

На правах рукописи



Ружин Игорь Мартынович

Математические методы популяционной генетики и их применение к вирусологии

1.5.8. Математическая биология, биоинформатика

Автореферат диссертации на соискание ученой степени

доктора физико-математических наук

Санкт-Петербург – 2024

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте эволюционной физиологии и биохимии им. И.М. Сеченова РАН, Санкт-Петербург

Официальные оппоненты:

Братусь Александр Сергеевич, доктор физико-математических наук, профессор, зав. кафедрой прикладной математики, Российский университет транспорта (МИИТ), Москва

Саакян Давид Багратович, доктор физико-математических наук, ведущий научный сотрудник, Ереванский физический институт, Ереван, Армения

Розанова Ольга Сергеевна, доктор физико-математических наук, доцент, профессор, Московский государственный университет имени М.В. Ломоносова

Ведущая организация: Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М.В. Ломоносова» (МГУ)

Защита состоится на заседании диссертационного совета 24.1.101.01 на базе Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН) по адресу: 127051, г. Москва, Большой Каретный переулок, д.19 стр. 1.

С диссертацией можно ознакомиться в библиотеке ФГБУН Института проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН), а также на сайте ИППИ РАН (www.iitp.ru)

Автореферат разослан “ “ 2024 г.

Ученый секретарь диссертационного совета

доктор биологических наук

Казенников О.В.

Общая характеристика работы

Актуальность исследований. Многие вирусы (Вирус Иммунодефицита Человека (ВИЧ), вирус гепатита С, вирус гриппа, SARS-CoV-2) эволюционируют чрезвычайно быстро и, как следствие, обладают огромным генетическим разнообразием (рисунок 1). 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, вирус вырабатывал устойчивость ко всем лекарствам в течение нескольких дней или недель после начала терапии.

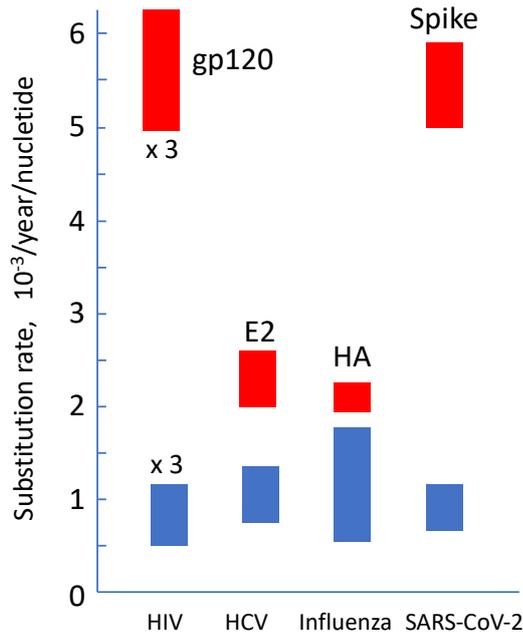


Рисунок 1. Скорость замещений в вирусах. Синие прямоугольники показывают интервалы медианных значений для наиболее быстро и наиболее медленно развивающихся подтипов ВИЧ, вируса гепатита С, вируса гриппа и SARS-CoV-2 для всего генома. Красные прямоугольники соответствуют внешним белкам вируса, на которые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

Причина такого беспрецедентного поведения, в то время, была неизвестна. В 1995 году появилась важная идея о причине неудач антиретровирусной терапии (Coffin, 1995), 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, «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 подтверждаются прямыми экспериментами, из-за огромного шума в данных непонятного происхождения. Более того, математическая теория многолокусной эволюции ранее полностью отсутствовала, из-за трудности описания стохастической динамики экспоненциально большого числа возможных последовательностей.

Цели диссертации. Адаптировать модели популяционной генетики к вирусологии, чтобы определить параметры генетической эволюции патогенных вирусов из геномных данных (главы 2, 6, 7). 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 нелеченных
2. Разработать и решить аналитически реалистичную модель эволюции многолокусных геномов
3. Предсказать влияние генетической рекомбинации на адаптацию многолокусных геномов
4. Разработать высокоточный метод определения адаптивного ландшафта из геномных данных.
5. Объяснить эволюционную выгоду латентного состояния ВИЧ для данного вируса
6. Разработать модель, объясняющую скорость прогрессирования ВИЧ-инфекции к СПИДу, и предсказать скорость прогрессии, как функцию параметров пациента
7. Выразить скорость эволюции антигенных участков вируса гриппа через иммунологические, эпидемиологические и вирусологические параметры
8. Предсказать, разовьют ли вирусы устойчивость к анти-вирусной терапии дефектными интерферирующими частицами.

Новизна. Впервые в вирусологии были использованы стохастические модели популяционной генетики с учетом естественного отбора, и эволюционные параметры ВИЧ оценены из геномных данных. Впервые биологически реалистичная модель многолокусной эволюции популяции бесполок организмов, учитывающая сильные эффекты генетического сцепления, решена аналитически и протестирована численно. Впервые показано, что редкая генетической рекомбинация в многолокусной модели ускоряет адаптацию во много раз. Эволюция вируса гриппа, вызванная иммунным ответом населения, впервые сведена аналитически к этим моделям многолокусной эволюции. Разработан первый высокоточный метод определения адаптивного ландшафта из геномных данных, на фоне сильного шума генетического сцепления. Предложено первое эволюционное преимущество латентности ВИЧ, подкрепленное моделью и данными.

Теоретическая значимость. Методы популяционной генетики введены в вирусологию. Создана новая научная область: мультилокусная популяционная генетика. Создан метод полудетерминистической бегущей волны для ее предсказания. Быстрая эволюция штаммов ВИЧ, устойчивых к лекарствам объяснена из наличия резистивных мутаций существующих еще до начала терапии. Установлена эволюционные причины прогрессии ВИЧ инфекции к СПИДу и латентности ВИЧ. Показано существование вероятности мутации оптимальной для адаптации.

Практическая значимость. Полученные оценки базовых эволюционных параметров ВИЧ, такие как размер популяции вируса, оказались полезными при разработке коктейлей вирусных ингибиторов. Разработан надежный метод оценки адаптивного ландшафта из геномных данных, применимый к любой адаптирующейся популяции. Показано что латентность ВИЧ это выгодный для вируса признак, что и объясняет провал всех терапий направленных на его устранение. Поставлен вопрос о функциональном излечении от ВИЧ. Предсказан, в общем виде, механизм и скорость эволюции респираторных вирусов, необходимый для разработки и обновления вакцин. Создан метод предсказания порядка мутаций иммунного ускользания, необходимого для

разработки консервативных эпитопов. Показано, что ВИЧ не может ускользнуть генетически от терапии при помощи дефектной интерферирующей частицы, если эпистаз не учитывается.

