetic variation in the pro gene of human immunodeficiency virus // *J Virol*. – 1999. – V. 73, № 10. – P. 8167-78.
7. Rouzine I. M., Wakeley J., Coffin J. M. The solitary wave of asexual evolution // *Proc Natl Acad Sci U S A*. – 2003. – V. 100, № 2. – P. 587-92.
8. Rouzine I. M., Coffin J. M. Evolution of human immunodeficiency virus under selection and weak recombination // *Genetics*. – 2005. – V. 170, № 1. – P. 7-18.
9. Gheorghiu-Svirschevski S., Rouzine I. M., Coffin J. M. Increasing sequence correlation limits the efficiency of recombination in a multisite evolution model // *Mol Biol Evol*. – 2007. – V. 24, № 2. – P. 574-86.
10. Rouzine I. M., Coffin J. M. Highly fit ancestors of a partly sexual haploid population // *Theor Popul Biol*. – 2007. – V. 71, № 2. – P. 239-50.
11. Rouzine I. M., Brunet E., Wilke C. O. The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation // *Theor Popul Biol*. – 2008. – V. 73, № 1. – P. 24-46.
12. Brunet E., Rouzine I. M., Wilke C. O. The stochastic edge in adaptive evolution // *Genetics*. – 2008. – V. 179, № 1. – P. 603-20.
13. Dutta R. N., Rouzine I. M., Smith S. D., Wilke C. O., Novella I. S. Rapid adaptive amplification of preexisting variation in an RNA virus // *J Virol*. – 2008. – V. 82, № 9. – P. 4354-62.
14. Rouzine I. M., Coffin J. M. Multi-site adaptation in the presence of infrequent recombination // *Theor Popul Biol*. – 2010. – V. 77, № 3. – P. 189-204.
15. Batorsky R., Kearney M. F., Palmer S. E., Maldarelli F., Rouzine I. M., Coffin J. M. Estimate of effective recombination rate and average selection coefficient for HIV in chronic infection // *Proc Natl Acad Sci U S A*. – 2011. – V. 108, № 14. – P. 5661-6.
16. Good B. H., Rouzine I. M., Balick D. J., Hallatschek O., Desai M. M. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations // *Proc Natl Acad Sci U S A*. – 2012. – V. 109, № 13. – P. 4950-5.
17. Rouzine I. M., Weinberger L. S. Design requirements for interfering particles to maintain co-adaptive stability with HIV-1. // *Journal of Virology*. – 2013. – V. 87, № 4. – P. 2081-2093.
18. Batorsky R., Sergeev R. A., Rouzine I. M. The route of HIV escape from immune response targeting multiple sites is determined by the cost-benefit tradeoff of escape mutations // *PLoS Comput Biol*. – 2014. – V. 10, № 10. – P. e1003878.
19. Rouzine I. M., Weinberger A. D., Weinberger L. S. An evolutionary role for HIV latency in enhancing viral transmission // *Cell*. – 2015. – V. 160, № 5. – P. 1002-12.
20. Razooky B. S., Pai A., Aull K., Rouzine I. M., Weinberger L. S. A hardwired HIV latency program // *Cell*. – 2015. – V. 160, № 5. – P. 990-1001.

21. Rast L. I., Rouzine I. M., Rozhnova G., Bishop L., Weinberger A. D., Weinberger L. S. Conflicting selection pressures will constrain viral escape from interfering particles: Principles for designing resistance-proof antivirals // *PLoS Comput Biol.* – 2016. – V. 12, № 5. – P. e1004799.
22. Xiao Y., Rouzine I. M., Bianco S., Acevedo A., Goldstein E. F., Farkov M., Brodsky L., Andino R. RNA recombination enhances adaptability and is required for virus spread and virulence // *Cell Host Microbe.* – 2016. – V. 19, № 4. – P. 493-503.
23. Lidsky P. V., Andino R., Rouzine I. M. Variability in viral pathogenesis: modeling the dynamic of acute and persistent infections // *Curr Opin Virol.* – 2017. – V. 23. – P. 120-124.
24. Pedruzzi G., Barlukova A., Rouzine I. M. Evolutionary footprint of epistasis // *PLoS Comput Biol.* – 2018. – V. 14, № 9. – P. e1006426.
25. Rouzine I. M., Rozhnova G. Antigenic evolution of viruses in host populations // *PLoS Pathog.* – 2018. – V. 14, № 9. – P. e1007291.
26. Pedruzzi G., Rouzine I. M. Epistasis detectably alters correlations between genomic sites in a narrow parameter window // *PLoS ONE.* – 2019. – V. 14. – P. e0214036.
27. Rouzine I. M. An Evolutionary Model of Progression to AIDS // *Microorganisms.* – 2020. – V. 8, № 11. – P. 1714.
28. Pedruzzi G., Rouzine I. M. An evolution-based high-fidelity method of epistasis measurement: Theory and application to influenza // *PLoS Pathog.* – 2021. – V. 17, № 6. – P. e1009669.
29. Barlukova A., Rouzine I. M. The evolutionary origin of the universal distribution of mutation fitness effect // *PLoS Comput Biol.* – 2021. – V. 17, № 3. – P. e1008822.
30. Shirogane Y., Rousseau E., Voznica J., Xiao Y., Su W., Catching A., Whitfield Z. J., Rouzine I. M., Bianco S., Andino R. Experimental and mathematical insights on the interactions between poliovirus and a defective interfering genome // *PLoS Pathog.* – 2021. – V. 17, № 9. – P. e1009277.
31. Likhachev I. V., Rouzine I. M. Measurement of selection coefficients from genomic samples of adapting populations by computer modeling // *STAR Protoc.* – 2023. – V. 4, № 1. – P. 101821.
32. Rouzine I. M. Long-range linkage effects in adapting sexual populations // *Sci Rep.* – 2023. – V. 13, № 1. – P. 12492.
33. Rouzine I. M., Rodrigo A., Coffin J. M. Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology // *Microbiol Mol Biol Rev.* – 2001. – V. 65, № 1. – P. 151-85.
34. Rouzine I. M., Weinberger L. The quantitative theory of within-host viral evolution [review] // *J. Stat. Mech.: Theory and Experiment.* – 2013. doi:10.1088/1742-5468/2013/01/P01009 № 1. – P. P01009.
35. Rouzine I. M., Coffin J. M., Weinberger L. S. Fifteen years later: hard and soft selection sweeps confirm a large population number for HIV in vivo // *PLoS Genet.* – 2014. – V. 10, № 2. – P. e1004179.
36. Rouzine I. M., Weinberger L. S. Reply to "Coadaptive stability of interfering particles with HIV-1 when there is an evolutionary conflict" // *J Virol.* – 2013. – V. 87, № 17. – P. 9960-2.
37. Rouzine I. M., Razooky B. S., Weinberger L. S. Stochastic variability in HIV affects viral eradication // *Proc Natl Acad Sci U S A.* – 2014. – V. 111, № 37. – P. 13251-2.
38. Wain-Hobson S. The fastest genome evolution ever described: HIV variation in situ // *Curr Opin Genet Dev.* – 1993. – V. 3, № 6. – P. 878-83.
39. Burns D. P., Desrosiers R. C. Envelope sequence variation, neutralizing antibodies, and primate lentivirus persistence // *Curr Top Microbiol Immunol.* – 1994. – V. 188. – P. 185-219.
40. Moore J. P., Cao Y., Leu J., Qin L., Korber B., Ho D. D. Inter- and intraclade neutralization of human immunodeficiency virus type 1: genetic clades do not correspond to neutralization serotypes but partially correspond to gp120 antigenic serotypes // *J Virol.* – 1996. – V. 70, № 1. – P. 427-44.
41. Wolfs T. F., de Jong J. J., Van den Berg H., Tijnagel J. M., Krone W. J., Goudsmit J. Evolution of sequences encoding the principal neutralization epitope of human immunodeficiency virus 1 is host dependent, rapid, and continuous // *Proc Natl Acad Sci U S A.* – 1990. – V. 87, № 24. – P. 9938-42.
42. Wolinsky S. M., Korber B. T., Neumann A. U., Daniels M., Kunstman K. J., Whetsell A. J., Furtado M. R., Cao Y., Ho D. D., Safrit J. T. Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection // *Science.* – 1996. – V. 272, № 5261. – P. 537-42.

43. Cleland A., Watson H. G., Robertson P., Ludlam C. A., Brown A. J. Evolution of zidovudine resistance-associated genotypes in human immunodeficiency virus type 1-infected patients // *J Acquir Immune Defic Syndr Hum Retrovirol.* – 1996. – V. 12, № 1. – P. 6-18.
44. Mayers D. L., McCutchan F. E., Sanders-Buell E. E., Merritt L. I., Dilworth S., Fowler A. K., Marks C. A., Ruiz N. M., Richman D. D., Roberts C. R., et al. Characterization of HIV isolates arising after prolonged zidovudine therapy // *J Acquir Immune Defic Syndr* (1988). – 1992. – V. 5, № 8. – P. 749-59.
45. Burns D. P., Desrosiers R. C. Selection of genetic variants of simian immunodeficiency virus in persistently infected rhesus monkeys // *J Virol.* – 1991. – V. 65, № 4. – P. 1843-54.
46. Pang S., Vinters H. V., Akashi T., O'Brien W. A., Chen I. S. HIV-1 env sequence variation in brain tissue of patients with AIDS-related neurologic disease // *J Acquir Immune Defic Syndr* (1988). – 1991. – V. 4, № 11. – P. 1082-92.
47. Simmonds P., Balfe P., Ludlam C. A., Bishop J. O., Brown A. J. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1 // *J Virol.* – 1990. – V. 64, № 12. – P. 5840-50.
48. Balfe P., Simmonds P., Ludlam C. A., Bishop J. O., Brown A. J. Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating mutations // *J Virol.* – 1990. – V. 64, № 12. – P. 6221-33.
49. Lamers S. L., Sleasman J. W., She J. X., Barrie K. A., Pomeroy S. M., Barrett D. J., Goodenow M. M. Independent variation and positive selection in env V1 and V2 domains within maternal-infant strains of human immunodeficiency virus type 1 in vivo // *J Virol.* – 1993. – V. 67, № 7. – P. 3951-60.
50. Kuiken C. L., Zwart G., Baan E., Coutinho R. A., van den Hoek J. A., Goudsmit J. Increasing antigenic and genetic diversity of the V3 variable domain of the human immunodeficiency virus envelope protein in the course of the AIDS epidemic // *Proc Natl Acad Sci U S A.* – 1993. – V. 90, № 19. – P. 9061-5.
51. McCutchan F. E., Artenstein A. W., Sanders-Buell E., Salminen M. O., Carr J. K., Mascola J. R., Yu X. F., Nelson K. E., Khamboonruang C., Schmitt D., Kieny M. P., McNeil J. G., Burke D. S. Diversity of the envelope glycoprotein among human immunodeficiency virus type 1 isolates of clade E from Asia and Africa // *J Virol.* – 1996. – V. 70, № 6. – P. 3331-8.
52. Yoshimura F. K., Diem K., Learn G. H., Jr., Riddell S., Corey L. Inpatient sequence variation of the gag gene of human immunodeficiency virus type 1 plasma virions // *J Virol.* – 1996. – V. 70, № 12. – P. 8879-87.
53. Lech W. J., Wang G., Yang Y. L., Chee Y., Dorman K., McCrae D., Lazzeroni L. C., Erickson J. W., Sinsheimer J. S., Kaplan A. H. In vivo sequence diversity of the protease of human immunodeficiency virus type 1: presence of protease inhibitor-resistant variants in untreated subjects // *J Virol.* – 1996. – V. 70, № 3. – P. 2038-43.
54. Najera I., Holguin A., Quinones-Mateu M. E., Munoz-Fernandez M. A., Najera R., Lopez-Galindez C., Domingo E. Pol gene quasispecies of human immunodeficiency virus: mutations associated with drug resistance in virus from patients undergoing no drug therapy // *J Virol.* – 1995. – V. 69, № 1. – P. 23-31.
55. Haase A. T. Population biology of HIV-1 infection: viral and CD4+ T cell demographics and dynamics in lymphatic tissues // *Annu Rev Immunol.* – 1999. – V. 17. – P. 625-56.
56. Mansky L. M., Temin H. M. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase // *J Virol.* – 1995. – V. 69, № 8. – P. 5087-94.
57. Leigh-Brown A. J. Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population // *Proc Natl Acad Sci U S A.* – 1997. – V. 94. – P. 1862-1865.
58. Coffin J. M. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy // *Science.* – 1995. – V. 267, № 5197. – P. 483-9.
59. The genetical theory of natural selection. / Fisher R. A. – Oxford, United Kingdom: Clarendon Press, 1958, 1930.
60. Muller H. Some genetic aspects of sex. // *Am. Nat.* . – 1932. – V. 66. – P. 118.

61. Lewontin R. C. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models // *Genetics*. – 1964. – V. 49, № 1. – P. 49-67.
62. Hill W. G., Robertson A. Linkage disequilibrium in finite populations // *Theor Appl Genet*. – 1968. – V. 38, № 6. – P. 226-31.
63. Population genetics, molecular evolution, and the neutral theory. Selected papers. / Kimura M.; Ed. Takahata N. – Chicago: The University of Chicago Press, 1994.
64. Kimura M. Diffusion models in population genetics // *J. Appl. Probab.* – 1964. – V. 1. – P. 177-232.
65. Gerrish P. J., Lenski R. E. The fate of competing beneficial mutations in an asexual population // *Genetica*. – 1998. – V. 102-103, № 1-6. – P. 127-44.
66. Rice W. R. Experimental tests of the adaptive significance of sexual recombination // *Nat Rev Genet*. – 2002. – V. 3, № 4. – P. 241-51.
67. The genetical theory of natural selection. / Fisher R. A. – Oxford, United Kingdom: Clarendon Press, 1958.
68. Otto S. P., Barton N. H. The evolution of recombination: removing the limits to natural selection // *Genetics*. – 1997. – V. 147, № 2. – P. 879-906.
69. Felsenstein J. The evolutionary advantage of recombination // *Genetics*. – 1974. – V. 78, № 2. – P. 737-56.
70. Muller H. J. Some genetic aspects of sex // *Am. Nat.* – 1932. – V. 66. – P. 118-128.
71. Maynard Smith J. What use is sex? // *J. Theor. Bio.* – 1971. – V. 30. – P. 319-335.
72. Desai M. M., Fisher D. S. Beneficial mutation selection balance and the effect of linkage on positive selection // *Genetics*. – 2007. – V. 176, № 3. – P. 1759-98.
73. Hallatschek O. The noisy edge of traveling waves // *Proc Natl Acad Sci U S A.* – 2011. – V. 108, № 5. – P. 1783-7.
74. Neher R. A., Kessinger T. A., Shraiman B. I. Coalescence and genetic diversity in sexual populations under selection // *Proc Natl Acad Sci U S A.* – 2013. – V. 110, № 39. – P. 15836-41.
75. Weissman D. B., Hallatschek O. The rate of adaptation in large sexual populations with linear chromosomes // *Genetics*. – 2014. – V. 196, № 4. – P. 1167-83.
76. Neher R. A., Shraiman B. I., Fisher D. S. Rate of adaptation in large sexual populations // *Genetics*. – 2010. – V. 184, № 2. – P. 467-81.
77. Neher R. A., Leitner T. Recombination rate and selection strength in HIV intra-patient evolution // *PLoS Comput Biol.* – 2010. – V. 6, № 1. – P. e1000660.
78. Hill W. G., Robertson A. The effect of linkage on limits to artificial selection // *Genet. Res.* – 1966. – V. 8. – P. 269-294.
79. Gerrish P. J., Lenski R. E. The fate of competing beneficial mutations in an asexual population // *Genetica*. – 1998. – V. 102/103. – P. 127-144.
80. Arjan J. A., Visser M., Zeyl C. W., Gerrish P. J., Blanchard J. L., Lenski R. E. Diminishing returns from mutation supply rate in asexual populations // *Science*. – 1999. – V. 283, № 5400. – P. 404-6.
81. Schiffels S., Szollosi G. J., Mustonen V., Lassig M. Emergent neutrality in adaptive asexual evolution // *Genetics*. – 2011. – V. 189, № 4. – P. 1361-75.
82. Whitlock M. C. Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles, and sexual selection // *Evolution*. – 2000. – V. 54, № 6. – P. 1855-61.
83. Lande R. Risk of population extinction from fixation of deleterious and reverse mutations // *Genetica*. – 1998. – V. 102/103. – P. 21-27.
84. Kimura M., Maruyama T. The mutational load with epistatic gene interactions in fitness // *Genetics*. – 1966. – V. 54, № 6. – P. 1337-51.
85. Haigh J. The accumulation of deleterious genes in a population - Muller's ratchet // *Theor. Popul. Biol.* – 1978. – V. 14. – P. 251-267.
86. Gordo I., Charlesworth B. The degeneration of asexual haploid populations and the speed of Muller's ratchet. // *Genetics*. – 2000. – V. 154. – P. 1379-1387.
87. Stephan W., Chao L., Smale J. G. The advance of Muller's ratchet in a haploid asexual population: approximate solutions based on diffusion theory // *Genet. Res.* – 1993. – V. 61, № 3. – P. 225-31.