Положения, выносимые на защиту:

1. Эффективный размер популяции у среднего нелеченого ВИЧ-инфицированного пациента составляет не менее миллиона активно-инфицированных клеток, поэтому невыгодные мутации, обеспечивающие устойчивость к монотерапии, существуют в популяции вируса уже до терапии.

2. Высокая скорость эволюции ВИЧ в хронической инфекции обусловлены мутациями ускользания в иммунологически важных областях и мутациями, возникающими для их компенсации. Средний коэффициент отбора компенсирующей мутации меньше одного процента. Порядок мутаций ускользания во времени определяется их положением в плоскости «стоимость-выгода».

3. 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 в мышах.

4. Даже редкая рекомбинация ускоряет адаптацию на много порядков величины. В хронической нелеченой ВИЧ-инфекции вероятность рекомбинации примерно 1%. Наличие общей родословной у гомологичных локусов снижает эффективность рекомбинации из-за межгеномных корреляций. 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. Использование оригинального метода 3-локусной корреляции позволяет достичь отличной надежности обнаружения 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 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.

Соответствие паспорту специальности. Содержание диссертации соответствует паспорту специальности 1.5.8. Математическая биология, биоинформатика, в частности, пунктам: п. 1 – Математическое и компьютерное моделирование живых систем: биомолекул, ферментативных реакций, метаболических и сигнальных путей, субклеточных структур, клеток, тканей, органов, систем органов, организмов, популяций, биоценозов; п. 8 – Математическое и компьютерное моделирование эволюционных процессов в живой природе; п. 12 – Разработка и применение новых вычислительных алгоритмов для анализа экспериментальных данных в биологии и медицине.

Публикации. По теме диссертации опубликовано 36 работ, в том числе 34 статьи в рецензируемых журналах и 2 монографии. Большинство статей опубликовано в журналах первой четверти по SCOPUS.

Личный вклад. И.М. Рузин был единственным, первым, последним автором или автором-корреспондентом в 30 из 36 опубликованных работ по теме диссертации. В работах, где он был последним автором или автором-корреспондентом, он отвечал за постановку задачи, разработку моделей и методов, аналитические выводы, и редактирование рукописи. В работах где он был первым автором, он также производил компьютерное моделирование, сравнение с данными, визуализацию результатов, и основную работу с рукописью. В работах где он был вторым автором, он выполнял аналитическое или компьютерное моделирование.

Апробация. Результаты представлены на 40 международных конференциях, включая NORDITA Program “Unifying the epidemiological and evolutionary dynamics of pathogens” (Stockholm, Sweden, 2023); QLife Winter School “Quantitative Viral Dynamics Across Scales” (Paris, France, 2022); NORDITA program “From Molecular Basis to Predictability and Control of Evolution” (Stockholm, Sweden, 2019); 3d IFOM Symposium “Evolution, Resistance, and Cancer”, IFOM (Milan, Italy, 2018); 5th Interdisciplinary Symposium “Genomics and Microbial Communities”, Institut de Biologie Paris-Seine

(UPMC, Paris, 2015); Program on Virus Evolution in Kavli Institute of Theoretical Physics, University of California (Santa Barbara, CA, USA 2014); Workshop “From Within-Host Dynamics to the Epidemiology of Infectious Disease”, Institute of Mathematical Biosciences (Columbus, Ohio, 2014); Workshop on Microbial Evolution, Kavli Institute for Theoretical Physics, University of California at Santa Barbara (Santa Barbara, CA, USA, 2014); International 3rd Eastern European and Central Asia AIDS Conference (Moscow, Russia, 2009); Workshop on Mathematical Population Genetics (Edinburgh, UK, 2006); Meeting on HIV Dynamics and Evolution, University of California (San Diego, USA, 1998).

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.

Методы диссертации

Стандартные: Линейные и нелинейные обыкновенные 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. Его обобщение при наличии рекомбинации и с учетом филогенетики 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.

Основное содержание

Глава 1. Обзор литературы

В целях удобства, обзор литературы разбит на разделы и подразделы 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.

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 (Fisher 1990, Wright 1931, Kimura 1994): 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. Динамическое взаимодействие между теорией и экспериментом разрешает этот временной парадокс (Rouzine and Coffin 1999). Разработчик выбирает несколько основных экспериментальных наблюдений, которые он собирается объяснить или подгонять и подбирает простейшую модель для их интерпретации.

Настоящая глава представляет три работы, где данная стратегия использовалась для определения доминирующих эволюционных факторов, действующих на популяцию ВИЧ в не принимающих лекарств инфицированных людях 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 (Rouzine and Coffin 1999) проанализированы последовательности белка протеазы ВИЧ из инфицированных лиц, ранее не получавших ингибиторы протеазы (Lech, et al. 1996). Генетически разнообразные (полиморфные) основания интерпретированы как локусы, подвергающиеся медленной адаптации в течение хронической инфекции.

Анализ данных. Набор 265 последовательностей гена протеазы был получен в пациентах, ранее не принимавших ингибиторы протеазы (Lech, et al. 1996). Для каждого человека и каждого основания генетическое расстояние внутри человека (полиморфизм, генетическое разнообразие) рассчитывалась как средняя доля локусов различных в случайно выбранной паре геномов. Основные выводы из этих данных следующие: (i) Полиморфизм сосредоточен на редких основаниях: 47 из 297 оснований. (ii) Такой средний локус полиморфен у 16% индивидуумов, со средним полиморфизмом $\langle T \rangle = 0.27$ на полиморфный локус, при максимуме 0.5. (iii) Синонимичные и несинонимичные полиморфные основания имеют одинаковый средний полиморфизм, но первые встречаются в два раза чаще.