88. Hallatschek O. The noisy edge of traveling waves // *Proc Natl Acad Sci U S A.* – 2010. – V. 108, № 5. – P. 1783-7.
89. Neher R. A., Hallatschek O. Genealogies of rapidly adapting populations // *Proc Natl Acad Sci U S A.* – 2013. – V. 110, № 2. – P. 437-42.
90. Walczak A. M., Nicolaisen L. E., Plotkin J. B., Desai M. M. The structure of genealogies in the presence of purifying selection: a fitness-class coalescent // *Genetics.* – 2012. – V. 190, № 2. – P. 753-79.
91. Imhof M., Schlotterer C. Fitness effects of advantageous mutations in evolving *Escherichia coli* populations // *Proc Natl Acad Sci U S A.* – 2001. – V. 98, № 3. – P. 1113-7.
92. Kassen R., Bataillon T. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria // *Nat Genet.* – 2006. – V. 38, № 4. – P. 484-8.
93. Acevedo A., Brodsky L., Andino R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing // *Nature.* – 2014. – V. 505, № 7485. – P. 686-90.
94. Stern A., Bianco S., Yeh M. T., Wright C., Butcher K., Tang C., Nielsen R., Andino R. Costs and benefits of mutational robustness in RNA viruses // *Cell Rep.* – 2014. – V. 8, № 4. – P. 1026-36.
95. Wrenbeck E. E., Azouz L. R., Whitehead T. A. Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded // *Nat Commun.* – 2017. – V. 8. – P. 15695.
96. Hegreness M., Shores N., Hartl D., Kishony R. An equivalence principle for the incorporation of favorable mutations in asexual populations // *Science.* – 2006. – V. 311. – P. 1615.
97. Fogle C. A., Nagle J. L., Desai M. M. Clonal interference, multiple mutations and adaptation in large asexual populations // *Genetics.* – 2008. – V. 180, № 4. – P. 2163-73.
98. Hey J. Selfish genes, pleiotropy and the origin of recombination // *Genetics.* – 1998. – V. 149, № 4. – P. 2089-97.
99. Barton N. H. Linkage and the limits to natural selection // *Genetics.* – 1995. – V. 140, № 2. – P. 821-41.
100. Maynard Smith J. M. What use is sex? // *J. Theor. Biol.* – 1971. – V. 30. – P. 319-335.
101. Pamilo P., Nei M., Li W. H. Accumulation of mutations in sexual and asexual populations // *Genet Res.* – 1987. – V. 49, № 2. – P. 135-46.
102. Charlesworth B. Mutation-selection balance and the evolutionary advantage of sex and recombination // *Genet Res.* – 1990. – V. 55, № 3. – P. 199-221.
103. Rouzine I. M. Long-range linkage effects in adapting sexual populations // *Scientific Reports.* – 2023. – V. in press.
104. Zuk O., Hechter E., Sunyaev S. R., Lander E. S. The mystery of missing heritability: Genetic interactions create phantom heritability // *Proc Natl Acad Sci U S A.* – 2012. – V. 109, № 4. – P. 1193-8.
105. Cordell H. J. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans // *Hum Mol Genet.* – 2002. – V. 11, № 20. – P. 2463-8.
106. Cordell H. J. Detecting gene-gene interactions that underlie human diseases // *Nature reviews. Genetics.* – 2009. – V. 10, № 6. – P. 392-404.
107. Wei W.-H., Hemani G., Haley C. S. Detecting epistasis in human complex traits // *Nature Reviews Genetics.* – 2014. – V. 15. – P. 722.
108. Alvarez-Castro J. M., Le Rouzic A., Andersson L., Siegel P. B., Carlborg O. Modelling of genetic interactions improves prediction of hybrid patterns--a case study in domestic fowl // *Genet Res (Camb).* – 2012. – V. 94, № 5. – P. 255-66.
109. Carlborg O., Jacobsson L., Ahgren P., Siegel P., Andersson L. Epistasis and the release of genetic variation during long-term selection // *Nat Genet.* – 2006. – V. 38, № 4. – P. 418-20.
110. Brem R. B., Storey J. D., Whittle J., Kruglyak L. Genetic interactions between polymorphisms that affect gene expression in yeast // *Nature.* – 2005. – V. 436. – P. 701.
111. Schuldiner M., Collins S. R., Thompson N. J., Denic V., Bhamidipati A., Punna T., Ihmels J., Andrews B., Boone C., Greenblatt J. F., Weissman J. S., Krogan N. J. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile // *Cell.* – 2005. – V. 123, № 3. – P. 507-19.

112. Segrè D., DeLuna A., Church G. M., Kishony R. Modular epistasis in yeast metabolism // *Nature Genetics*. – 2004. – V. 37. – P. 77.
113. Nijhuis M., Schuurman R., de Jong D., Erickson J., Gustchina E., Albert J., Schipper P., Gulnik S., Boucher C. A. Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy // *AIDS*. – 1999. – V. 13, № 17. – P. 2349-59.
114. Cong M. E., Heneine W., Garcia-Lerma J. G. The fitness cost of mutations associated with human immunodeficiency virus type 1 drug resistance is modulated by mutational interactions // *J Virol*. – 2007. – V. 81, № 6. – P. 3037-41.
115. Handel A., Regoes R. R., Antia R. The role of compensatory mutations in the emergence of drug resistance // *PLoS Comput Biol*. – 2006. – V. 2, № 10. – P. e137.
116. Piana S., Carloni P., Rothlisberger U. Drug resistance in HIV-1 protease: Flexibility-assisted mechanism of compensatory mutations // *Protein Sci*. – 2002. – V. 11, № 10. – P. 2393-402.
117. Levin B. R., Perrot V., Walker N. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria // *Genetics*. – 2000. – V. 154, № 3. – P. 985-97.
118. Wang D., Salah El-Basyoni I., Stephen Baenziger P., Crossa J., Eskridge K. M., Dweikat I. Prediction of genetic values of quantitative traits with epistatic effects in plant breeding populations // *Heredity (Edinb)*. – 2012. – V. 109, № 5. – P. 313-9.
119. Dudley J., Johnson G. R. Epistatic models improve prediction of performance in corn. // *Crop Sci* – 2009. – V. 49. – P. 763-770.
120. Hu Z., Li Y., Song X., Han Y., Cai X., Xu S., Li W. Genomic value prediction for quantitative traits under the epistatic model // *BMC Genet*. – 2011. – V. 12. – P. 15.
121. Wan X., Yang C., Yang Q., Xue H., Fan X., Tang N. L. S., Yu W. BOOST: A Fast Approach to Detecting Gene-Gene Interactions in Genome-wide Case-Control Studies // *The American Journal of Human Genetics*. – 2010. – V. 87, № 3. – P. 325-340.
122. Lippert C., Listgarten J., Davidson R. I., Baxter S., Poon H., Kadie C. M., Heckerman D. An exhaustive epistatic SNP association analysis on expanded Wellcome Trust data // *Sci Rep*. – 2013. – V. 3. – P. 1099.
123. Zhu Z., Tong X., Zhu Z., Liang M., Cui W., Su K., Li M. D., Zhu J. Development of GMDR-GPU for Gene-Gene Interaction Analysis and Its Application to WTCCC GWAS Data for Type 2 Diabetes // *PLOS ONE*. – 2013. – V. 8, № 4. – P. e61943.
124. Lu Q., Wei C., Ye C., Li M., Elston R. C. A Likelihood Ratio-Based Mann-Whitney Approach Finds Novel Replicable Joint Gene Action for Type 2 Diabetes // *Genetic epidemiology*. – 2012. – V. 36, № 6. – P. 583-593.
125. Bell J. T., Timpson N. J., Rayner N. W., Zeggini E., Frayling T. M., Hattersley A. T., Morris A. P., M.I. M. Genome-Wide Association Scan Allowing for Epistasis in Type 2 Diabetes // *Annals of Human Genetics*. – 2011. – V. 75, № 1. – P. 10-19.
126. Génin E., Coustet B., Allanore Y., Ito I., Teruel M., Constantin A., Schaefferbeke T., Ruysse-Witrand A., Tohma S., Cantagrel A., Vittecoq O., Barnetche T., Le Loët X., Fardellone P., Furukawa H., Meyer O., Fernández-Gutiérrez B., Balsa A., González-Gay M. A., Chiochia G., Tsuchiya N., Martin J., Dieudé P. Epistatic Interaction between BANK1 and BLK in Rheumatoid Arthritis: Results from a Large Trans-Ethnic Meta-Analysis // *PLOS ONE*. – 2013. – V. 8, № 4. – P. e61044.
127. Kölsch H., Lehmann D. J., Ibrahim-Verbaas C. A., Combarros O., van Duijn C. M., Hammond N., Belbin O., Cortina-Borja M., Lehmann M. G., Aulchenko Y. S., Schuur M., Breteler M., Wilcock G. K., Brown K., Kehoe P. G., Barber R., Coto E., Alvarez V., Deloukas P., Mateo I., Maier W., Morgan K., Warden D. R., Smith A. D., Heun R. Interaction of insulin and PPAR- α genes in Alzheimer's disease: the Epistasis Project // *Journal of Neural Transmission*. – 2012. – V. 119, № 4. – P. 473-479.
128. Bullock J. M., Medway C., Cortina-Borja M., Turton J. C., Prince J. A., Ibrahim-Verbaas C. A., Schuur M., Breteler M. M., van Duijn C. M., Kehoe P. G., Barber R., Coto E., Alvarez V., Deloukas P., Hammond N., Combarros O., Mateo I., Warden D. R., Lehmann M. G., Belbin O., Brown K., Wilcock G. K., Heun R., Kölsch H., Smith A. D., Lehmann D. J., Morgan K. Discovery by the Epistasis Project of an epistatic interaction between the GSTM3 gene and the HHEX/IDE/KIF11 locus in the risk of Alzheimer's disease // *Neurobiol Aging*. – 2013. – V. 34, № 4. – P. 1309 e1-7.

129. Combarros O., van Duijn C. M., Hammond N., Belbin O., Arias-Vasquez A., Cortina-Borja M., Lehmann M. G., Aulchenko Y. S., Schuur M., Kolsch H., Heun R., Wilcock G. K., Brown K., Kehoe P. G., Harrison R., Coto E., Alvarez V., Deloukas P., Mateo I., Gwilliam R., Morgan K., Warden D. R., Smith A. D., Lehmann D. J. Replication by the Epistasis Project of the interaction between the genes for IL-6 and IL-10 in the risk of Alzheimer's disease // *J Neuroinflammation*. – 2009. – V. 6. – P. 22.
130. Rhinn H., Fujita R., Qiang L., Cheng R., Lee J. H., Abeliovich A. Integrative genomics identifies APOE epsilon4 effectors in Alzheimer's disease // *Nature*. – 2013. – V. 500, № 7460. – P. 45-50.
131. Weissman D. B., Desai M. M., Fisher D. S., Feldman M. W. The rate at which asexual populations cross fitness valleys // *Theor Popul Biol*. – 2009. – V. 75, № 4. – P. 286-300.
132. Gonzalez-Ortega E., Ballana E., Badia R., Clotet B., Este J. A. Compensatory mutations rescue the virus replicative capacity of VIRIP-resistant HIV-1 // *Antiviral Res*. – 2011. – V. 92, № 3. – P. 479-83.
133. Noviello C. M., Lopez C. S., Kukull B., McNett H., Still A., Eccles J., Sloan R., Barklis E. Second-site compensatory mutations of HIV-1 capsid mutations // *J Virol*. – 2011. – V. 85, № 10. – P. 4730-8.
134. Meher B. R., Wang Y. Interaction of I50V mutant and I50L/A71V double mutant HIV-protease with inhibitor TMC114 (darunavir): molecular dynamics simulation and binding free energy studies // *J Phys Chem B*. – 2012. – V. 116, № 6. – P. 1884-900.
135. Yu Y., Wang J., Shao Q., Shi J., Zhu W. Effects of drug-resistant mutations on the dynamic properties of HIV-1 protease and inhibition by Amprenavir and Darunavir // *Scientific Reports*. – 2015. – V. 5. – P. 10517.
136. Huang Y., Wuchty S., Przytycka T. M. eQTL Epistasis - Challenges and Computational Approaches // *Front Genet*. – 2013. – V. 4. – P. 51.
137. McKinney B. A., Pajewski N. M. Six Degrees of Epistasis: Statistical Network Models for GWAS // *Front Genet*. – 2011. – V. 2. – P. 109.
138. Pang X., Wang Z., Yap J. S., Wang J., Zhu J., Bo W., Lv Y., Xu F., Zhou T., Peng S., Shen D., Wu R. A statistical procedure to map high-order epistasis for complex traits // *Brief Bioinform*. – 2013. – V. 14, № 3. – P. 302-14.
139. Ritchie M. D. Using biological knowledge to uncover the mystery in the search for epistasis in genome-wide association studies // *Ann Hum Genet*. – 2011. – V. 75, № 1. – P. 172-82.
140. Zhang Y., Jiang B., Zhu J., Liu J. S. Bayesian models for detecting epistatic interactions from genetic data // *Ann Hum Genet*. – 2011. – V. 75, № 1. – P. 183-93.
141. Chen C. C., Schwender H., Keith J., Nunkesser R., Mengersen K., Macrossan P. Methods for identifying SNP interactions: a review on variations of Logic Regression, Random Forest and Bayesian logistic regression // *IEEE/ACM Trans Comput Biol Bioinform*. – 2011. – V. 8, № 6. – P. 1580-91.
142. García-Magariños M., López-de-Ullibarri I., Cao R., Salas A. Evaluating the Ability of Tree-Based Methods and Logistic Regression for the Detection of SNP-SNP Interaction // *Annals of Human Genetics*. – 2009. – V. 73, № 3. – P. 360-369.
143. Zhang Y., Liu J. S. Bayesian inference of epistatic interactions in case-control studies // *Nat Genet*. – 2007. – V. 39, № 9. – P. 1167-73.
144. Tang W., Wu X., Jiang R., Li Y. Epistatic module detection for case-control studies: a Bayesian model with a Gibbs sampling strategy // *PLoS Genet*. – 2009. – V. 5, № 5. – P. e1000464.
145. Ueki M., Cordell H. J. Improved statistics for genome-wide interaction analysis // *PLoS Genet*. – 2012. – V. 8, № 4. – P. e1002625.
146. Hoh J., Ott J. Mathematical multi-locus approaches to localizing complex human trait genes // *Nat Rev Genet*. – 2003. – V. 4, № 9. – P. 701-9.
147. Combarros O., Cortina-Borja M., Smith A. D., Lehmann D. J. Epistasis in sporadic Alzheimer's disease // *Neurobiol Aging*. – 2009. – V. 30, № 9. – P. 1333-49.
148. Kimura M. Attainment of Quasi Linkage Equilibrium When Gene Frequencies Are Changing by Natural Selection // *Genetics*. – 1965. – V. 52, № 5. – P. 875-90.
149. Neher R. A., Shraiman B. I. Competition between recombination and epistasis can cause a transition from allele to genotype selection // *Proc Natl Acad Sci U S A*. – 2009. – V. 106, № 16. – P. 6866-71.
150. Neher R. A., Shraiman B. I. Statistical genetics and evolution of quantitative traits // *Rev. Mod. Phys.* – 2011. – V. 83. – P. 1283.

151. Turelli M., Barton N. H. Will population bottlenecks and multilocus epistasis increase additive genetic variance? // *Evolution*. – 2006. – V. 60, № 9. – P. 1763-76.
152. Cocco S., Feinauer C., Figliuzzi M., Monasson R., Weigt M. Inverse statistical physics of protein sequences: a key issues review // *Rep Prog Phys*. – 2018. – V. 81, № 3. – P. 032601.
153. Weigt M., White R. A., Szurmant H., Hoch J. A., Hwa T. Identification of direct residue contacts in protein-protein interaction by message passing // *Proc Natl Acad Sci U S A*. – 2009. – V. 106, № 1. – P. 67-72.
154. Frost S. D., Nijhuis M., Schuurman R., Boucher C. A., Brown A. J. Evolution of lamivudine resistance in human immunodeficiency virus type 1-infected individuals: the relative roles of drift and selection // *J Virol*. – 2000. – V. 74, № 14. – P. 6262-8.
155. Illingworth C. J., Mustonen V. Components of selection in the evolution of the influenza virus: linkage effects beat inherent selection // *PLoS Pathog*. – 2012. – V. 8, № 12. – P. e1003091.
156. Keightley P. D., Eyre-Walker A. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies // *Genetics*. – 2007. – V. 177, № 4. – P. 2251-61.
157. Gillespie J. H. A Randomized Sas Cff Model of Natural-Selection in a Random Environment // *Theoretical Population Biology*. – 1982. – V. 21, № 2. – P. 219-237.
158. Orr H. A. The distribution of fitness effects among beneficial mutations // *Genetics*. – 2003. – V. 163, № 4. – P. 1519-26.
159. Joyce P., Rokyta D. R., Beisel C. J., Orr H. A. A general extreme value theory model for the adaptation of DNA sequences under strong selection and weak mutation // *Genetics*. – 2008. – V. 180, № 3. – P. 1627-43.
160. Barre-Sinoussi F., Chermann J. C., Rey F., Nugeyre M. T., Chamaret S., Gruest J., Dauguet C., Axler-Blin C., Vezinet-Brun F., Rouzioux C., Rozenbaum W., Montagnier L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS) // *Science*. – 1983. – V. 220, № 4599. – P. 868-71.
161. Gallo R. C., Salahuddin S. Z., Popovic M., Shearer G. M., Kaplan M., Haynes B. F., Palker T. J., Redfield R., Oleske J., Safai B., et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS // *Science*. – 1984. – V. 224, № 4648. – P. 500-3.
162. Levy J. A., Hoffman A. D., Kramer S. M., Landis J. A., Shimabukuro J. M., Oshiro L. S. Isolation of lymphocytotropic retroviruses from San Francisco patients with AIDS // *Science*. – 1984. – V. 225, № 4664. – P. 840-2.
163. Quagliariello V. The Acquired Immunodeficiency Syndrome: current status // *Yale J Biol Med*. – 1982. – V. 55, № 5-6. – P. 443-52.
164. Gottlieb M. S., Schroff R., Schanker H. M., Weisman J. D., Fan P. T., Wolf R. A., Saxon A. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency // *N Engl J Med*. – 1981. – V. 305, № 24. – P. 1425-31.
165. Masur H., Michelis M. A., Greene J. B., Onorato I., Stouwe R. A., Holzman R. S., Wormser G., Brettman L., Lange M., Murray H. W., Cunningham-Rundles S. An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction // *N Engl J Med*. – 1981. – V. 305, № 24. – P. 1431-8.
166. Small C. B., Klein R. S., Friedland G. H., Moll B., Emeson E. E., Spigland I. Community-acquired opportunistic infections and defective cellular immunity in heterosexual drug abusers and homosexual men // *Am J Med*. – 1983. – V. 74, № 3. – P. 433-41.
167. Vieira J., Frank E., Spira T. J., Landesman S. H. Acquired immune deficiency in Haitians: opportunistic infections in previously healthy Haitian immigrants // *N Engl J Med*. – 1983. – V. 308, № 3. – P. 125-9.
168. Klatzmann D., Barre-Sinoussi F., Nugeyre M. T., Danquet C., Vilmer E., Griscelli C., Brun-Veziret F., Rouzioux C., Gluckman J. C., Chermann J. C., et al. Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes // *Science*. – 1984. – V. 225, № 4657. – P. 59-63.

169. Masur H., Ognibene F. P., Yarchoan R., Shelhamer J. H., Baird B. F., Travis W., Suffredini A. F., Deyton L., Kovacs J. A., Falloon J., et al. CD4 counts as predictors of opportunistic pneumonias in human immunodeficiency virus (HIV) infection // *Ann Intern Med.* – 1989. – V. 111, № 3. – P. 223-31.
170. Ho D. D., Neumann A. U., Perelson A. S., Chen W., Leonard J. M., Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection // *Nature.* – 1995. – V. 373, № 6510. – P. 123-6.
171. Wei X., Ghosh S. K., Taylor M. E., Johnson V. A., Emimi E. A., Deutsch P., Lifson J. D., Bonhoeffer S., Nowak M. A., Hahn B. H., et al. Viral dynamics in human immunodeficiency virus type 1 infection // *Nature.* – 1995. – V. 373, № 6510. – P. 117-22.
172. Reimann K. A., Li J. T., Veazey R., Halloran M., Park I. W., Karlsson G. B., Sodroski J., Letvin N. L. A chimeric simian/human immunodeficiency virus expressing a primary patient human immunodeficiency virus type 1 isolate env causes an AIDS-like disease after in vivo passage in rhesus monkeys // *J Virol.* – 1996. – V. 70, № 10. – P. 6922-8.
173. Nishimura Y., Brown C. R., Mattapallil J. J., Igarashi T., Buckler-White A., Lafont B. A., Hirsch V. M., Roederer M., Martin M. A. Resting naive CD4+ T cells are massively infected and eliminated by X4-tropic simian-human immunodeficiency viruses in macaques // *Proc Natl Acad Sci U S A.* – 2005. – V. 102, № 22. – P. 8000-5.
174. Letvin N. L., Mascola J. R., Sun Y., Gorgone D. A., Buzby A. P., Xu L., Yang Z. Y., Chakrabarti B., Rao S. S., Schmitz J. E., Montefiori D. C., Barker B. R., Bookstein F. L., Nabel G. J. Preserved CD4+ central memory T cells and survival in vaccinated SIV-challenged monkeys // *Science.* – 2006. – V. 312, № 5779. – P. 1530-3.
175. Sun Y., Schmitz J. E., Buzby A. P., Barker B. R., Rao S. S., Xu L., Yang Z. Y., Mascola J. R., Nabel G. J., Letvin N. L. Virus-specific cellular immune correlates of survival in vaccinated monkeys after simian immunodeficiency virus challenge // *J Virol.* – 2006. – V. 80, № 22. – P. 10950-6.
176. Chun T. W., Carruth L., Finzi D., Shen X., DiGiuseppe J. A., Taylor H., Hermankova M., Chadwick K., Margolick J., Quinn T. C., Kuo Y. H., Brookmeyer R., Zeiger M. A., Barditch-Crovo P., Siliciano R. F. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection // *Nature.* – 1997. – V. 387, № 6629. – P. 183-8.
177. Grivel J. C., Penn M. L., Eckstein D. A., Schramm B., Speck R. F., Abbey N. W., Herndier B., Margolis L., Goldsmith M. A. Human immunodeficiency virus type 1 coreceptor preferences determine target T-cell depletion and cellular tropism in human lymphoid tissue // *J Virol.* – 2000. – V. 74, № 11. – P. 5347-51.
178. Brenchley J. M., Schacker T. W., Ruff L. E., Price D. A., Taylor J. H., Beilman G. J., Nguyen P. L., Khoruts A., Larson M., Haase A. T., Douek D. C. CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract // *J Exp Med.* – 2004. – V. 200, № 6. – P. 749-59.
179. Picker L. J., Hagen S. I., Lum R., Reed-Inderbitzin E. F., Daly L. M., Sylwester A. W., Walker J. M., Siess D. C., Piatak M., Jr., Wang C., Allison D. B., Maino V. C., Lifson J. D., Kodama T., Axthelm M. K. Insufficient production and tissue delivery of CD4+ memory T cells in rapidly progressive simian immunodeficiency virus infection // *J Exp Med.* – 2004. – V. 200, № 10. – P. 1299-314.
180. Mohri H., Bonhoeffer S., Monard S., Perelson A. S., Ho D. D. Rapid turnover of T lymphocytes in SIV-infected rhesus macaques // *Science.* – 1998. – V. 279, № 5354. – P. 1223-7.
181. Rouzine I., Coffin J. M. Interplay between experiment and theory in development of a working model for HIV-1 population dynamics // *Origin and evolution of viruses / Domingo E. и др.* – London, United Kingdom: Academic Press Ltd., 1999.
182. Ribeiro R. M., Mohri H., Ho D. D., Perelson A. S. In vivo dynamics of T cell activation, proliferation, and death in HIV-1 infection: why are CD4+ but not CD8+ T cells depleted? // *Proc Natl Acad Sci U S A.* – 2002. – V. 99, № 24. – P. 15572-7.
183. Ribeiro R. M., Mohri H., Ho D. D., Perelson A. S. Modeling deuterated glucose labeling of T-lymphocytes // *Bull. Math. Biol.* – 2002. – V. 64, № 2. – P. 385-405.
184. Rouzine I. M., Coffin J. M. T cell turnover in SIV infection [comment] // *Science* – 1999. – V. 284. – P. 555b.

185. Grossman Z., Herberman R., Dimitrov D. S. T cell turnover in SIV infection [comment] // *Science* – 1999. – V. 284. – P. 555a.
186. Giorgi J. V., Hultin L. E., McKeating J. A., Johnson T. D., Owens B., Jacobson L. P., Shih R., Lewis J., Wiley D. J., Phair J. P., Wolinsky S. M., Detels R. Shorter survival in advanced human immunodeficiency virus type 1 infection is more closely associated with T lymphocyte activation than with plasma virus burden or virus chemokine coreceptor usage // *J Infect Dis.* – 1999. – V. 179, № 4. – P. 859-70.
187. Hazenberg M. D., Otto S. A., van Benthem B. H., Roos M. T., Coutinho R. A., Lange J. M., Hamann D., Prins M., Miedema F. Persistent immune activation in HIV-1 infection is associated with progression to AIDS // *AIDS.* – 2003. – V. 17, № 13. – P. 1881-8.
188. Deeks S. G., Kitchen C. M., Liu L., Guo H., Gascon R., Narvaez A. B., Hunt P., Martin J. N., Kahn J. O., Levy J., McGrath M. S., Hecht F. M. Immune activation set point during early HIV infection predicts subsequent CD4+ T-cell changes independent of viral load // *Blood.* – 2004. – V. 104, № 4. – P. 942-7.
189. Rodriguez B., Sethi A. K., Cheruvu V. K., Mackay W., Bosch R. J., Kitahata M., Boswell S. L., Mathews W. C., Bangsberg D. R., Martin J., Whalen C. C., Sieg S., Yadavalli S., Deeks S. G., Lederman M. M. Predictive value of plasma HIV RNA level on rate of CD4 T-cell decline in untreated HIV infection // *JAMA.* – 2006. – V. 296, № 12. – P. 1498-506.
190. De Boer R. J., Perelson A. S. Target cell limited and immune control models of HIV infection: a comparison // *J Theor Biol.* – 1998. – V. 190, № 3. – P. 201-14.
191. *Virus Dynamics: Mathematical Principles of Immunology and Virology.* / Nowak M. A., May R. – New York, NY, USA; Oxford, UK, 2000: Oxford University Press, 2000.
192. Chahroudi A., Bosinger S. E., Vanderford T. H., Paiardini M., Silvestri G. Natural SIV hosts: showing AIDS the door // *Science.* – 2012. – V. 335, № 6073. – P. 1188-93.
193. Goldstein S., Brown C. R., Ourmanov I., Pandrea I., Buckler-White A., Erb C., Nandi J. S., Foster G. J., Autissier P., Schmitz J. E., Hirsch V. M. Comparison of simian immunodeficiency virus SIVagmVer replication and CD4+ T-cell dynamics in vervet and sabaeus African green monkeys // *J Virol.* – 2006. – V. 80, № 10. – P. 4868-77.
194. Gordon S. N., Dunham R. M., Engram J. C., Estes J., Wang Z., Klatt N. R., Paiardini M., Pandrea I. V., Apetrei C., Sodora D. L., Lee H. Y., Haase A. T., Miller M. D., Kaur A., Staprans S. I., Perelson A. S., Feinberg M. B., Silvestri G. Short-lived infected cells support virus replication in sooty mangabeys naturally infected with simian immunodeficiency virus: implications for AIDS pathogenesis // *J Virol.* – 2008. – V. 82, № 7. – P. 3725-35.
195. Pandrea I., Ribeiro R. M., Gautam R., Gaufin T., Pattison M., Barnes M., Monjure C., Stoulig C., Dufour J., Cyprian W., Silvestri G., Miller M. D., Perelson A. S., Apetrei C. Simian immunodeficiency virus SIVagm dynamics in African green monkeys // *J Virol.* – 2008. – V. 82, № 7. – P. 3713-24.
196. Silvestri G., Sodora D. L., Koup R. A., Paiardini M., O'Neil S. P., McClure H. M., Staprans S. I., Feinberg M. B. Nonpathogenic SIV infection of sooty mangabeys is characterized by limited bystander immunopathology despite chronic high-level viremia // *Immunity.* – 2003. – V. 18, № 3. – P. 441-52.
197. Norley S., Beer B., Holzammer S., zur Megede J., Kurth R. Why are the natural hosts of SIV resistant to AIDS? // *Immunol Lett.* – 1999. – V. 66, № 1-3. – P. 47-52.
198. Milush J. M., Mir K. D., Sundaravaradan V., Gordon S. N., Engram J., Cano C. A., Reeves J. D., Anton E., O'Neill E., Butler E., Hancock K., Cole K. S., Brenchley J. M., Else J. G., Silvestri G., Sodora D. L. Lack of clinical AIDS in SIV-infected sooty mangabeys with significant CD4+ T cell loss is associated with double-negative T cells // *J Clin Invest.* – 2011. – V. 121, № 3. – P. 1102-10.
199. Dapp M. J., Kober K. M., Chen L., Westfall D. H., Wong K., Zhao H., Hall B. M., Deng W., Sibley T., Ghorai S., Kim K., Chen N., McHugh S., Au L., Cohen M., Anastos K., Mullins J. I. Patterns and rates of viral evolution in HIV-1 subtype B infected females and males // *PLoS One.* – 2017. – V. 12, № 10. – P. e0182443.
200. Patino-Galindo J. A., Gonzalez-Candelas F. The substitution rate of HIV-1 subtypes: a genomic approach // *Virus Evol.* – 2017. – V. 3, № 2. – P. vex029.