Детерминистическая адаптация вируса в хозяине. Объяснение этих фактов таково: если человек В заражен вирусом менее приспособленным к данному человеку, вирус постепенно адаптируется и возвращается к более приспособленному (рисунок 2). В процессе адаптации, локус имеет высокое генетическое разнообразие. Детерминистические уравнения для этой модели имеют вид

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

$$\frac{dn_2}{dt} = (1 - \mu_f)\kappa_2 n_2 + \mu_r \kappa_1 n_1 - \frac{n_2}{t_{rep}} \quad (2)$$

Здесь n_1 и n_2 — количество невыгодных и выгодных аллелей соответственно; κ_1 и κ_2 — их коэффициенты репликации; t_{rep} — время цикла репликации; μ_f и μ_r — вероятность прямой и обратной мутации соответственно. Цикл репликации $t_{rep} \approx 2$ дня (Haase 1999, Ho, et al. 1995, Perelson, et al. 1996, Wei, et al. 1995) (позже уточнено до 1 дня). Коэффициент отбора s — относительная разность приспособленности, $s = \kappa_2/\kappa_1 - 1$, которая удовлетворяет двойному неравенству $\mu_{f(r)} \ll s \ll 1$.

Время полуадаптации локуса с начальной невыгодной аллелью, t_{50} , определяется условием $n_1 = n_2$. Из уравнений выше, можно показать что $t_{50} = (t_{rep}/s)\log(s/\mu_r)$. В интервале времени $t = t_{50} \pm 2t_{rep}/s$, полиморфизм популяции высокий, $T > 0.015$. Оценивая среднее время выборки в

пациенте $\langle t \rangle t_{rep} \approx 5$ лет (Smith, et al. 1997), подгонка дает оценку $s = 0.006 - 0.008$ (Rouzine and Coffin 1999).

Частота появления локусов с начальным невыгодным аллелем у индивидуумов, 16%, объяснена наличием иммунного ответа CD8 Т-клеток (Kuroda, et al. 1999, Schmitz, et al. 1999), распознающих несколько десятков участков генома из 8-9 аминокислот (эпитопов) (Murphy 2011). Протеаза быстро накапливает антигенные ускользящие мутации в эпитопах STL, отменяя их распознавание вирусом, а клетки иммунной памяти гарантируют, что эти изменения станут постоянными.

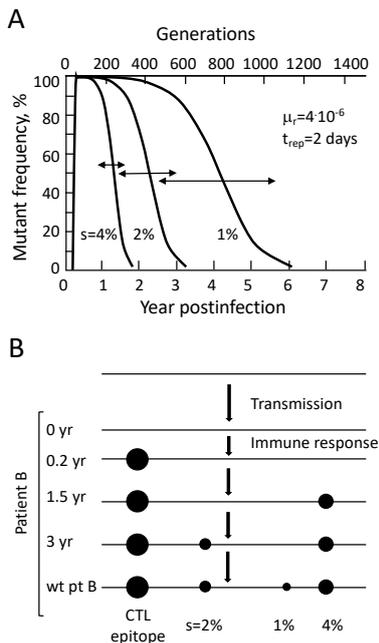


Рисунок 2. Модель эволюции ВИЧ, обусловленной иммунными ускользящими мутациями и компенсаторными мутациями. Быстрое начальное переключение доминантных эпитопов переопределяет выгодный аллель для ряда оснований, которые затем эволюционируют в сторону нового выгодного аллеля со скоростью, обратно пропорциональной 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 (Rouzine and Coffin 1999).

Эти фиксированные мутации снижают репродукцию вируса, что приводит к мутациям компенсации вне эпитопов, 10-20 мутаций на эпитоп, где адаптация происходит с разной скоростью (рисунок 2). Некоторые замены не меняют аминокислот, и для них естественный отбор действует на уровне мРНК или транскрипции. Расположение эпитопов CTL в геноме определяется индивидуальной иммуногенетикой, а именно шестью подтипами HLA, каждый из которых существует в сотне вариантов (Murphy 2011). Это объясняет, почему полиморфные основания отличаются между индивидуальными пациентами. Позже данная модель была подтверждена наблюдениями за антигенным ускользанием в эпитопах CTL в пациентах (Goonetilleke, et al. 2009, Batorsky, et al. 2014).

§ 2.2 Оценка эффективного размера популяции

В конце 1990х, было неясно, достаточно ли велика *эффективная* численность ВИЧ-популяции, чтобы содержать устойчивые к лекарственным средствам мутации еще до применения лекарства (§ 2.1). Положительный ответ, $N(\mu/s) \gg 1$, объяснил бы быстрое появление устойчивости ВИЧ к любой монотерапии (Coffin 1995). Рассмотренная ниже работа (Rouzine and Coffin 1999) ответила на этот вопрос.

Двухлокусная модель и тест неравновесия по сцеплению. Для оценки эффективного значения N у среднего нелеченого пациента из геномных данных, использовался следующий тест на основе двух-локусной модели включающей стохастические факторы: мутацию и случайный генетический дрейф. Строго говоря, эволюционную динамику можно предсказывать только в вероятностном смысле. При этом, важность стохастических факторов зависит от размера популяции и степени ее полиморфизма. Общим аналитическим методом является уравнение Колмогорова (Rouzine, et al. 2001). Простым способом, использованным в (Rouzine and Coffin 1999), является численное моделирование методом Монте-Карло.

На ранних стадиях инфекции, популяция ВИЧ обычно однородна генетически, так как вероятность передачи мала, и передается обычно один вирус (Holmes, et al. 1992, Zhang, et al. 1993). Некоторые локусы адаптируются к хозяину (Rouzine and Coffin 1999) (§ 2.1). Для двух таких локусов, все геномы в популяции классифицируются в четыре группы (гаплотипы): *ab*, *Ab*, *aB* и *AB*. Исходная и конечная популяция состоят только из наименее и наиболее приспособленного гаплотипа *ab* и *AB*, соответственно.