201. Yewdell J. W. Antigenic drift: Understanding COVID-19 // *Immunity*. – 2021. – V. 54, № 12. – P. 2681-2687.
202. Rouzine I. M. An evolutionary model of progression to AIDS // *Microorganisms*. – 2020. – V. 8, № 11.
203. Chun T. W., Stuyver L., Mizell S. B., Ehler L. A., Mican J. A., Baseler M., Lloyd A. L., Nowak M. A., Fauci A. S. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy // *Proc Natl Acad Sci U S A*. – 1997. – V. 94, № 24. – P. 13193-7.
204. Finzi D., Hermankova M., Pierson T., Carruth L. M., Buck C., Chaisson R. E., Quinn T. C., Chadwick K., Margolick J., Brookmeyer R., Gallant J., Markowitz M., Ho D. D., Richman D. D., Siliciano R. F. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy // *Science*. – 1997. – V. 278. – P. 1295-1300.
205. Pierson T., McArthur J., Siliciano R. F. Reservoirs for HIV-1: mechanisms for viral persistence in the presence of antiviral immune responses and antiretroviral therapy // *Annual review of immunology*. – 2000. – V. 18. – P. 665-708.
206. Richman D. D., Margolis D. M., Delaney M., Greene W. C., Hazuda D., Pomerantz R. J. The challenge of finding a cure for HIV infection // *Science*. – 2009. – V. 323, № 5919. – P. 1304-7.
207. Coffin J., Swanstrom R. HIV pathogenesis: dynamics and genetics of viral populations and infected cells // *Cold Spring Harb Perspect Med*. – 2013. – V. 3, № 1. – P. a012526.
208. Han Y., Wind-Rotolo M., Yang H. C., Siliciano J. D., Siliciano R. F. Experimental approaches to the study of HIV-1 latency // *Nature reviews. Microbiology*. – 2007. – V. 5, № 2. – P. 95-106.
209. Eisele E., Siliciano R. F. Redefining the viral reservoirs that prevent HIV-1 eradication // *Immunity*. – 2012. – V. 37, № 3. – P. 377-88.
210. Whitney J. B., Hill A. L., Sanisetty S., Penaloza-MacMaster P., Liu J., Shetty M., Parenteau L., Cabral C., Shields J., Blackmore S., Smith J. Y., Brinkman A. L., Peter L. E., Mathew S. I., Smith K. M., Borducchi E. N., Rosenbloom D. I., Lewis M. G., Hattersley J., Li B., Hesselgesser J., Geleziunas R., Robb M. L., Kim J. H., Michael N. L., Barouch D. H. Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys // *Nature*. – 2014. – V. 512, № 7512. – P. 74-7.
211. Kuroda M. J., Schmitz J. E., Charini W. A., Nickerson C. E., Lifton M. A., Lord C. I., Forman M. A., Letvin N. L. Emergence of CTL coincides with clearance of virus during primary simian immunodeficiency virus infection in rhesus monkeys // *J Immunol*. – 1999. – V. 162, № 9. – P. 5127-33.
212. Schmitz J. E., Kuroda M. J., Santra S., Sasseville V. G., Simon M. A., Lifton M. A., Racz P., Tenner-Racz K., Dalesandro M., Scallon B. J., Ghayeb J., Forman M. A., Montefiori D. C., Rieber E. P., Letvin N. L., Reimann K. A. Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes // *Science*. – 1999. – V. 283, № 5403. – P. 857-60.
213. Dahabieh M. S., Ooms M., Simon V., Sadowski I. A double-fluorescent HIV-1 reporter shows that the majority of integrated HIV-1 is latent shortly after infection // *Journal of virology*. – 2013.10.1128/JVI.03478-12.
214. Calvanese V., Chavez L., Laurent T., Ding S., Verdin E. Dual-color HIV reporters trace a population of latently infected cells and enable their purification // *Virology*. – 2013. – V. 446, № 1-2. – P. 283-92.
215. Weinberger L. S., Dar R. D., Simpson M. L. Transient-mediated fate determination in a transcriptional circuit of HIV // *Nat Genet*. – 2008. – V. 40, № 4. – P. 466-70.
216. Burnett J. C., Miller-Jensen K., Shah P. S., Arkin A. P., Schaffer D. V. Control of stochastic gene expression by host factors at the HIV promoter // *PLoS pathogens*. – 2009. – V. 5, № 1. – P. e1000260.
217. Weinberger L. S., Burnett J. C., Toettcher J. E., Arkin A. P., Schaffer D. V. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity // *Cell*. – 2005. – V. 122, № 2. – P. 169-82.
218. Fraser H. B., Hirsh A. E., Giaever G., Kumm J., Eisen M. B. Noise minimization in eukaryotic gene expression // *PLoS Biol*. – 2004. – V. 2, № 6. – P. e137.
219. Batada N. N., Hurst L. D. Evolution of chromosome organization driven by selection for reduced gene expression noise // *Nat Genet*. – 2007. – V. 39, № 8. – P. 945-9.

220. Cohen D. Optimizing reproduction in a randomly varying environment // *J Theor Biol.* – 1966. – V. 12, № 1. – P. 119-29.
221. Arkin A., Ross J., McAdams H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells // *Genetics.* – 1998. – V. 149, № 4. – P. 1633-48.
222. Balaban N. Q. Persistence: mechanisms for triggering and enhancing phenotypic variability // *Curr Opin Genet Dev.* – 2011. – V. 21, № 6. – P. 768-75.
223. Rouzine I. M., Sergeev R. A., Glushtsov A. I. Two types of cytotoxic lymphocyte regulation explain kinetics of immune response to human immunodeficiency virus // *Proc Natl Acad Sci U S A.* – 2006. – V. 103, № 3. – P. 666-71.
224. Rouzine I. M. A role for CD4 + helper cells in HIV control and progression // *AIDS.* – 2022. – V. 36, № 11. – P. 1501-1510.
225. Fraser C., Hollingsworth T. D., Chapman R., de Wolf F., Hanage W. P. Variation in HIV-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis // *Proc Natl Acad Sci U S A.* – 2007. – V. 104, № 44. – P. 17441-6.
226. Wawer M. J., Gray R. H., Sewankambo N. K., Serwadda D., Li X., Laeyendecker O., Kiwanuka N., Kigozi G., Kiddugavu M., Lutalo T., Nalugoda F., Wabwire-Mangen F., Meehan M. P., Quinn T. C. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda // *J Infect Dis.* – 2005. – V. 191, № 9. – P. 1403-9.
227. Gray R. H., Wawer M. J., Brookmeyer R., Sewankambo N. K., Serwadda D., Wabwire-Mangen F., Lutalo T., Li X., vanCott T., Quinn T. C., Rakai Project T. Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda // *Lancet.* – 2001. – V. 357, № 9263. – P. 1149-53.
228. Keele B. F., Giorgi E. E., Salazar-Gonzalez J. F., Decker J. M., Pham K. T., Salazar M. G., Sun C., Grayson T., Wang S., Li H., Wei X., Jiang C., Kirchherr J. L., Gao F., Anderson J. A., Ping L. H., Swanstrom R., Tomaras G. D., Blattner W. A., Goepfert P. A., Kilby J. M., Saag M. S., Delwart E. L., Busch M. P., Cohen M. S., Montefiori D. C., Haynes B. F., Gaschen B., Athreya G. S., Lee H. Y., Wood N., Seoighe C., Perelson A. S., Bhattacharya T., Korber B. T., Hahn B. H., Shaw G. M. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection // *Proc Natl Acad Sci U S A.* – 2008. – V. 105, № 21. – P. 7552-7.
229. Kearney M., Maldarelli F., Shao W., Margolick J. B., Daar E. S., Mellors J. W., Rao V., Coffin J. M., Palmer S. Human immunodeficiency virus type 1 population genetics and adaptation in newly infected individuals // *J Virol.* – 2009. – V. 83, № 6. – P. 2715-27.
230. Zhang Z., Schuler T., Zupancic M., Wietgreffe S., Staskus K. A., Reimann K. A., Reinhart T. A., Rogan M., Cavert W., Miller C. J., Veazey R. S., Notermans D., Little S., Danner S. A., Richman D. D., Havlir D., Wong J., Jordan H. L., Schacker T. W., Racz P., Tenner-Racz K., Letvin N. L., Wolinsky S., Haase A. T. Sexual transmission and propagation of SIV and HIV in resting and activated CD4+ T cells // *Science.* – 1999. – V. 286, № 5443. – P. 1353-7.
231. Haase A. T. Early events in sexual transmission of HIV and SIV and opportunities for interventions // *Annual review of medicine.* – 2011. – V. 62. – P. 127-39.
232. Miller C. J., Li Q., Abel K., Kim E. Y., Ma Z. M., Wietgreffe S., La Franco-Scheuch L., Compton L., Duan L., Shore M. D., Zupancic M., Busch M., Carlis J., Wolinsky S., Haase A. T. Propagation and dissemination of infection after vaginal transmission of simian immunodeficiency virus // *J Virol.* – 2005. – V. 79, № 14. – P. 9217-27.
233. Santiago M. L., Range F., Keele B. F., Li Y., Bailes E., Bibollet-Ruche F., Fruteau C., Noe R., Peeters M., Brookfield J. F., Shaw G. M., Sharp P. M., Hahn B. H. Simian immunodeficiency virus infection in free-ranging sooty mangabeys (*Cercocebus atys atys*) from the Tai Forest, Cote d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2 // *J Virol.* – 2005. – V. 79, № 19. – P. 12515-27.
234. Markowitz M., Louie M., Hurley A., Sun E., Di Mascio M., Perelson A. S., Ho D. D. A novel antiviral intervention results in more accurate assessment of human immunodeficiency virus type 1 replication dynamics and T-cell decay in vivo // *J Virol.* – 2003. – V. 77, № 8. – P. 5037-8.

235. Elena S. F., Sanjuan R. RNA viruses as complex adaptive systems // *Biosystems*. – 2005. – V. 81, № 1. – P. 31-41.
236. Holland J., Spindler K., Horodyski F., Grabau E., Nichol S., VandePol S. Rapid evolution of RNA genomes // *Science*. – 1982. – V. 215, № 4540. – P. 1577-85.
237. Steinhauer D. A., Holland J. J. Rapid evolution of RNA viruses // *Annu Rev Microbiol*. – 1987. – V. 41. – P. 409-33.
238. Domingo E., Holland J. J. RNA virus mutations and fitness for survival // *Annu Rev Microbiol*. – 1997. – V. 51. – P. 151-78.
239. Hartfield M., Otto S. P., Keightley P. D. The role of advantageous mutations in enhancing the evolution of a recombination modifier // *Genetics*. – 2010. – V. 184, № 4. – P. 1153-64.
240. Elena S. F., Sanjuan R. Adaptive value of high mutation rates of RNA viruses: separating causes from consequences // *J Virol*. – 2005. – V. 79, № 18. – P. 11555-8.
241. Domingo E., Escarmis C., Sevilla N., Moya A., Elena S. F., Quer J., Novella I. S., Holland J. J. Basic concepts in RNA virus evolution // *FASEB J*. – 1996. – V. 10, № 8. – P. 859-64.
242. Holmes E. C. RNA virus genomics: a world of possibilities // *J Clin Invest*. – 2009. – V. 119, № 9. – P. 2488-95.
243. Cooper T. F. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli* // *PLoS Biol*. – 2007. – V. 5, № 9. – P. e225.
244. Worobey M., Holmes E. C. Evolutionary aspects of recombination in RNA viruses // *J Gen Virol*. – 1999. – V. 80 (Pt 10). – P. 2535-2543.
245. Chao L. Fitness of RNA virus decreased by Muller's ratchet // *Nature*. – 1990. – V. 348, № 6300. – P. 454-5.
246. Duarte E., Clarke D., Moya A., Domingo E., Holland J. Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet // *Proc Natl Acad Sci U S A*. – 1992. – V. 89, № 13. – P. 6015-9.
247. Roze D., Barton N. H. The Hill-Robertson effect and the evolution of recombination // *Genetics*. – 2006. – V. 173, № 3. – P. 1793-811.
248. Zeyl C., Bell G. The advantage of sex in evolving yeast populations // *Nature*. – 1997. – V. 388, № 6641. – P. 465-8.
249. Bonhoeffer S., Chappey C., Parkin N. T., Whitcomb J. M., Petropoulos C. J. Evidence for positive epistasis in HIV-1 // *Science*. – 2004. – V. 306, № 5701. – P. 1547-50.
250. Burch C. L., Turner P. E., Hanley K. A. Patterns of epistasis in RNA viruses: a review of the evidence from vaccine design // *J Evol Biol*. – 2003. – V. 16, № 6. – P. 1223-35.
251. Racaniello V. R., Baltimore D. Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome // *Proc Natl Acad Sci U S A*. – 1981. – V. 78, № 8. – P. 4887-91.
252. Agol V. I. Molecular mechanisms of poliovirus variation and evolution // *Curr Top Microbiol Immunol*. – 2006. – V. 299. – P. 211-59.
253. Kirkegaard K., Baltimore D. The mechanism of RNA recombination in poliovirus // *Cell*. – 1986. – V. 47, № 3. – P. 433-43.
254. Lowry K., Woodman A., Cook J., Evans D. J. Recombination in enteroviruses is a biphasic replicative process involving the generation of greater-than genome length 'imprecise' intermediates // *PLoS Pathog*. – 2014. – V. 10, № 6. – P. e1004191.
255. Runckel C., Westesson O., Andino R., DeRisi J. L. Identification and manipulation of the molecular determinants influencing poliovirus recombination // *PLoS Pathog*. – 2013. – V. 9, № 2. – P. e1003164.
256. Sztuba-Solinska J., Urbanowicz A., Figlerowicz M., Bujarski J. J. RNA-RNA recombination in plant virus replication and evolution // *Annu Rev Phytopathol*. – 2011. – V. 49. – P. 415-43.
257. Brunet E., Derrida B., Mueller A. H., Munier S. Effect of selection on ancestry: An exactly soluble case and its phenomenological generalization // *Physical Review E*. – 2007. – V. 76. – P. 041104-1.
258. Gerrish P. J., Colato A., Sniegowski P. D. Genomic mutation rates that neutralize adaptive evolution and natural selection // *J R Soc Interface*. – 2013. – V. 10, № 85. – P. 20130329.

259. Smith D. J., Lapedes A. S., de Jong J. C., Bestebroer T. M., Rimmelzwaan G. F., Osterhaus A. D., Fouchier R. A. Mapping the antigenic and genetic evolution of influenza virus // *Science*. – 2004. – V. 305, № 5682. – P. 371-6.
260. Rambaut A., Pybus O. G., Nelson M. I., Viboud C., Taubenberger J. K., Holmes E. C. The genomic and epidemiological dynamics of human influenza A virus // *Nature*. – 2008. – V. 453, № 7195. – P. 615-9.
261. Russell C. A., Jones T. C., Barr I. G., Cox N. J., Garten R. J., Gregory V., Gust I. D., Hampson A. W., Hay A. J., Hurt A. C., de Jong J. C., Kelso A., Klimov A. I., Kageyama T., Komadina N., Lapedes A. S., Lin Y. P., Mosterin A., Obuchi M., Odagiri T., Osterhaus A. D., Rimmelzwaan G. F., Shaw M. W., Skepner E., Stohr K., Tashiro M., Fouchier R. A., Smith D. J. Influenza vaccine strain selection and recent studies on the global migration of seasonal influenza viruses // *Vaccine*. – 2008. – V. 26 Suppl 4. – P. D31-4.
262. Bedford T., Riley S., Barr I. G., Broor S., Chadha M., Cox N. J., Daniels R. S., Gunasekaran C. P., Hurt A. C., Kelso A., Klimov A., Lewis N. S., Li X., McCauley J. W., Odagiri T., Potdar V., Rambaut A., Shu Y., Skepner E., Smith D. J., Suchard M. A., Tashiro M., Wang D., Xu X., Lemey P., Russell C. A. Global circulation patterns of seasonal influenza viruses vary with antigenic drift // *Nature*. – 2015. – V. 523, № 7559. – P. 217-20.
263. Koel B. F., Burke D. F., Bestebroer T. M., van der Vliet S., Zondag G. C., Vervaet G., Skepner E., Lewis N. S., Spronken M. I., Russell C. A., Eropkin M. Y., Hurt A. C., Barr I. G., de Jong J. C., Rimmelzwaan G. F., Osterhaus A. D., Fouchier R. A., Smith D. J. Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution // *Science*. – 2013. – V. 342, № 6161. – P. 976-9.
264. Fonville J. M., Wilks S. H., James S. L., Fox A., Ventresca M., Aban M., Xue L., Jones T. C., Le N. M. H., Pham Q. T., Tran N. D., Wong Y., Mosterin A., Katzelnick L. C., Labonte D., Le T. T., van der Net G., Skepner E., Russell C. A., Kaplan T. D., Rimmelzwaan G. F., Masurel N., de Jong J. C., Palache A., Beyer W. E. P., Le Q. M., Nguyen T. H., Wertheim H. F. L., Hurt A. C., Osterhaus A., Barr I. G., Fouchier R. A. M., Horby P. W., Smith D. J. Antibody landscapes after influenza virus infection or vaccination // *Science*. – 2014. – V. 346, № 6212. – P. 996-1000.
265. Neher R. A., Bedford T., Daniels R. S., Russell C. A., Shraiman B. I. Prediction, dynamics, and visualization of antigenic phenotypes of seasonal influenza viruses // *Proc Natl Acad Sci U S A*. – 2016. – V. 113, № 12. – P. E1701-9.
266. Lin J., Andreasen V., Casagrandi R., Levin S. A. Traveling waves in a model of influenza A drift // *J Theor Biol*. – 2003. – V. 222, № 4. – P. 437-45.
267. Bedford T., Rambaut A., Pascual M. Canalization of the evolutionary trajectory of the human influenza virus // *BMC Biol*. – 2012. – V. 10. – P. 38.
268. Strelkova N., Lassig M. Clonal interference in the evolution of influenza // *Genetics*. – 2012. – V. 192, № 2. – P. 671-82.
269. Bedford T., Suchard M. A., Lemey P., Dudas G., Gregory V., Hay A. J., McCauley J. W., Russell C. A., Smith D. J., Rambaut A. Integrating influenza antigenic dynamics with molecular evolution // *Elife*. – 2014. – V. 3. – P. e01914.
270. Luksza M., Lassig M. A predictive fitness model for influenza // *Nature*. – 2014. – V. 507, № 7490. – P. 57-61.
271. Yan L., Neher R. A., Shraiman B. I. Phylodynamic theory of persistence, extinction and speciation of rapidly adapting pathogens // *Elife*. – 2019. – V. 8. – P. e44205.
272. Marchi J., Lassig M., Walczak A. M., Mora T. Antigenic waves of virus-immune coevolution // *Proc Natl Acad Sci U S A*. – 2021. – V. 118, № 27. – P. e2103398118.
273. Potter S. J., Lacabaratz C., Lambotte O., Perez-Patrigéon S., Vingert B., Sinet M., Colle J. H., Urrutia A., Scott-Algara D., Boufassa F., Delfraissy J. F., Theze J., Venet A., Chakrabarti L. A. Preserved central memory and activated effector memory CD4⁺ T-cell subsets in human immunodeficiency virus controllers: an ANRS EP36 study // *J Virol*. – 2007. – V. 81, № 24. – P. 13904-15.