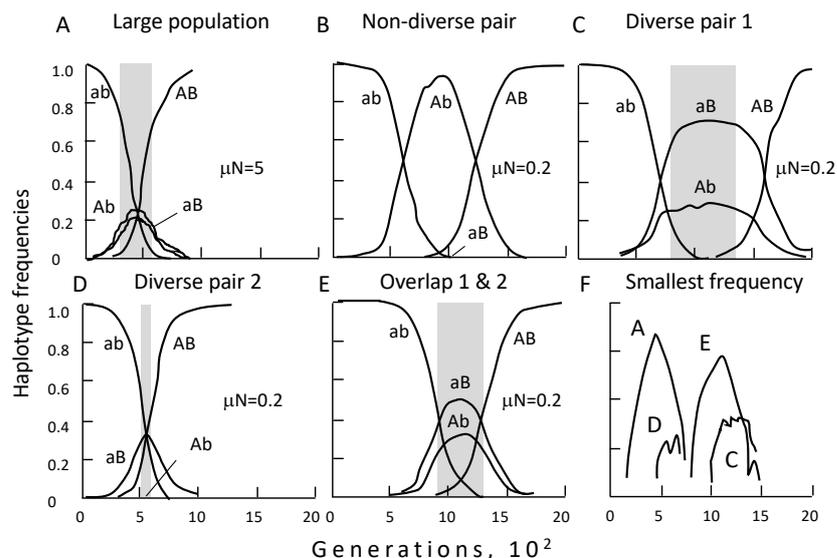


Рисунок 3. Компьютерное моделирование адаптации на двух участках в режиме отбор-дрейф, $1/s \ll N \ll 1/\mu$. А–Е. Частоты четырех гаплотипов. А и В обозначают первый и второй локус, соответственно, А и а обозначают дикий тип и мутант. Коэффициент отбора s одинаков для двух локусов. Значения параметров: 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 (A) и (C–E). По работе (Rouzine and Coffin 1999).

В детерминистическом пределе $N\mu \gg 1$, все четыре гаплотипа хорошо представлены в едином временном интервале (рисунок 3а). Однако если популяция мала, $N\mu \ll 1$, два локуса обычно адаптируются в разные случайные моменты времени, даже при одинаковых коэффициентах отбора. Почти одновременная адаптация может происходить, с небольшой вероятностью, по одному из редких сценариев (рис 3с-д), когда только три гаплотипа представлены. Зависимость от времени доли наименее представленного гаплотипа для прогонов Монте-Карло (рисунок 3е) усреднялась по 100 прогонам моделирования при разных N . Соответствующая экспериментальная оценка была получена путем объединения данных 11 пациентов по гену протеазы (Lech, et al. 1996) и оболочечного гена (Holmes, et al. 1992). Сравнение показывает, что наиболее вероятное значение μN больше, чем 7.0, что соответствует $N > 5 \cdot 10^5$ инфицированных клеток (рисунок 4). Было показано, что рекомбинация в типичном пациенте довольно редка и существенно не меняет результат.

Заключение. Эта работа показала, что популяция, $N_{\text{eff}} \sim 10^6$ зараженных клеток, достаточно велика чтобы содержать резистивные аллели на одном локусе, еще до начала лечения. Данный результат сильно отличается от предыдущей оценки $N_{\text{eff}} \sim 10^2 - 10^3$ (Leigh-Brown 1997), где был применен «тест нейтральности» Тажимы (Tajima 1989). Он был позже подтвержден другими методами другими группами (Frost, et al. 2000, Pennings, et al. 2014, Rouzine, et al. 2014).

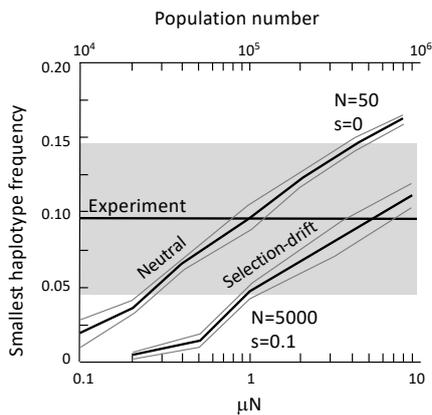


Рисунок 4. Предсказанная зависимость средней частоты наименее представленного гаплотипа от численности популяции. Переход от детерминистического $N \gg 1/\mu$ к стохастическому $N \ll 1/s$ режиму показан для двух значений коэффициента отбора: $s = 0.1$ и $s < 10^{-5}$. Для усреднения используются только прогоны моделирования, в которых генетический состав на каждом участке находится в интервале 25–75% во временном интервале, 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 $\mu = 10^{-5}$ (Mansky and Temin 1995). Толстая горизонтальная линия и заштрихованная полоса представляют собой экспериментальное среднее значение и область достоверности 95%, полученную из данных (Lech, et al. 1996, Holmes, et al. 1992). По работе (Rouzine and Coffin 1999).

§ 2.3 Оценка вероятности рекомбинации ВИЧ

Целью следующей работы (Batorsky, et al. 2011) была оценка эффективной вероятности рекомбинации генома ВИЧ с другим геномом, у нелеченых пациентов.

Модель. Модель, реализованная в виде стандартного алгоритма Райта-Фишера с дискретными поколениями (Batorsky, et al. 2011), представляет собой обобщение двухлокусной модели (§ 2.2) на произвольное большое количество локусов L и присутствие рекомбинации. Рассматривается популяция из N бинарных последовательностей $[K_i]$, например, 0111010101, где каждый локус с номером $i = 1, 2, \dots, L$ имеет два возможных варианта: $K_i = 0$ или $K_i = 1$. Число потомков генома является случайным, а среднее значение определяется относительной приспособленностью $\exp(W)$ которая зависит от количества полезных аллелей, где

$$W = \sum_{i=1}^L s_i K_i \quad (3)$$

Значения s_i выбираются в соответствии с заданным случайным распределением, и их значения фиксируются до начала прогонов моделирования. Перед случайной выборкой потомства, из популяции случайным образом отбирают rN пар геномов. 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.

Меры корреляции аллелей. Используются два следующих коррелятора между аллелями на парах локусов. Отобраны «очень полиморфные» пары локусов, где аллельная доля f , для обоих локусов, находится в промежутке $0.25 < f < 0.75$. Первый коррелятор определен как $1 - \langle f_{AB}^{LRH} / (f_A f_B) \rangle$, где f_{AB}^{LRH} — частота наименее представленных гаплотипов для локусов А и В (§ 2.2), $f_A f_B$ — произведение частот соответствующих аллелей, а $\langle \dots \rangle$ означает усреднение по «очень полиморфным» парам. Второй коррелятор определен как доля таких пар, для которых $f_{AB}^{LRH} < 0.04$.

У 4 пациентов, не получавших лечения, были отобраны девять наборов последовательностей генов ВИЧ-полимеразы, в 2-3 временных точках, в среднем 25 геномов на набор, с длиной от 1200 до 1250 нуклеотидов. После сравнения с консенсусом полимеразы (1100 оснований), каждая последовательность была преобразована в бинарную строку.

Оценка скорости рекомбинации и среднего коэффициента отбора. Используя подогнанное значение s_0 (ниже), две смоделированных коррелятора, построенные как функция скорости рекомбинации r , были сопоставлены с диапазоном корреляторов в данных пациентов (рисунок 5А и В). Интервал r , который соответствует данным, составляет от 0.05 до 0.15 с центром в $r = 0.01$. Приведенные выше оценки сделаны для $M = 10$ (Levy, et al. 2004). Уменьшение M в 3 раза увеличивает конечную оценку r в 2–3 раза. Оценка $s_0 = 0.003-0.005$ была сделана двумя методами (рисунок 5С).

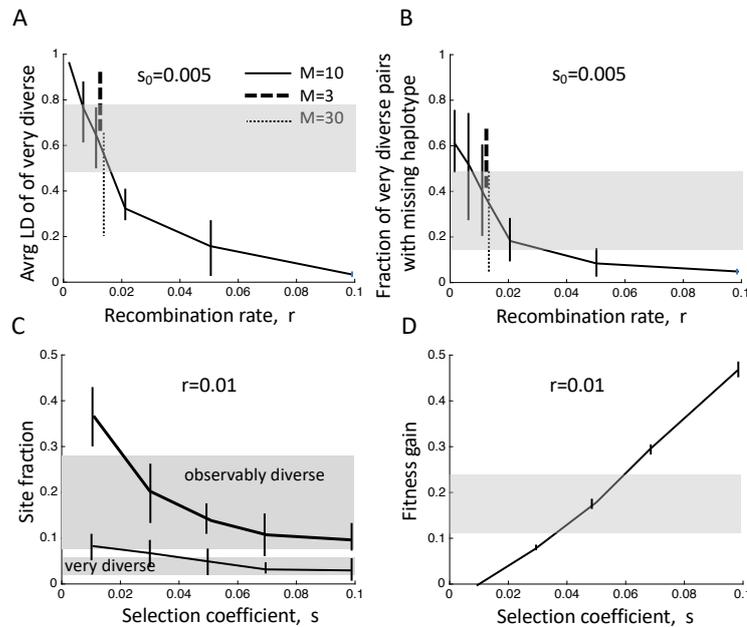


Рисунок 5. Оценка параметров модели хронической ВИЧ-инфекции. Скорость рекомбинации r и средний коэффициент отбора s_0 оцениваются путем сравнения четырех наблюдаемых величин в данных с их предсказанными значениями при моделировании. Серые области показывают среднее \pm одно стандартное отклонение величин, рассчитанных по данным гена *pol* у 4 пациентов. 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) Коррелятор $1 - \langle f_{AB}^{LRH} / f_A f_B \rangle$ для $s_0 = 0.005$. (В) 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 $r = 0.01$. Параметры $s_0 = 0.005$, $M = 10$, $L = 2000$, $N = 10^5$, $\mu = 3 \cdot 10^{-5}$. По работе (Batorsky, et al. 2011).

Закключение. Таким образом, вероятность рекомбинации у среднего нелеченного ВИЧ-позитивного человека составляет $r \approx 1\%$. Средняя скорость адаптации меньше, чем предсказывает однолокусная модель, в 4 раза (Batorsky, et al. 2011). Тем не менее, даже при столь редкой

рекомбинации, адаптация ускорена в 2 раза. Похожие оценки для этих двух параметров, $r \approx 0.02$ и $s_0 \approx 0.005$, были получены другой группой другим методом (Neher and Leitner 2010).

Глава 3. Мультилокусные модели бесполок популяций

Реальные популяции имеют большое количество полиморфных локусов, которые эволюционируют одновременно. Эволюция этих локусов, вообще говоря, не является независимой, так как они имеют общих предков, если только не разделены кроссоверами рекомбинации. Модель с одним локусом (§ 2.1), вообще говоря, не применима из-за генетического сцепления: генетическая информация на локусах, не разделенных рекомбинацией, передается от родителя к потомству вместе (Fisher 1930, Muller 1932, Hill and Robertson 1966). Примером последствия генетического сцепления является «клональная интерференция» (Fisher 1930, Muller 1932): растущий клон с выгодной мутацией на одном локусе вытесняет из популяции клоны с более слабыми мутациями. Задачей большого практического значения являлось предсказание скорости многолокусной эволюции через заданные постоянные параметры системы. До работ, представленных ниже, эта задача была решена только для двух или трех локусов. Ниже скорость эволюции выражена через параметры модели для произвольного количества локусов L для однородного коэффициента отбора (Rouzine, et al. 2003, Rouzine, et al. 2008) (§ 3.1) и коэффициента отбора, который изменяется между локусами (Good, et al. 2012) (§ 3.2).

§ 3.1 Многолокусная эволюция с фиксированным коэффициентом отбора

Строгое описание этой системы требует решения 2^L сцепленных уравнений Колмогорова, что невозможно практически. Для преодоления этой трудности было предложено классифицировать геномы по их приспособленности и применить детерминистическую динамику к этим классам приспособленности (Tsimring, et al. 1996). Асимптотически точная реализация этой идеи в полустохастическом варианте была получена в (Rouzine, et al. 2003, Rouzine, et al. 2008).

Модель и метод. При смене поколений более приспособленный аллель может мутировать в менее приспособленный с вероятностью $\mu \ll 1$, и наоборот. Мутация меняет приспособленность на s или $-s$. Последовательности классифицируются по числу невыгодных аллелей k (рисунок 6А).

Главное наблюдение состоит в том, что почти любой фитнес-класс достаточно велик по размеру, чтобы его можно было рассматривать детерминистически (рисунок 6В). Только наиболее приспособленный класс, $k = k_0$, должен рассматриваться стохастически. Точность такого подхода была проверена численно, с помощью моделирования Монте Карло методом описанным в § 2.3 с $r = 0$. Результат—это волна плотности фитнес-классов, бегущая к более высокой или более низкой приспособленности, в зависимости от начального состояния и параметров модели: N , s , L и μL . Рано или поздно волна останавливается, и система приходит в равновесие (Rouzine, et al. 2003, Rouzine, et al. 2008).