274. Vingert B., Perez-Patrigeon S., Jeannin P., Lambotte O., Boufassa F., Lemaitre F., Kwok W. W., Theodorou I., Delfraissy J. F., Theze J., Chakrabarti L. A., Group A. E. H. C. S. HIV controller CD4+ T cells respond to minimal amounts of Gag antigen due to high TCR avidity // *PLoS Pathog.* – 2010. – V. 6, № 2. – P. e1000780.
275. Leslie A., Kavanagh D., Honeyborne I., Pfafferott K., Edwards C., Pillay T., Hilton L., Thobakgale C., Ramduth D., Draenert R., Le Gall S., Luzzi G., Edwards A., Brander C., Sewell A. K., Moore S., Mullins J., Moore C., Mallal S., Bhardwaj N., Yusim K., Phillips R., Klenerman P., Korber B., Kiepiela P., Walker B., Goulder P. Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA // *J Exp Med.* – 2005. – V. 201, № 6. – P. 891-902.
276. Friedrich T. C., Dodds E. J., Yant L. J., Vojnov L., Rudersdorf R., Cullen C., Evans D. T., Desrosiers R. C., Mothe B. R., Sidney J., Sette A., Kunstman K., Wolinsky S., Piatak M., Lifson J., Hughes A. L., Wilson N., O'Connor D. H., Watkins D. I. Reversion of CTL escape-variant immunodeficiency viruses in vivo // *Nat Med.* – 2004. – V. 10, № 3. – P. 275-81.
277. Troyer R. M., McNevin J., Liu Y., Zhang S. C., Krizan R. W., Abraha A., Tebit D. M., Zhao H., Avila S., Lobritz M. A., McElrath M. J., Le Gall S., Mullins J. I., Arts E. J. Variable fitness impact of HIV-1 escape mutations to cytotoxic T lymphocyte (CTL) response // *PLoS Pathog.* – 2009. – V. 5, № 4. – P. e1000365.
278. Turnbull E. L., Wong M., Wang S., Wei X., Jones N. A., Conrod K. E., Aldam D., Turner J., Pellegrino P., Keele B. F., Williams I., Shaw G. M., Borrow P. Kinetics of expansion of epitope-specific T cell responses during primary HIV-1 infection // *J Immunol.* – 2009. – V. 182, № 11. – P. 7131-45.
279. Karlsson A. C., Iversen A. K., Chapman J. M., de Oliveira T., Spotts G., McMichael A. J., Davenport M. P., Hecht F. M., Nixon D. F. Sequential broadening of CTL responses in early HIV-1 infection is associated with viral escape // *PLoS One.* – 2007. – V. 2, № 2. – P. e225.
280. Liu M. K., Hawkins N., Ritchie A. J., Ganusov V. V., Whale V., Brackenridge S., Li H., Pavlicek J. W., Cai F., Rose-Abrahams M., Treurnicht F., Hraber P., Riou C., Gray C., Ferrari G., Tanner R., Ping L. H., Anderson J. A., Swanstrom R., B. C. C., Cohen M., Karim S. S., Haynes B., Borrow P., Perelson A. S., Shaw G. M., Hahn B. H., Williamson C., Korber B. T., Gao F., Self S., McMichael A., Goonetilleke N. Vertical T cell immunodominance and epitope entropy determine HIV-1 escape // *J Clin Invest.* – 2013. – V. 123, № 1. – P. 380-93.
281. Jin X., Bauer D. E., Tuttleton S. E., Lewin S., Gettie A., Blanchard J., Irwin C. E., Safrit J. T., Mittler J., Weinberger L., Kostrikis L. G., Zhang L., Perelson A. S., Ho D. D. Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques // *J Exp Med.* – 1999. – V. 189, № 6. – P. 991-8.
282. Liu Y., McNevin J. P., Holte S., McElrath M. J., Mullins J. I. Dynamics of viral evolution and CTL responses in HIV-1 infection // *PLoS One.* – 2011. – V. 6, № 1. – P. e15639.
283. Henn M. R., Boutwell C. L., Charlebois P., Lennon N. J., Power K. A., Macalalad A. R., Berlin A. M., Malboeuf C. M., Ryan E. M., Gnerre S., Zody M. C., Erlich R. L., Green L. M., Berical A., Wang Y., Casali M., Streeck H., Bloom A. K., Dudek T., Tully D., Newman R., Axten K. L., Gladden A. D., Battis L., Kemper M., Zeng Q., Shea T. P., Gujja S., Zedlack C., Gasser O., Brander C., Hess C., Gunthard H. F., Brumme Z. L., Brumme C. J., Bazner S., Rychert J., Tinsley J. P., Mayer K. H., Rosenberg E., Pereyra F., Levin J. Z., Young S. K., Jessen H., Altfeld M., Birren B. W., Walker B. D., Allen T. M. Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection // *PLoS Pathog.* – 2012. – V. 8, № 3. – P. e1002529.
284. Althaus C. L., De Boer R. J. Dynamics of immune escape during HIV/SIV infection // *PLoS Comput Biol.* – 2008. – V. 4, № 7. – P. e1000103.
285. Mostowy R., Kouyos R. D., Fouchet D., Bonhoeffer S. The role of recombination for the coevolutionary dynamics of HIV and the immune response // *PLoS One.* – 2011. – V. 6, № 2. – P. e16052.
286. Ganusov V. V., Goonetilleke N., Liu M. K., Ferrari G., Shaw G. M., McMichael A. J., Borrow P., Korber B. T., Perelson A. S. Fitness costs and diversity of the cytotoxic T lymphocyte (CTL) response determine the rate of CTL escape during acute and chronic phases of HIV infection // *J Virol.* – 2011. – V. 85, № 20. – P. 10518-28.

287. van Deutekom H. W., Wijnker G., de Boer R. J. The rate of immune escape vanishes when multiple immune responses control an HIV infection // *J Immunol.* – 2013. – V. 191, № 6. – P. 3277-86.
288. Fryer H. R., Frater J., Duda A., Roberts M. G., Investigators S. T., Phillips R. E., McLean A. R. Modelling the evolution and spread of HIV immune escape mutants // *PLoS Pathog.* – 2010. – V. 6, № 11. – P. e1001196.
289. Kelleher A. D., Long C., Holmes E. C., Allen R. L., Wilson J., Conlon C., Workman C., Shaunak S., Olson K., Goulder P., Brander C., Ogg G., Sullivan J. S., Dyer W., Jones I., McMichael A. J., Rowland-Jones S., Phillips R. E. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses // *J Exp Med.* – 2001. – V. 193, № 3. – P. 375-86.
290. Crawford H., Matthews P. C., Schaefer M., Carlson J. M., Leslie A., Kilembe W., Allen S., Ndung'u T., Heckerman D., Hunter E., Goulder P. J. The hypervariable HIV-1 capsid protein residues comprise HLA-driven CD8+ T-cell escape mutations and covarying HLA-independent polymorphisms // *J Virol.* – 2011. – V. 85, № 3. – P. 1384-90.
291. Song H., Pavlicek J. W., Cai F., Bhattacharya T., Li H., Iyer S. S., Bar K. J., Decker J. M., Goonetilleke N., Liu M. K., Berg A., Hora B., Drinker M. S., Eudailey J., Pickeral J., Moody M. A., Ferrari G., McMichael A., Perelson A. S., Shaw G. M., Hahn B. H., Haynes B. F., Gao F. Impact of immune escape mutations on HIV-1 fitness in the context of the cognate transmitted/founder genome // *Retrovirology.* – 2012. – V. 9. – P. 89.
292. Boutwell C. L., Carlson J. M., Lin T. H., Seese A., Power K. A., Peng J., Tang Y., Brumme Z. L., Heckerman D., Schneidewind A., Allen T. M. Frequent and variable cytotoxic-T-lymphocyte escape-associated fitness costs in the human immunodeficiency virus type 1 subtype B Gag proteins // *J Virol.* – 2013. – V. 87, № 7. – P. 3952-65.
293. Cale E. M., Hraber P., Giorgi E. E., Fischer W., Bhattacharya T., Leitner T., Yeh W. W., Gleasner C., Green L. D., Han C. S., Korber B., Letvin N. L. Epitope-specific CD8+ T lymphocytes cross-recognize mutant simian immunodeficiency virus (SIV) sequences but fail to contain very early evolution and eventual fixation of epitope escape mutations during SIV infection // *J Virol.* – 2011. – V. 85, № 8. – P. 3746-57.
294. Ganusov V. V., De Boer R. J. Estimating Costs and Benefits of CTL Escape Mutations in SIV/HIV Infection // *PLoS Comput Biol.* – 2006. – V. 2, № 3. – P. e24.
295. Mostowy R., Kouyos R. D., Hoof I., Hinkley T., Haddad M., Whitcomb J. M., Petropoulos C. J., Kesmir C., Bonhoeffer S. Estimating the fitness cost of escape from HLA presentation in HIV-1 protease and reverse transcriptase // *PLoS Comput Biol.* – 2012. – V. 8, № 5. – P. e1002525.
296. Schneidewind A., Brockman M. A., Sidney J., Wang Y. E., Chen H., Suscovich T. J., Li B., Adam R. I., Allgaier R. L., Mothe B. R., Kuntzen T., Oniangue-Ndza C., Trocha A., Yu X. G., Brander C., Sette A., Walker B. D., Allen T. M. Structural and functional constraints limit options for cytotoxic T-lymphocyte escape in the immunodominant HLA-B27-restricted epitope in human immunodeficiency virus type 1 capsid // *J Virol.* – 2008. – V. 82, № 11. – P. 5594-605.
297. Kawashima Y., Pfafferoth K., Frater J., Matthews P., Payne R., Addo M., Gatanaga H., Fujiwara M., Hachiya A., Koizumi H., Kuse N., Oka S., Duda A., Prendergast A., Crawford H., Leslie A., Brumme Z., Brumme C., Allen T., Brander C., Kaslow R., Tang J., Hunter E., Allen S., Mulenga J., Branch S., Roach T., John M., Mallal S., Ogwu A., Shapiro R., Prado J. G., Fidler S., Weber J., Pybus O. G., Klenerman P., Ndung'u T., Phillips R., Heckerman D., Harrigan P. R., Walker B. D., Takiguchi M., Goulder P. Adaptation of HIV-1 to human leukocyte antigen class I // *Nature.* – 2009. – V. 458, № 7238. – P. 641-5.
298. Matthews P. C., Koyanagi M., Kloverpris H. N., Harndahl M., Stryhn A., Akahoshi T., Gatanaga H., Oka S., Juarez Molina C., Valenzuela Ponce H., Avila Rios S., Cole D., Carlson J., Payne R. P., Ogwu A., Bere A., Ndung'u T., Gounder K., Chen F., Riddell L., Luzzi G., Shapiro R., Brander C., Walker B., Sewell A. K., Reyes Teran G., Heckerman D., Hunter E., Buus S., Takiguchi M., Goulder P. J. Differential clade-specific HLA-B*3501 association with HIV-1 disease outcome is linked to immunogenicity of a single Gag epitope // *J Virol.* – 2012. – V. 86, № 23. – P. 12643-54.

299. Holland J. Generation and replication of defective viral genomes // *Virology* / BN F., DM K. – New York: Raven Press, 1990. – C. 77-99.
300. Huang A. S., Baltimore D. Defective viral particles and viral disease processes // *Nature*. – 1970. – V. 226, № 5243. – P. 325-7.
301. Holland J. J., Villarreal L. P. Purification of defective interfering T particles of vesicular stomatitis and rabies viruses generated in vivo in brains of newborn mice // *Virology*. – 1975. – V. 67, № 2. – P. 438-49.
302. Chattopadhyay S. K., Morse H. C., 3rd, Makino M., Ruscetti S. K., Hartley J. W. Defective virus is associated with induction of murine retrovirus-induced immunodeficiency syndrome // *Proc Natl Acad Sci U S A*. – 1989. – V. 86, № 10. – P. 3862-6.
303. McLain L., Armstrong S. J., Dimmock N. J. One defective interfering particle per cell prevents influenza virus-mediated cytopathology: an efficient assay system // *J Gen Virol*. – 1988. – V. 69 (Pt 6). – P. 1415-9.
304. Voynow S. L., Coffin J. M. Truncated gag-related proteins are produced by large deletion mutants of Rous sarcoma virus and form virus particles // *J. Virol*. – 1985. – V. 55. – P. 79-85.
305. Li D., Lott W. B., Lowry K., Jones A., Thu H. M., Aaskov J. Defective interfering viral particles in acute dengue infections // *PLoS One*. – 2011. – V. 6, № 4. – P. e19447.
306. Aaskov J., Buzacott K., Thu H. M., Lowry K., Holmes E. C. Long-term transmission of defective RNA viruses in humans and *Aedes* mosquitoes // *Science*. – 2006. – V. 311, № 5758. – P. 236-8.
307. Kubo Y., Kakimi K., Higo K., Kobayashi H., Ono T., Iwama Y., Kuribayashi K., Hiai H., Adachi A., Ishimoto A. Possible origin of murine AIDS (MAIDS) virus: conversion of an endogenous retroviral p12gag sequence to a MAIDS-inducing sequence by frameshift mutations // *J Virol*. – 1996. – V. 70, № 9. – P. 6405-9.
308. Paun A., Shaw K., Fisher S., Sammels L. M., Watson M. W., Beilharz M. W. Quantitation of defective and ecotropic viruses during LP-BM5 infection by real time PCR and RT-PCR // *J Virol Methods*. – 2005. – V. 124, № 1-2. – P. 57-63.
309. Cook W. J., Green K. A., Obar J. J., Green W. R. Quantitative analysis of LP-BM5 murine leukemia retrovirus RNA using real-time RT-PCR // *J Virol Methods*. – 2003. – V. 108, № 1. – P. 49-58.
310. Levine B. L., Humeau L. M., Boyer J., MacGregor R. R., Rebello T., Lu X., Binder G. K., Slepishkin V., Lemiale F., Mascola J. R., Bushman F. D., Dropulic B., June C. H. Gene transfer in humans using a conditionally replicating lentiviral vector // *Proc Natl Acad Sci U S A*. – 2006. – V. 103, № 46. – P. 17372-7.
311. Cave D. R., Hendrickson F. M., Huang A. S. Defective interfering virus particles modulate virulence // *J Virol*. – 1985. – V. 55, № 2. – P. 366-73.
312. Marriott A. C., Dimmock N. J. Defective interfering viruses and their potential as antiviral agents // *Rev Med Virol*. – 2010. – V. 20, № 1. – P. 51-62.
313. Dimmock N. J. Defective interfering viruses: modulators of infection // *Microbiol Sci*. – 1985. – V. 2, № 1. – P. 1-7.
314. Barrett A. D., Dimmock N. J. Defective interfering viruses and infections of animals // *Curr Top Microbiol Immunol*. – 1986. – V. 128. – P. 55-84.
315. Weinberger L. S., Schaffer D. V., Arkin A. P. Theoretical design of a gene therapy to prevent AIDS but not human immunodeficiency virus type 1 infection // *J Virol*. – 2003. – V. 77, № 18. – P. 10028-36.
316. D'Costa J., Mansfield S. G., Humeau L. M. Lentiviral vectors in clinical trials: Current status // *Curr Opin Mol Ther*. – 2009. – V. 11, № 5. – P. 554-64.
317. Vignuzzi M., Lopez C. B. Defective viral genomes are key drivers of the virus-host interaction // *Nat Microbiol*. – 2019. – V. 4, № 7. – P. 1075-1087.
318. Chaturvedi S., Vasen G., Pablo M., Chen X., Beutler N., Kumar A., Tanner E., Illouz S., Rahgoshay D., Burnett J., Holguin L., Chen P. Y., Ndjamien B., Ott M., Rodick R., Rogers T., Smith D. M., Weinberger L. S. Identification of a therapeutic interfering particle-A single-dose SARS-CoV-2 antiviral intervention with a high barrier to resistance // *Cell*. – 2021. – V. 184, № 25. – P. 6022-6036 e18.