Детерминистическое уравнение для фитнес-классов: бегущая волна. Далее, эти результаты получены в общей аналитической форме (Rouzine, et al. 2003, Rouzine, et al. 2008). Доля класса с k вредных аллелями в популяции $f_k(t)$ описывается конечно-разностным уравнением вида

$$f_k(t + 1) - f_k(t) = U_d f_{k-1}(t) + U_b f_{k+1}(t) - \{U_d + U_b + s(k - k_{av})\} f_k(t) \quad (4)$$

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, $U_d \equiv \mu(L - k_{av})$, $U_b \equiv \mu k_{av}$ соответственно, и $k_{av}(t) = \sum_k k f_k(t)$ есть среднее число аллелей делающее уравнение нелинейным по $f_k(t)$. По определению, $\sum_k f_k(t) \equiv 1$. При малом коэффициенте отбора $s \ll 1$, уравнение (4) имеет решение вида бегущей волны

$$\log f(k, t) \equiv \phi(x), x = k - k_{av}(t) \quad (5)$$

$$\sigma x = (1 - \alpha) e^{-\phi'(x)} + \alpha e^{\phi'(x)} + v \phi'(x) - 1 \quad (6)$$

Здесь $v \equiv \frac{(U_b + U_d)^{-1} d k_{av}(t)}{dt}$ — перенормированная скорость движения волны в сторону ухудшения приспособленности, которая соответствует суммарной частоте накопления геномных мутаций

(вредные минус полезные), $\sigma \equiv s/(U_b + U_d)$, $\alpha \equiv k_{av}/L$. Скорость v может быть как отрицательной (адаптация), так и положительной (накопление вредных мутаций). Из формулы (6) выводятся все важные свойства волны. Вблизи пика волны, $|x| \ll 1/\sigma$, можно написать

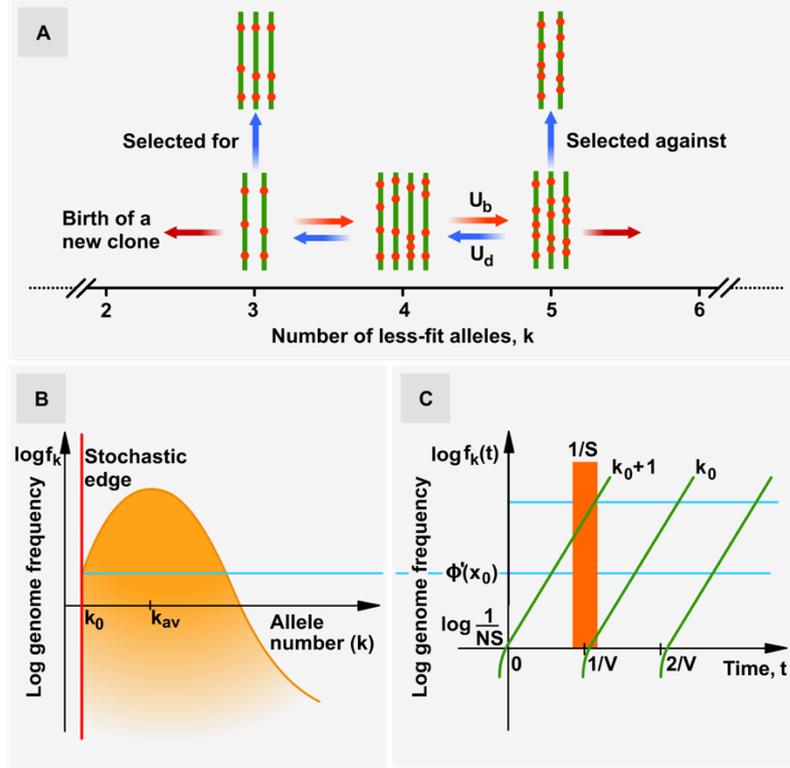


Рисунок 6. Адаптация большого числа сцепленных локусов в бесполой популяции представляет собой бегущую волну приспособленности, скорость которой определяется стохастической динамикой наиболее приспособленных индивидуумов. (А) Последовательности (зеленые линии) классифицируются по дискретным классам приспособленности 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) Динамика доли класса f_k определяется из детерминистического уравнения (4). 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 (Rouzine, et al. 2003).

$$\phi(x) \approx -\frac{x^2}{2 \text{Var}[x]} - \frac{1}{2} \log(2\pi \text{Var}[x]), \quad \text{Var}[x] = \frac{1-2\alpha-v}{\sigma} > 0 \quad (7)$$

Волна имеет край при $x = x_0$

$$x_0 = -\frac{1}{\sigma} (1 - 2\alpha u - v \ln u - v) < 0, \quad u \equiv e^{\phi'(x_0)} = \frac{1}{2\alpha} \left[-v + \sqrt{v^2 + 4\alpha(1-\alpha)} \right] \quad (8)$$

за которым нет геномов: все более приспособленные классы, $x < x_0$, пусты. Перепад высоты волны между пиком и краем

$$\phi(0) - \phi(x_0) = \frac{1}{\sigma} \left\{ 1 - 2\alpha - \frac{v}{2} [\ln^2(eu) + 1] - 2\alpha u \log u \right\} \quad (9)$$

Вероятностный подход к наиболее приспособленным. Чтобы выбрать правильную скорость v , надо знать граничное условие $\phi(x_0)$. Для этого рассмотрена динамика наиболее приспособленного класса, $k = k_0$ в рамках однолокусной двухаллельной модели, описываемой уравнением Колмогорова (Rouzine, et al. 2001). Лучший класс рассматривается, как малочисленная выгодная аллель с эффективным коэффициентом отбора $S \equiv U(\sigma|x_0| - 1)$ и эффективная частотой мутаций $M(t) \equiv U\alpha f_{k_0+1}(t)$. Размер наиболее приспособленного класса зависит от времени пилообразным образом (рисунок 6С).

В режиме адаптации, $S > 0$, который реализуется при низкой начальной приспособленности, наиболее приспособленный класс растет до тех пор, пока полезная мутация внутри него не приведет к рождению нового наиболее приспособленного класса. Динамика лучшего

класса с основной частью волны согласуется как

$$\phi(x_0) = \frac{1}{2} \left[\log \frac{1}{N|s|} + \log f_{max} \right] \quad (10)$$

где f_{max} — максимальная частота лучшего класса перед рождением следующего (рисунок 6С). Выражение $f_{max} = 1/(NU_b)$ для f_{max} в (10) находится из условия, что лучший класс приобретает за один зубец «пилы» (рисунок 6С). Поправка к данному приближению из-за дискретности k получена в работе (Rouzine, et al. 2008). Окончательный результат для средней частоты замен, когда популяция далека от стационарного состояния, имеет вид (Rouzine, et al. 2008)

$$V = -(U_b + U_d)v \approx \frac{2s \log(N\sqrt{sU_b})}{\log^2 \left[\frac{s}{U_b} \log(N\sqrt{sU_b}) \right]} \gg U_b + U_d \quad (11)$$

Когда популяция начинает с максимальной приспособленности, $S < 0$, средняя приспособленность будет снижаться. Полагая $\alpha = 0$ и делая аналогичное стохастическое рассмотрение получается (Rouzine, et al. 2008)

$$\sigma \log(NU_d\sigma^2) \approx \left[1 - \frac{v}{2} \left(\log^2 \frac{e}{v} + 1 \right) \right] - \sigma \log \left[\sqrt{\frac{v^3}{1-v}} \frac{\log(e/v)}{1-v \log(e/v) + 5\sigma/6} \right] \quad (12)$$

Выше предполагалось, что популяция по приспособленности далека от 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 (Rouzine, et al. 2003)

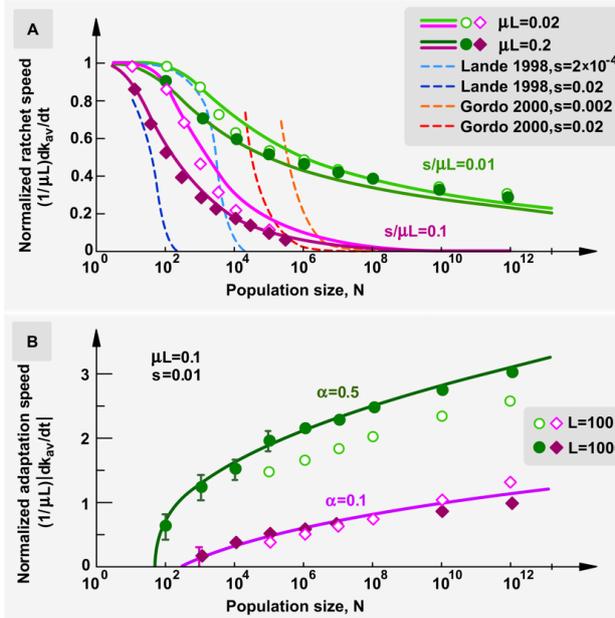


Рисунок 7. Сравнение аналитических результатов для частоты замен с результатами Монте Карло. (А) Нормализованная скорость хrapовика как функция размера популяции N : результаты анализа (сплошная линия) в сравнении с численного моделированием (символы). Полезные мутации отсутствуют, $\alpha = 0$. Фиолетовый: результаты для $\sigma = s/U_d = 0.1$, формула (12). Штриховые линии — асимптотика больших N из работы (Gordo and Charlesworth 2000) (формулы (3a) и (3b) в ней), и асимптотика малых N из работы (Lande 1998) (выражение (2c) умноженное на NU_d). Зеленый: результаты для $\sigma = s/U_d = 0.01$. Параметры показаны. (В) Нормализованная скорость адаптации как функция N из данных аналитических результатов, формула 13 (сплошная линия), и численного 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 (Rouzine, et al. 2003).

$$\sigma \log \left(\frac{N}{N^*} \right) \approx 1 - 2\alpha - v - \frac{v}{2} \log^2 u - v \log u - 2\alpha u \log u, v < 1 - 2\alpha \quad (13)$$

где префактор N^* является степенной функцией $Var[k]$, U , α и v [(Rouzine, et al. 2003), приложение, формулы (19)-(21)]. Лидер фитнес-волны x_0 определяется формулой (8). В конце концов, популяция достигает устойчивого состояния при значении α , где процессы адаптации и хrapовика уравнивают друг друга. Аналитические результаты подтверждаются численной симуляцией методом из § 2.3 с $r = 0$ (рисунок 7).

Закключение. Скорость адаптации гораздо меньше предсказания одно-локусной модели (§2.1) и очень медленно растет с численностью популяции N . Переход к режиму описываемому одно-локусной моделью происходит при размерах популяции которые, для умеренно большого числа локусов ($L > 20$), нереально велики. Последующие исследования подтвердили точность результатов (Desai and Fisher 2007, Brunet, et al. 2008, Goyal, et al. 2012).

§ 3.2 Многолокусная эволюция с распределенным коэффициентом отбора

В предыдущем параграфе предполагалось, что все мутации имеют одинаковый эффект на приспособленность вируса, s . В реальных организмах, значение s меняется между локусами. Заранее не очевидно, усредняются ли локусы до некоего эффективного значения s , и какие мутации являются наиболее важными, с большими или малыми значениями s . Работа (Good, et al. 2012), представленная ниже, выполнена в сотрудничестве с тремя теоретическими группами. Все сотрудники получили одни и те же результаты параллельно четырьмя независимыми методами, из них три аналитические и один численный.

Модель. Предполагается, что коэффициенты отбора выбраны из случайного распределения с плотностью $\rho(s)$. Вредные мутации отсутствуют. (Good, et al. 2012) рассмотрели случай общего распределения $\rho(s)$, но ниже рассматривается пример экспоненциального распределения

$$\rho(s) = \frac{1}{\sigma} e^{-\frac{s}{\sigma}}, \quad \sigma \gg U_b \quad (14)$$

Как показано выше в § 3.1, популяция развивается в уединенную волну, которая движется в координате приспособленности X со «скоростью адаптации» $v \equiv d\bar{X}(t)/dt$ и плотностью классов

$$f(x) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{x^2}{2v}} \quad (15)$$

где перемасштабированы обозначения $-\sigma v \rightarrow v$, $\sigma x \rightarrow x$, $x = X - \bar{X}(t)$ — относительная приспособленность индивидуума по отношению к среднему.