319. Rezelj V. V., Carrau L., Merwaiss F., Levi L. I., Erazo D., Tran Q. D., Henrion-Lacritick A., Gausson V., Suzuki Y., Shengjuler D., Meyer B., Vallet T., Weger-Lucarelli J., Bernhauerova V., Titievsky A., Sharov V., Pietropaoli S., Diaz-Salinas M. A., Legros V., Pardigon N., Barba-Spaeth G., Brodsky L., Saleh M. C., Vignuzzi M. Defective viral genomes as therapeutic interfering particles against flavivirus infection in mammalian and mosquito hosts // *Nat Commun.* – 2021. – V. 12, № 1. – P. 2290.
320. Sharov V., Rezelj V. V., Galatenko V. V., Titievsky A., Panov J., Chumakov K., Andino R., Vignuzzi M., Brodsky L. Intra- and Inter-cellular Modeling of Dynamic Interaction between Zika Virus and Its Naturally Occurring Defective Viral Genomes // *J Virol.* – 2021. – V. 95, № 22. – P. e0097721.
321. Xiao Y., Lidsky P. V., Shirogane Y., Aviner R., Wu C. T., Li W., Zheng W., Talbot D., Catching A., Doitsh G., Su W., Gekko C. E., Nayak A., Ernst J. D., Brodsky L., Brodsky E., Rousseau E., Capponi S., Bianco S., Nakamura R., Jackson P. K., Frydman J., Andino R. A defective viral genome strategy elicits broad protective immunity against respiratory viruses // *Cell.* – 2021. – V. 184, № 25. – P. 6037-6051 e14.
322. O'Hara P. J., Horodyski F. M., Nichol S. T., Holland J. J. Vesicular stomatitis virus mutants resistant to defective-interfering particles accumulate stable 5'-terminal and fewer 3'-terminal mutations in a stepwise manner // *J Virol.* – 1984. – V. 49, № 3. – P. 793-8.
323. Giachetti C., Holland J. J. Altered replicase specificity is responsible for resistance to defective interfering particle interference of an Sdi- mutant of vesicular stomatitis virus // *J Virol.* – 1988. – V. 62, № 10. – P. 3614-21.
324. DePolo N. J., Holland J. J. Very rapid generation/amplification of defective interfering particles by vesicular stomatitis virus variants isolated from persistent infection // *J Gen Virol.* – 1986. – V. 67 (Pt 6). – P. 1195-8.
325. Li Y., Kappes J. C., Conway J. A., Price R. W., Shaw G. M., Hahn B. H. Molecular characterization of human immunodeficiency virus type 1 cloned directly from uncultured human brain tissue: identification of replication-competent and -defective viral genomes // *J Virol.* – 1991. – V. 65, № 8. – P. 3973-85.
326. Dull T., Zufferey R., Kelly M., Mandel R. J., Nguyen M., Trono D., Naldini L. A third-generation lentivirus vector with a conditional packaging system // *J Virol.* – 1998. – V. 72, № 11. – P. 8463-71.
327. Perrault J., Holland J. J. Absence of transcriptase activity or transcription-inhibiting ability in defective interfering particles of vesicular stomatitis virus // *Virology.* – 1972. – V. 50, № 1. – P. 150-70.
328. Steinhauer D. A., de la Torre J. C., Meier E., Holland J. J. Extreme heterogeneity in populations of vesicular stomatitis virus // *J Virol.* – 1989. – V. 63, № 5. – P. 2072-80.
329. Horodyski F. M., Holland J. J. Viruses isolated from cells persistently infected with vesicular stomatitis virus show altered interactions with defective interfering particles // *J Virol.* – 1980. – V. 36, № 2. – P. 627-31.
330. DePolo N. J., Giachetti C., Holland J. J. Continuing coevolution of virus and defective interfering particles and of viral genome sequences during undiluted passages: virus mutants exhibiting nearly complete resistance to formerly dominant defective interfering particles // *J Virol.* – 1987. – V. 61, № 2. – P. 454-64.
331. Thompson K. A., Yin J. Population dynamics of an RNA virus and its defective interfering particles in passage cultures // *Virol J.* – 2010. – V. 7. – P. 257.
332. Kirkwood T. B., Bangham C. R. Cycles, chaos, and evolution in virus cultures: a model of defective interfering particles // *Proc Natl Acad Sci U S A.* – 1994. – V. 91, № 18. – P. 8685-9.
333. Nelson G. W., Perelson A. S. Modeling defective interfering virus therapy for AIDS: conditions for DIV survival // *Math Biosci.* – 1995. – V. 125, № 2. – P. 127-53.
334. Metzger V. T., Lloyd-Smith J. O., Weinberger L. S. Autonomous targeting of infectious superspreaders using engineered transmissible therapies // *PLoS Comput Biol.* – 2011. – V. 7, № 3. – P. e1002015.
335. Krakauer D. C., Komarova N. L. Levels of selection in positive-strand virus dynamics // *J Evol Biol.* – 2003. – V. 16, № 1. – P. 64-73.

336. Haseltine E. L., Yin J., Rawlings J. B. Implications of decoupling the intracellular and extracellular levels in multi-level models of virus growth // *Biotechnol Bioeng.* – 2008. – V. 101, № 4. – P. 811-20.
337. Guedj J., Neumann A. U. Understanding hepatitis C viral dynamics with direct-acting antiviral agents due to the interplay between intracellular replication and cellular infection dynamics // *J Theor Biol.* – 2010. – V. 267, № 3. – P. 330-40.
338. Woodhouse D. E., Rothenberg R. B., Potterat J. J., Darrow W. W., Muth S. Q., Klovdahl A. S., Zimmerman H. P., Rogers H. L., Maldonado T. S., Muth J. B., et al. Mapping a social network of heterosexuals at high risk for HIV infection // *AIDS.* – 1994. – V. 8, № 9. – P. 1331-6.
339. Lloyd-Smith J. O., Schreiber S. J., Kopp P. E., Getz W. M. Superspreading and the effect of individual variation on disease emergence // *Nature.* – 2005. – V. 438, № 7066. – P. 355-9.
340. Baggaley R. F., Garnett G. P., Ferguson N. M. Modelling the impact of antiretroviral use in resource-poor settings // *PLoS medicine.* – 2006. – V. 3, № 4. – P. e124.
341. Fisher R. A. On the dominance ratio. 1922 // *Bull Math Biol.* – 1990. – V. 52, № 1-2. – P. 297-318; discussion 201-7.
342. Wright S. Evolution in Mendelian Populations // *Genetics.* – 1931. – V. 16, № 2. – P. 97-159.
343. Population genetics, molecular evolution, and the neutral theory : selected papers. / Kimura M. – Chicago: University of Chicago Press, 1994. – xviii, 686 p. p.
344. Barnes W. M. The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion // *Gene.* – 1992. – V. 112, № 1. – P. 29-35.
345. Tindall K. R., Kunkel T. A. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase // *Biochemistry.* – 1988. – V. 27, № 16. – P. 6008-13.
346. Perelson A. S., Neumann A. U., Markowitz M., Leonard J. M., Ho D. D. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time // *Science.* – 1996. – V. 271, № 5255. – P. 1582-6.
347. Smith M. W., Dean M., Carrington M., Winkler C., Huttley G. A., Lomb D. A., Goedert J. J., O'Brien T. R., Jacobson L. P., Kaslow R., Buchbinder S., Vittinghoff E., Vlahov D., Hoots K., Hilgartner M. W., O'Brien S. J. Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC), ALIVE Study // *Science.* – 1997. – V. 277, № 5328. – P. 959-65.
348. Wyand M. S., Manson K. H., Garcia-Moll M., Montefiori D., Desrosiers R. C. Vaccine protection by a triple deletion mutant of simian immunodeficiency virus // *J Virol.* – 1996. – V. 70, № 6. – P. 3724-33.
349. Zhang L. Q., MacKenzie P., Cleland A., Holmes E. C., Brown A. J., Simmonds P. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection // *J Virol.* – 1993. – V. 67, № 6. – P. 3345-56.
350. Putfzner A., Dietrich U., von Eichel U., von Briesen H., Brede H. D., Maniar J. K., Rubsamen-Waigmann H. HIV-1 and HIV-2 infections in a high-risk population in Bombay, India: evidence for the spread of HIV-2 and presence of a divergent HIV-1 subtype // *J Acquir Immune Defic Syndr (1988).* – 1992. – V. 5, № 10. – P. 972-7.
351. Pieniazek D., Janini L. M., Ramos A., Tanuri A., Schechter M., Peralta J. M., Vicente A. C., Pieniazek N. K., Schochetman G., Rayfield M. A. HIV-1 patients may harbor viruses of different phylogenetic subtypes: implications for the evolution of the HIV/AIDS pandemic // *Emerg Infect Dis.* – 1995. – V. 1, № 3. – P. 86-8.
352. Groenink M., Andeweg A. C., Fouchier R. A., Broersen S., van der Jagt R. C., Schuitemaker H., de Goede R. E., Bosch M. L., Huisman H. G., Tersmette M. Phenotype-associated env gene variation among eight related human immunodeficiency virus type 1 clones: evidence for in vivo recombination and determinants of cytotropism outside the V3 domain // *J Virol.* – 1992. – V. 66, № 10. – P. 6175-80.
353. Sabino E. C., Shpaer E. G., Morgado M. G., Korber B. T., Diaz R. S., Bongertz V., Cavalcante S., Galvao-Castro B., Mullins J. I., Mayer A. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil // *J Virol.* – 1994. – V. 68, № 10. – P. 6340-6.

354. Wong J. K., Strain M. C., Porrata R., Reay E., Sankaran-Walters S., Ignacio C. C., Russell T., Pillai S. K., Looney D. J., Dandekar S. In vivo CD8+ T-cell suppression of siv viremia is not mediated by CTL clearance of productively infected cells // *PLoS Pathog.* – 2010. – V. 6, № 1. – P. e1000748.
355. Janeway's Immunobiology, Eighth Edition. / Murphy K. – London, New York: Garland Science, 2011.
356. Goonetilleke N., Liu M. K., Salazar-Gonzalez J. F., Ferrari G., Giorgi E., Ganusov V. V., Keele B. F., Learn G. H., Turnbull E. L., Salazar M. G., Weinhold K. J., Moore S., B C. C. C., Letvin N., Haynes B. F., Cohen M. S., Hraber P., Bhattacharya T., Borrow P., Perelson A. S., Hahn B. H., Shaw G. M., Korber B. T., McMichael A. J. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection // *J Exp Med.* – 2009. – V. 206, № 6. – P. 1253-72.
357. Condra J. H., Holder D. J., Schleif W. A., Blahy O. M., Danovich R. M., Gabryelski L. J., Graham D. J., Laird D., Quintero J. C., Rhodes A., Robbins H. L., Roth E., Shivaprakash M., Yang T., Chodakewitz J. A., Deutsch P. J., Leavitt R. Y., Massari F. E., Mellors J. W., Squires K. E., Steigbigel R. T., Tepler H., Emini E. A. Genetic correlates of in vivo viral resistance to indinavir, a human immunodeficiency virus type 1 protease inhibitor // *J Virol.* – 1996. – V. 70, № 12. – P. 8270-6.
358. Holmes E. C., Zhang L. Q., Simmonds P., Ludlam C. A., Brown A. J. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient // *Proc Natl Acad Sci U S A.* – 1992. – V. 89, № 11. – P. 4835-9.
359. Delwart E. L., Sheppard H. W., Walker B. D., Goudsmit J., Mullins J. I. Human immunodeficiency virus type 1 evolution in vivo tracked by DNA heteroduplex mobility assays // *J Virol.* – 1994. – V. 68, № 10. – P. 6672-83.
360. Liu S. L., Schacker T., Musey L., Shriner D., McElrath M. J., Corey L., Mullins J. I. Divergent patterns of progression to AIDS after infection from the same source: human immunodeficiency virus type 1 evolution and antiviral responses // *J Virol.* – 1997. – V. 71, № 6. – P. 4284-95.
361. Hu W. S., Temin H. M. Retroviral recombination and reverse transcription // *Science.* – 1990. – V. 250, № 4985. – P. 1227-33.
362. Robertson D. L., Sharp P. M., McCutchan F. E., Hahn B. H. Recombination in HIV-1 // *Nature.* – 1995. – V. 374, № 6518. – P. 124-6.
363. Haase A. T., Henry K., Zupancic M., Sedgewick G., Faust R. A., Melroe H., Cavert W., Gebhard K., Staskus K., Zhang Z. Q., Dailey P. J., Balfour H. H., Jr., Erice A., Perelson A. S. Quantitative image analysis of HIV-1 infection in lymphoid tissue // *Science.* – 1996. – V. 274, № 5289. – P. 985-9.
364. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism // *Genetics.* – 1989. – V. 123, № 3. – P. 585-95.
365. Pennings P. S., Kryazhimskiy S., Wakeley J. Loss and recovery of genetic diversity in adapting populations of HIV // *PLoS Genet.* – 2014. – V. 10, № 1. – P. e1004000.
366. Maldarelli F., Kearney M., Palmer S., Stephens R., Mican J., Polis M. A., Davey R. T., Kovacs J., Shao W., Rock-Kress D., Metcalf J. A., Rehm C., Greer S. E., Lucey D. L., Danley K., Alter H., Mellors J. W., Coffin J. M. HIV populations are large and accumulate high genetic diversity in a nonlinear fashion // *J Virol.* – 2013. – V. 87, № 18. – P. 10313-23.
367. Rouzine I. M., Coffin J. M., Weinberger L. S. Fifteen Years Later: Hard and Soft Selection Sweeps Confirm a Large Population Number for HIV In Vivo // *PLoS Genetics.* – 2014. – V. 10 – P. e1004179.
368. An introduction to population genetics: Theory and applications. / Nielsen R., Slatkin M. – 1st edition: Sinaur Associates, Inc, 2013. – 298 p.
369. Brandin E., Thorstensson R., Bonhoeffer S., Albert J. Rapid viral decay in simian immunodeficiency virus-infected macaques receiving quadruple antiretroviral therapy // *J Virol.* – 2006. – V. 80, № 19. – P. 9861-4.
370. Levy D. N., Aldrovandi G. M., Kutsch O., Shaw G. M. Dynamics of HIV-1 recombination in its natural target cells // *Proc Natl Acad Sci U S A.* – 2004. – V. 101, № 12. – P. 4204-9.
371. Palmer S., Kearney M., Maldarelli F., Halvas E. K., Bixby C. J., Bazmi H., Rock D., Falloon J., Davey R. T., Jr., Dewar R. L., Metcalf J. A., Hammer S., Mellors J. W., Coffin J. M. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis // *J Clin Microbiol.* – 2005. – V. 43, № 1. – P. 406-13.

372. Nowak M. A., Bangham C. R. Population dynamics of immune responses to persistent viruses // *Science*. – 1996. – V. 272, № 5258. – P. 74-9.
373. Muller V., Marea A. F., De Boer R. J. Small variations in multiple parameters account for wide variations in HIV-1 set-points: a novel modelling approach // *Proc Biol Sci*. – 2001. – V. 268, № 1464. – P. 235-42.
374. Wodarz D. Helper-dependent vs. helper-independent CTL responses in HIV infection: implications for drug therapy and resistance // *J Theor Biol*. – 2001. – V. 213, № 3. – P. 447-59.
375. Sergeev R. A., Batorsky R. E., Rouzine I. M. Model with two types of CTL regulation and experiments on CTL dynamics // *J Theor Biol*. – 2010. – V. 263, № 3. – P. 369-84.
376. Ayme V., Petit-Pierre J., Souche S., Palloix A., Moury B. Molecular dissection of the potato virus Y VPg virulence factor reveals complex adaptations to the pvr2 resistance allelic series in pepper // *J Gen Virol*. – 2007. – V. 88, № Pt 5. – P. 1594-1601.
377. Novella I. S. Negative effect of genetic bottlenecks on the adaptability of vesicular stomatitis virus // *J Mol Biol*. – 2004. – V. 336, № 1. – P. 61-7.
378. Weinreich D. M., Watson R. A., Chao L. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories // *Evolution*. – 2005. – V. 59, № 6. – P. 1165-74.
379. Poon A. F., Swenson L. C., Dong W. W., Deng W., Kosakovsky Pond S. L., Brumme Z. L., Mullins J. I., Richman D. D., Harrigan P. R., Frost S. D. Phylogenetic analysis of population-based and deep sequencing data to identify coevolving sites in the nef gene of HIV-1 // *Mol Biol Evol*. – 2010. – V. 27, № 4. – P. 819-32.
380. Kessler D. A., Levine H., Ridgway D., Tsimring L. Evolution on a smooth landscape // *J. Stat. Phys.* – 1997. – V. 87. – P. 519-544.
381. Tsimring L. S., Levine H., Kessler D. RNA virus evolution via a fitness-space model // *Phys. Rev. Lett.* – 1996. – V. 76. – P. 4440-4443.
382. Mansky L. M., Temin H. M. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase // *J. Virol.* – 1995. – V. 69. – P. 5087-5094.
383. Haldane J. B. S. A mathematical theory of natural and artificial selection. Part V: Selection and mutation // *Proc. Camb. Phil. Soc.* – 1927. – V. 23. – P. 838-844.
384. Kimura M. On the probability of fixation of mutant genes in a population // *Genetics*. – 1962. – V. 47. – P. 713-719.
385. Goyal S., Balick D. J., Jerison E. R., Neher R. A., Shraiman B. I., Desai M. M. Dynamic mutation-selection balance as an evolutionary attractor // *Genetics*. – 2012. – V. 191, № 4. – P. 1309-19.
386. Hallatschek O., Nelson D. R. Gene surfing in expanding populations // *Theor Popul Biol*. – 2008. – V. 73, № 1. – P. 158-70.
387. Excoffier L., Ray N. Surfing during population expansions promotes genetic revolutions and structuration // *Trends Ecol Evol*. – 2008. – V. 23, № 7. – P. 347-51.
388. *Mathematical Modeling of Evolution. Volume 1: One-Locus and Multi-Locus Theory and Recombination. De Gruyter Series in Mathematics and Life Sciences.* / Rouzine I. – Berlin/Boston: De Gruyter, 2020. De Gruyter Series in Mathematics and Life Sciences. – 185 p.
389. *The genetical theory of natural selection.* / Fisher D. S. – Oxford, United Kingdom: Clarendon Press, 1930.
390. Desai M. M., Walczak A. M., Fisher D. S. Genetic diversity and the structure of genealogies in rapidly adapting populations // *Genetics*. – 2013. – V. 193, № 2. – P. 565-85.
391. Kingman J. F. C. The coalescent // *Stochastic Processes and their Applications*. – 1982. – V. 13, № 3. – P. 235-248.
392. Neuhauser C., Krone S. M. The genealogy of samples in models with selection // *Genetics*. – 1997. – V. 145, № 2. – P. 519-34.
393. Kingman J. F. Origins of the coalescent. 1974-1982 // *Genetics*. – 2000. – V. 156, № 4. – P. 1461-3.
394. Brunet E., Derrida B., Simon D. Universal tree structures in directed polymers and models of evolving populations // *Phys Rev E Stat Nonlin Soft Matter Phys*. – 2008. – V. 78, № 6 Pt 1. – P. 061102.

395. Brunet E., Derrida B., Mueller A. H., Munier S. Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization // *Phys Rev E Stat Nonlin Soft Matter Phys.* – 2007. – V. 76, № 4 Pt 1. – P. 041104.
396. Good B. H., Walczak A. M., Neher R. A., Desai M. M. Genetic diversity in the interference selection limit // *PLoS Genet.* – 2014. – V. 10, № 3. – P. e1004222.
397. Eguia R. T., Crawford K. H. D., Stevens-Ayers T., Kelnhofer-Millevolte L., Greninger A. L., Englund J. A., Boeckh M. J., Bloom J. D. A human coronavirus evolves antigenically to escape antibody immunity // *PLoS Pathog.* – 2021. – V. 17, № 4. – P. e1009453.
398. Rochman N. D., Wolf Y. I., Faure G., Mutz P., Zhang F., Koonin E. V. Ongoing global and regional adaptive evolution of SARS-CoV-2 // *Proc Natl Acad Sci U S A.* – 2021. – V. 118, № 29.
399. Haynes W. A., Kamath K., Bozekowski J., Baum-Jones E., Campbell M., Casanovas-Massana A., Daugherty P. S., Dela Cruz C. S., Dhal A., Farhadian S. F., Fitzgibbons L., Fournier J., Jhatro M., Jordan G., Klein J., Lucas C., Kessler D., Luchsinger L. L., Martinez B., Catherine Muenker M., Pischel L., Reifert J., Sawyer J. R., Waitz R., Wunder E. A., Jr., Zhang M., Yale I. T., Iwasaki A., Ko A., Shon J. C. High-resolution epitope mapping and characterization of SARS-CoV-2 antibodies in large cohorts of subjects with COVID-19 // *Commun Biol.* – 2021. – V. 4, № 1. – P. 1317.
400. Greaney A. J., Loes A. N., Crawford K. H. D., Starr T. N., Malone K. D., Chu H. Y., Bloom J. D. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies // *Cell Host Microbe.* – 2021. – V. 29, № 3. – P. 463-476 e6.
401. Greaney A. J., Starr T. N., Gilchuk P., Zost S. J., Binshtein E., Loes A. N., Hilton S. K., Huddleston J., Eguia R., Crawford K. H. D., Dingens A. S., Nargi R. S., Sutton R. E., Suryadevara N., Rothlauf P. W., Liu Z., Whelan S. P. J., Carnahan R. H., Crowe J. E., Jr., Bloom J. D. Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition // *Cell Host Microbe.* – 2021. – V. 29, № 1. – P. 44-57 e9.
402. Rouzine I. M., Rozhnova G. Evolutionary implications of SARS-CoV-2 vaccination for the future design of vaccination strategies // *Commun Medicine.* – 2023. – V. 3. – P. 86.
403. SARS-CoV-2 variants of concern as of 27 January 2022 // European Centre for Disease Prevention and Control. – 2022. <https://www.ecdc.europa.eu/en/covid-19/variants-concern>.
404. Martin D., Weaver S., H T., San E., SD S., Wilkinson E., Lucaci A., Giandhari G., Naidoo S., Pillay Y., Singh L., Lessells R., NGS-SA, COVID-19 Genomics UK (COG-UK), Ravindra K., JO W., Nekturenko A, B M., GW H., P L., OA M., DL R., T O., S K.-P. The emergence and ongoing convergent evolution of the N501Y lineages coincides with a major global shift in the SARS-CoV-2 selective landscape // *Cell.* – 2021. – V. 184. – P. P5189-5200.e7.
405. Ghafari M., Liu Q., Dhillon, A, Katzourakis A., Weissman D. Investigating the evolutionary origins of the first three SARS-CoV-2 variants of concern // *Frontiers in Virology.* – 2022. – V. 2.
406. Markov P. V., Ghafari M., Beer M., Lythgoe K., Simmonds P., Stilianakis N. I., Katzourakis A. The evolution of SARS-CoV-2 // *Nat Rev Microbiol.* – 2023. – V. 21, № 6. – P. 361-379.
407. Tay J. H., Porter A. F., Wirth W., Duchene S. The Emergence of SARS-CoV-2 Variants of Concern Is Driven by Acceleration of the Substitution Rate // *Mol Biol Evol.* – 2022. – V. 39, № 2.
408. Otto S. P., Day T., Arino J., Colijn C., Dushoff J., Li M., Mechai S., Van Domselaar G., Wu J., Earn D. J. D., Ogden N. H. The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic // *Curr Biol.* – 2021. – V. 31, № 14. – P. R918-R929.
409. Kemp S. A., Collier D. A., Datir R. P., Ferreira I., Gayed S., Jahun A., Hosmillo M., Rees-Spear C., Mlcochova P., Lumb I. U., Roberts D. J., Chandra A., Temperton N., Collaboration C.-N. B. C.-., Consortium C.-G. U., Sharrocks K., Blane E., Modis Y., Leigh K. E., Briggs J. A. G., van Gils M. J., Smith K. G. C., Bradley J. R., Smith C., Doffinger R., Ceron-Gutierrez L., Barcenas-Morales G., Pollock D. D., Goldstein R. A., Smielewska A., Skittrall J. P., Gouliouris T., Goodfellow I. G., Gkrania-Klotsas E., Illingworth C. J. R., McCoy L. E., Gupta R. K. SARS-CoV-2 evolution during treatment of chronic infection // *Nature.* – 2021. – V. 592, № 7853. – P. 277-282.
410. Corey L., Beyrer C., Cohen M. S., Michael N. L., Bedford T., Rolland M. SARS-CoV-2 Variants in Patients with Immunosuppression // *N Engl J Med.* – 2021. – V. 385, № 6. – P. 562-566.

411. Ignatieva A., Hein J., Jenkins P. A. Ongoing Recombination in SARS-CoV-2 Revealed through Genealogical Reconstruction // *Mol Biol Evol.* – 2022. – V. 39, № 2.
412. Jackson B., Boni M. F., Bull M. J., Collieran A., Colquhoun R. M., Darby A. C., Haldenby S., Hill V., Lucaci A., McCrone J. T., Nicholls S. M., O'Toole A., Pacchiarini N., Poplawski R., Scher E., Todd F., Webster H. J., Whitehead M., Wierzbicki C., Consortium C.-G. U., Loman N. J., Connor T. R., Robertson D. L., Pybus O. G., Rambaut A. Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic // *Cell.* – 2021. – V. 184, № 20. – P. 5179-5188 e8.
413. Yi H. 2019 Novel Coronavirus Is Undergoing Active Recombination // *Clin Infect Dis.* – 2020. – V. 71, № 15. – P. 884-887.
414. Lau M. S. Y., Grenfell B., Thomas M., Bryan M., Nelson K., Lopman B. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA // *Proc Natl Acad Sci U S A.* – 2020. – V. 117, № 36. – P. 22430-22435.
415. Liu Y., Eggo R. M., Kucharski A. J. Secondary attack rate and superspreading events for SARS-CoV-2 // *Lancet.* – 2020. – V. 395, № 10227. – P. e47.
416. Good B. H., Desai M. M. The impact of macroscopic epistasis on long-term evolutionary dynamics // *Genetics.* – 2015. – V. 199, № 1. – P. 177-90.
417. Jerison E. R., Desai M. M. Genomic investigations of evolutionary dynamics and epistasis in microbial evolution experiments // *Curr Opin Genet Dev.* – 2015. – V. 35. – P. 33-9.
418. Balding D. J. A tutorial on statistical methods for population association studies // *Nat Rev Genet.* – 2006. – V. 7, № 10. – P. 781-91.
419. Kryazhimskiy S., Dushoff J., Bazykin G. A., Plotkin J. B. Prevalence of epistasis in the evolution of influenza A surface proteins // *PLoS Genet.* – 2011. – V. 7, № 2. – P. e1001301.
420. Seibert C. W., Rahmat S., Krammer F., Palese P., Bouvier N. M. Efficient transmission of pandemic H1N1 influenza viruses with high-level oseltamivir resistance // *J Virol.* – 2012. – V. 86, № 9. – P. 5386-9.
421. Byarugaba D. K., Erima B., Millard M., Kibuuka H., Lkwago L., Bwogi J., Mimbe D., Kiconco J. B., Tugume T., Mworosi E. A., Turner J., McKenzie P. P., Webby R. R., Webster R. G., Foret C., Ducatez M. F., Coldren R., Wabwire-Mangen F., Krauss S. Whole-genome analysis of influenza A(H1N1)pdm09 viruses isolated in Uganda from 2009 to 2011 // *Influenza Other Respir Viruses.* – 2016. – V. 10, № 6. – P. 486-492.
422. Otte A., Marriott A. C., Dreier C., Dove B., Mooren K., Klingent T. R., Sauter M., Thompson K. A., Bennett A., Klingel K., van Riel D., McHardy A. C., Carroll M. W., Gabriel G. Evolution of 2009 H1N1 influenza viruses during the pandemic correlates with increased viral pathogenicity and transmissibility in the ferret model // *Sci Rep.* – 2016. – V. 6. – P. 28583.
423. Takahashi T., Song J., Suzuki T., Kawaoka Y. Mutations in NA that induced low pH-stability and enhanced the replication of pandemic (H1N1) 2009 influenza A virus at an early stage of the pandemic // *PLoS One.* – 2013. – V. 8, № 5. – P. e64439.
424. Hom N., Gentles L., Bloom J. D., Lee K. K. Deep Mutational Scan of the Highly Conserved Influenza A Virus M1 Matrix Protein Reveals Substantial Intrinsic Mutational Tolerance // *J Virol.* – 2019. – V. 93, № 13.
425. Lee J. M., Huddleston J., Doud M. B., Hooper K. A., Wu N. C., Bedford T., Bloom J. D. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants // *Proc Natl Acad Sci U S A.* – 2018. – V. 115, № 35. – P. E8276-E8285.
426. Shah P., McCandlish D. M., Plotkin J. B. Contingency and entrenchment in protein evolution under purifying selection // *Proc Natl Acad Sci U S A.* – 2015. – V. 112, № 25. – P. E3226-35.
427. McCandlish D. M., Shah P., Plotkin J. B. Epistasis and the Dynamics of Reversion in Molecular Evolution // *Genetics.* – 2016. – V. 203, № 3. – P. 1335-51.
428. Berg J., Lassig M., Wagner A. Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications // *BMC Evol Biol.* – 2004. – V. 4. – P. 51.
429. Sella G., Hirsh A. E. The application of statistical physics to evolutionary biology // *Proc Natl Acad Sci U S A.* – 2005. – V. 102, № 27. – P. 9541-6.

430. Rice D. P., Good B. H., Desai M. M. The evolutionarily stable distribution of fitness effects // *Genetics*. – 2015. – V. 200, № 1. – P. 321-9.
431. Okoye A. A., Picker L. J. CD4(+) T-cell depletion in HIV infection: mechanisms of immunological failure // *Immunol Rev*. – 2013. – V. 254, № 1. – P. 54-64.
432. Perelson A. S., Essunger P., Cao Y., Vesanen M., Hurley A., Saksela K., Markowitz M., Ho D. D. Decay characteristics of HIV-1-infected compartments during combination therapy // *Nature*. – 1997. – V. 387. – P. 188-191.
433. Lambotte O., Boufassa F., Madec Y., Nguyen A., Goujard C., Meyer L., Rouzioux C., Venet A., Delfraissy J. F., Group S.-H. S. HIV controllers: a homogeneous group of HIV-1-infected patients with spontaneous control of viral replication // *Clin Infect Dis*. – 2005. – V. 41, № 7. – P. 1053-6.
434. Rouzine I. M., Murali-Krishna K., Ahmed R. Generals die in friendly fire, or modeling immune response to HIV // *J. Comput. Appl. Math.* – 2005. – V. 184. – P. 258–274.
435. Sergeev R. A., Batorsky R. E., Coffin J. M., Rouzine I. M. Interpreting the effect of vaccination on steady state infection in animals challenged with Simian immunodeficiency virus // *J Theor Biol*. – 2010. – V. 263, № 3. – P. 385-92.
436. Arnaout R. A., Lloyd A. L., O'Brien T. R., Goedert J. J., Leonard J. M., Nowak M. A. A simple relationship between viral load and survival time in HIV-1 infection // *Proc Natl Acad Sci U S A*. – 1999. – V. 96, № 20. – P. 11549-53.
437. Pearson J. E., Krapivsky P., Perelson A. S. Stochastic theory of early viral infection: continuous versus burst production of virions // *PLoS Comput Biol*. – 2011. – V. 7, № 2. – P. e1001058.
438. Nowak M. A., Lloyd A. L., Vasquez G. M., Wiltout T. A., Wahl L. M., Bischofberger N., Williams J., Kinter A., Fauci A. S., Hirsch V. M., Lifson J. D. Viral dynamics of primary viremia and antiretroviral therapy in simian immunodeficiency virus infection // *J. Virol*. – 1997. – V. 71, № 10. – P. 7518-7525.
439. Ho Y. C., Shan L., Hosmane N. N., Wang J., Laskey S. B., Rosenbloom D. I., Lai J., Blankson J. N., Siliciano J. D., Siliciano R. F. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure // *Cell*. – 2013. – V. 155, № 3. – P. 540-51.
440. Sedaghat A. R., Siliciano R. F., Wilke C. O. Low-level HIV-1 replication and the dynamics of the resting CD4+ T cell reservoir for HIV-1 in the setting of HAART // *BMC infectious diseases*. – 2008. – V. 8. – P. 2.
441. Sedaghat A. R., Siliciano J. D., Brennan T. P., Wilke C. O., Siliciano R. F. Limits on replenishment of the resting CD4+ T cell reservoir for HIV in patients on HAART // *PLoS Pathog*. – 2007. – V. 3, № 8. – P. e122.
442. Rong L., Perelson A. S. Modeling HIV persistence, the latent reservoir, and viral blips // *Journal of theoretical biology*. – 2009. – V. 260, № 2. – P. 308-31.
443. Rong L., Perelson A. S. Modeling latently infected cell activation: viral and latent reservoir persistence, and viral blips in HIV-infected patients on potent therapy // *PLoS computational biology*. – 2009. – V. 5, № 10. – P. e1000533.
444. Li Q., Duan L., Estes J. D., Ma Z. M., Rourke T., Wang Y., Reilly C., Carlis J., Miller C. J., Haase A. T. Peak SIV replication in resting memory CD4+ T cells depletes gut lamina propria CD4+ T cells // *Nature*. – 2005. – V. 434, № 7037. – P. 1148-52.
445. Stafford M. A., Corey L., Cao Y., Daar E. S., Ho D. D., Perelson A. S. Modeling plasma virus concentration during primary HIV infection // *Journal of theoretical biology*. – 2000. – V. 203, № 3. – P. 285-301.
446. Klatt N. R., Shudo E., Ortiz A. M., Engram J. C., Paiardini M., Lawson B., Miller M. D., Else J., Pandrea I., Estes J. D., Apetrei C., Schmitz J. E., Ribeiro R. M., Perelson A. S., Silvestri G. CD8+ lymphocytes control viral replication in SIVmac239-infected rhesus macaques without decreasing the lifespan of productively infected cells // *PLoS Pathog*. – 2010. – V. 6, № 1. – P. e1000747.
447. Chun T. W., Engel D., Mizell S. B., Ehler L. A., Fauci A. S. Induction of HIV-1 replication in latently infected CD4+ T cells using a combination of cytokines // *The Journal of experimental medicine*. – 1998. – V. 188, № 1. – P. 83-91.
448. Siliciano R. F., Greene W. C. HIV latency // *Cold Spring Harb Perspect Med*. – 2011. – V. 1, № 1. – P. a007096.

449. Davenport M. P., Ribeiro R. M., Perelson A. S. Kinetics of virus-specific CD8⁺ T cells and the control of HIV infection // *J. Virol.* – 2004. – V. 78. – P. 10096-10103.
450. De Boer R. J. Understanding the failure of CD8⁺ T-cell vaccination against simian/human immunodeficiency virus // *Journal of virology.* – 2007. – V. 81, № 6. – P. 2838-48.
451. De Boer R. J., Homann D., Perelson A. S. Different dynamics of CD4 and CD8 T cell responses during and after acute lymphocytic choriomeningitis virus infection // *J. Immunol.* – 2003. – V. 171. – P. 3928-3935.
452. Ogg G. S., Jin X., Bonhoeffer S., Dunbar P. R., Nowak M. A., Monard S., Segal J. P., Cao Y., Rowland-Jones S. L., Cerundolo V., Hurley A., Markowitz M., Ho D. D., Nixon D. F., McMichael A. J. Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA // *Science.* – 1998. – V. 279, № 5359. – P. 2103-6.
453. Doitsh G., Cavrois M., Lassen K. G., Zepeda O., Yang Z., Santiago M. L., Hebbeler A. M., Greene W. C. Abortive HIV infection mediates CD4 T cell depletion and inflammation in human lymphoid tissue // *Cell.* – 2010. – V. 143, № 5. – P. 789-801.
454. Finzi D., Blankson J., Siliciano J. D., Margolick J. B., Chadwick K., Pierson T., Smith K., Lisziewicz J., Lori F., Flexner C., Quinn T. C., Chaisson R. E., Rosenberg E., Walker B., Gange S., Gallant J., Siliciano R. F. Latent infection of CD4⁺ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy // *Nat Med.* – 1999. – V. 5, № 5. – P. 512-7.
455. Matano T., Shibata R., Siemon C., Connors M., Lane H. C., Martin M. A. Administration of an anti-CD8 monoclonal antibody interferes with the clearance of chimeric simian/human immunodeficiency virus during primary infections of rhesus macaques // *Journal of virology.* – 1998. – V. 72, № 1. – P. 164-9.
456. Peressin M., Proust A., Schmidt S., Su B., Lambotin M., Biedma M. E., Laumond G., Decoville T., Holl V., Moog C. Efficient transfer of HIV-1 in trans and in cis from Langerhans dendritic cells and macrophages to autologous T lymphocytes // *AIDS.* – 2014.10.1097/QAD.000000000000193.
457. Archin N. M., Vaidya N. K., Kuruc J. D., Liberty A. L., Wiegand A., Kearney M. F., Cohen M. S., Coffin J. M., Bosch R. J., Gay C. L., Eron J. J., Margolis D. M., Perelson A. S. Immediate antiviral therapy appears to restrict resting CD4⁺ cell HIV-1 infection without accelerating the decay of latent infection // *Proceedings of the National Academy of Sciences of the United States of America.* – 2012. – V. 109, № 24. – P. 9523-8.
458. Deeks S. G. HIV: Shock and kill // *Nature.* – 2012. – V. 487, № 7408. – P. 439-40.
459. Steinhauer D. A., Domingo E., Holland J. J. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase // *Gene.* – 1992. – V. 122, № 2. – P. 281-8.
460. Vignuzzi M., Stone J. K., Arnold J. J., Cameron C. E., Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population // *Nature.* – 2006. – V. 439, № 7074. – P. 344-8.
461. Pfeiffer J. K., Kirkegaard K. A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity // *Proc Natl Acad Sci U S A.* – 2003. – V. 100, № 12. – P. 7289-94.
462. Korboukh V. K., Lee C. A., Acevedo A., Vignuzzi M., Xiao Y., Arnold J. J., Hemperly S., Graci J. D., August A., Andino R., Cameron C. E. RNA virus population diversity, an optimum for maximal fitness and virulence // *J Biol Chem.* – 2014. – V. 289, № 43. – P. 29531-44.
463. Acevedo A., Andino R. Library preparation for highly accurate population sequencing of RNA viruses // *Nat Protoc.* – 2014. – V. 9, № 7. – P. 1760-9.
464. Whitton J. L., Cornell C. T., Feuer R. Host and virus determinants of picornavirus pathogenesis and tropism // *Nat Rev Microbiol.* – 2005. – V. 3, № 10. – P. 765-76.
465. Biebricher C. K., Eigen M. The error threshold // *Virus Res.* – 2005. – V. 107, № 2. – P. 117-27.
466. Eigen M. Error catastrophe and antiviral strategy // *Proc Natl Acad Sci U S A.* – 2002. – V. 99, № 21. – P. 13374-6.
467. Clarke D. K., Duarte E. A., Elena S. F., Moya A., Domingo E., Holland J. The red queen reigns in the kingdom of RNA viruses // *Proc Natl Acad Sci U S A.* – 1994. – V. 91, № 11. – P. 4821-4.

468. Gog J. R., Rimmelzwaan G. F., Osterhaus A. D., Grenfell B. T. Population dynamics of rapid fixation in cytotoxic T lymphocyte escape mutants of influenza A // *Proc Natl Acad Sci U S A.* – 2003. – V. 100, № 19. – P. 11143-7.
469. Grenfell B. T., Pybus O. G., Gog J. R., Wood J. L., Daly J. M., Mumford J. A., Holmes E. C. Unifying the epidemiological and evolutionary dynamics of pathogens // *Science.* – 2004. – V. 303, № 5656. – P. 327-32.
470. Carrat F., Vergu E., Ferguson N. M., Lemaître M., Cauchemez S., Leach S., Valleron A. J. Time lines of infection and disease in human influenza: a review of volunteer challenge studies // *Am J Epidemiol.* – 2008. – V. 167, № 7. – P. 775-85.
471. Biggerstaff M., Cauchemez S., Reed C., Gambhir M., Finelli L. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature // *BMC Infect Dis.* – 2014. – V. 14. – P. 480.
472. Principles of Plant Virology. / Astier S.: Science Publishers, 2007.
473. Evolutionary Dynamics: Exploring the Equations of Life. / Nowak M. A. – Cambridge, USA: Harvard University Press, 2006.
474. Evolutionary Ecology of Parasites. / Poulin R.: Princeton University Press, 2007.
475. Evolutionary Theory: Mathematical and Conceptual Foundations. / Rice S.: Sinauer Associated, 2004.
476. Park A. W., Daly J. M., Lewis N. S., Smith D. J., Wood J. L., Grenfell B. T. Quantifying the impact of immune escape on transmission dynamics of influenza // *Science.* – 2009. – V. 326, № 5953. – P. 726-8.
477. Fischer W., Ganusov V. V., Giorgi E. E., Hraber P. T., Keele B. F., Leitner T., Han C. S., Gleasner C. D., Green L., Lo C. C., Nag A., Wallstrom T. C., Wang S., McMichael A. J., Haynes B. F., Hahn B. H., Perelson A. S., Borrow P., Shaw G. M., Bhattacharya T., Korber B. T. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing // *PLoS One.* – 2010. – V. 5, № 8. – P. e12303.
478. Ogg G. S., Jin X., Bonhoeffer S., Moss P., Nowak M. A., Monard S., Segal J. P., Cao Y., Rowland-Jones S. L., Hurley A., Markowitz M., Ho D. D., McMichael A. J., Nixon D. F. Decay kinetics of human immunodeficiency virus-specific effector cytotoxic T lymphocytes after combination antiretroviral therapy // *J Virol.* – 1999. – V. 73, № 1. – P. 797-800.
479. Moore M. D., Fu W., Nikolaitchik O., Chen J., Ptak R. G., Hu W. S. Dimer initiation signal of human immunodeficiency virus type 1: its role in partner selection during RNA copackaging and its effects on recombination // *J Virol.* – 2007. – V. 81, № 8. – P. 4002-11.
480. Chen J., Nikolaitchik O., Singh J., Wright A., Bencsics C. E., Coffin J. M., Ni N., Lockett S., Pathak V. K., Hu W. S. High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis // *Proc Natl Acad Sci U S A.* – 2009. – V. 106, № 32. – P. 13535-40.
481. Josefsson L., King M. S., Makitalo B., Brannstrom J., Shao W., Maldarelli F., Kearney M. F., Hu W. S., Chen J., Gaines H., Mellors J. W., Albert J., Coffin J. M., Palmer S. E. Majority of CD4+ T cells from peripheral blood of HIV-1-infected individuals contain only one HIV DNA molecule // *Proc Natl Acad Sci U S A.* – 2011. – V. 108, № 27. – P. 11199-204.
482. Koldej R. M., Anson D. S. Refinement of lentiviral vector for improved RNA processing and reduced rates of self inactivation repair // *BMC Biotechnol.* – 2009. – V. 9. – P. 86.
483. An D. S., Morizono K., Li Q. X., Mao S. H., Lu S., Chen I. S. An inducible human immunodeficiency virus type 1 (HIV-1) vector which effectively suppresses HIV-1 replication // *J Virol.* – 1999. – V. 73, № 9. – P. 7671-7.
484. Bukovsky A. A., Song J. P., Naldini L. Interaction of human immunodeficiency virus-derived vectors with wild-type virus in transduced cells // *J Virol.* – 1999. – V. 73, № 8. – P. 7087-92.
485. Wang D., Hicks C. B., Goswami N. D., Tafoya E., Ribeiro R. M., Cai F., Perelson A. S., Gao F. Evolution of drug-resistant viral populations during interruption of antiretroviral therapy // *J Virol.* – 2011. – V. 85, № 13. – P. 6403-15.

486. Hinkley T., Martins J., Chappey C., Haddad M., Stawiski E., Whitcomb J. M., Petropoulos C. J., Bonhoeffer S. A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase // *Nat Genet.* – 2011. – V. 43, № 5. – P. 487-9.
487. Messer P. W., Neher R. A. Estimating the strength of selective sweeps from deep population diversity data // *Genetics.* – 2012. – V. 191, № 2. – P. 593-605.
488. Paillart J. C., Shehu-Xhilaga M., Marquet R., Mak J. Dimerization of retroviral RNA genomes: an inseparable pair // *Nat Rev Microbiol.* – 2004. – V. 2, № 6. – P. 461-72.
489. Ke R., Lloyd-Smith J. O. Evolutionary analysis of human immunodeficiency virus type 1 therapies based on conditionally replicating vectors // *PLoS Comput Biol.* – 2012. – V. 8, № 10. – P. e1002744.
490. Rouzine I. M., McKenzie F. E. Link between immune response and parasite synchronization in malaria // *Proc Natl Acad Sci U S A.* – 2003. – V. 100, № 6. – P. 3473-8.
491. Chen C. J., Banerjee A. C., Harmison G. G., Haglund K., Schubert M. Multitarget-ribozyme directed to cleave at up to nine highly conserved HIV-1 env RNA regions inhibits HIV-1 replication--potential effectiveness against most presently sequenced HIV-1 isolates // *Nucleic Acids Res.* – 1992. – V. 20, № 17. – P. 4581-9.
492. Klimatcheva E., Planelles V., Day S. L., Fulreader F., Renda M. J., Rosenblatt J. Defective lentiviral vectors are efficiently trafficked by HIV-1 and inhibit its replication // *Mol Ther.* – 2001. – V. 3, № 6. – P. 928-39.
493. Shirreff G., Pellis L., Laeyendecker O., Fraser C. Transmission selects for HIV-1 strains of intermediate virulence: a modelling approach // *PLoS computational biology.* – 2011. – V. 7, № 10. – P. e1002185.
494. *Infectious Diseases of Humans : Dynamics and Control* /Anderson B., May R. M., Anderson R. M.: Oxford Science Publications, 1992.
495. Re F., Braaten D., Franke E. K., Luban J. Human immunodeficiency virus type 1 vpr arrests the cell cycle in G2 by inhibiting activation of p34cdc2-cyclin B // *J. Virol.* – 1995. – V. 69. – P. 6859-6864.
496. Frankel A. D., Young J. A. HIV-1: fifteen proteins and an RNA // *Annu Rev Biochem.* – 1998. – V. 67. – P. 1-25.
497. Bamdres J. D., Shaw A. S., Ratner L. HIV-1 nef protein downregulation of CD4 surface expression: Relevance of the Ick binding domain of CD4 // *Virology.* – 1994. – V. 207. – P. 338-341.
498. Rankin D. J., Bargum K., Kokko H. The tragedy of the commons in evolutionary biology // *Trends Ecol Evol.* – 2007. – V. 22, № 12. – P. 643-51.