Вероятность установления генетической линии. Предположим, некий геном содержит полезный аллель. Чтобы не прерваться, генетическая линия этого аллеля должна претерпеть множественные выгодные мутации для увеличения приспособленности на фоне конкуренции с постоянно увеличивающейся средней приспособленностью $\bar{X}(t)$. Вероятность невымирания $w(x)$ удовлетворяет уравнению (Good, et al. 2012)

$$v \frac{dw}{dx} = xw(x) - w(x)^2 + U_b \int_0^{\infty} ds \rho(s)[w(x+s) - w(x)] \quad (16)$$

Условие самосогласования скорости эволюции. Вероятность фиксации аллели с коэффициентом отбора s , обозначаемая $\pi(s)$, получена путем усреднения по распределению приспособленности генома $f(x)$ в котором этот аллель появился

$$\pi(s) = \int_{-\infty}^{\infty} dx w(x) f(x-s) \quad (17)$$

Скорость увеличения приспособленности v равно средней частоте фиксации новых мутаций, взвешенной по их эффекту приспособленности

$$v = NU_b \int_{-\infty}^{\infty} ds s \rho_f(s), \quad \rho_f(s) = \text{const } \pi(s) \rho(s) \quad (18)$$

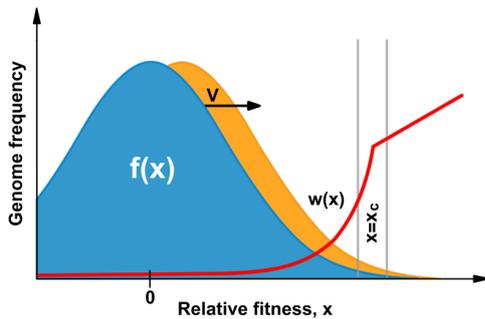


Рисунок 8. Процесс адаптации (схема). Плотность распределение геномов по приспособленности $f(x)$ движется с постоянной скоростью v . Вероятность фиксации $w(x)$, формула (19), резко возрастает с увеличением x , пока не достигает тонкого пограничного слоя $x = x_c$, после чего переходит к стандартному результату теории одного локуса (глава 2), $w(x) = x$. По работе (Good, et al. 2012).

Вероятность фиксации и скорость адаптации. Когда размеры популяции велики, а мутации редки, решение (2.26) для вероятности фиксации $w(x)$ претерпевает резкое изменение при пороговой приспособленности

$x = x_c$, ниже которого оно экспоненциально затухает

$$w(x) \approx x_c e^{-\frac{x^2 - x_c^2}{2v}}, \quad 0 < x < x_c \quad w(x) \approx x, \quad x > x_c \quad (19)$$

где x_c определяется дополнительным условием найденным в (Good, et al. 2012). Лидер x_c аналогичный x_0 в § 3.1, есть граница приспособленности, выше которой клональная интерференция не снижает вероятность фиксации (рисунок 8). Предполагая экспоненциальное распределение эффектов приспособленности (14), и неравенство $NU_b \gg 1$, все интегралы по s в (16) и (18) имеют острый пик в точке

$$s = s^* = x_c - \frac{v}{\sigma} \quad (20)$$

что дает искомый эффективный коэффициент отбора. Это приводит к двум связанным уравнениям для v и x_c (Good, et al. 2012). Их асимптотические решения получены в следующих двух важных пределах.

Если размер популяции не слишком большой, волна узкая, и доминантные мутации представляют собой большие скачки приспособленности сравнимые с лидером волны, $s^* \sim x_c$. В этом случае, окончательный ответ для скорости адаптации имеет вид (Good, et al. 2012)

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

Здесь логарифмические коэффициенты внутри больших логарифмов, как всегда, опущены.

В случае больших NU_b , волна широкая, и адаптация происходит путем сравнительно небольших скачков, $s^* \ll x_c$, что эквивалентно многолокусной модели с фиксированным s в § 3.1, где s и U_b заменяются эффективными значениями

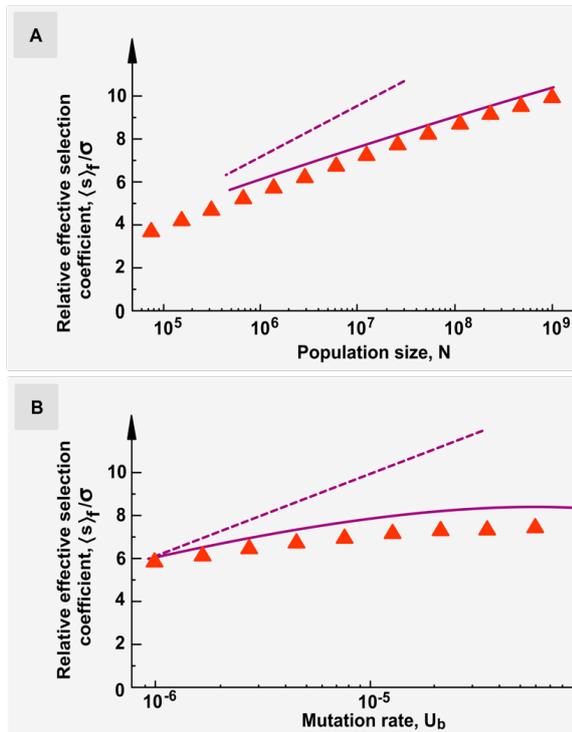


Рисунок 9. Средний эффект приспособленности фиксированной мутации s^* , в зависимости от параметров модели. Использована формула (20). (А) s^* как функция размера популяции N и (В) s^* как частоты полезных мутаций U_b . Оранжевые треугольники и фиолетовая сплошная кривая показывают результаты моделирования и анализа, соответственно. Фиолетовая пунктирная линия: результаты интерференционной модели с двумя клонами (Gerrish and Lenski 1998). Параметры по умолчанию $N = 10^7, U_b = 10^{-5}, \sigma = 0.01$. По работе (Good, et al. 2012).

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

$$\log(NU_b) \gg 2 \log\left(\frac{\sigma}{U_b}\right)$$

Закключение. Для более быстрого затухания распределения s , чем экспоненциальное (или для экспоненциального но в большой популяции), важны множественные локусы с малым эффектом мутации, и результат сводится к результатам в § 3.1, с некоторым эффективным значением s . Если же спад хвоста происходит медленнее экспоненциального, режим сводится к более

простой модели интерференции двух выгодных мутаций за раз (Gerrish and Lenski 1998).

Глава 4. Многолокусная эволюция частично половой популяции

Главным отличием половых организмов от бесполой является наличие рекомбинации. Во время спаривания, генетическая информация объединяется из двух родительских ДНК, по половине от каждого, и передается потомству. В пяти работах на которых основана данная глава (Rouzine and Coffin 2005, Gheorghiu-Svirshchevski, et al. 2007, Rouzine and Coffin 2007, Rouzine and Coffin 2010, Rouzine 2023), предполагается, что начальная популяция содержит небольшую долю f_0 полезных аллелей на рассматриваемых локусах, и новой мутацией можно пренебречь. Так бывает, если окружающая среда претерпела недавние изменения. Целью является выражение скорости адаптации через параметры системы и изучение эффектов генетического сцепления. В § 4.1, этот результат получен аналитически, пренебрегая корреляцией геномов из-за филогенетического родства (Rouzine and Coffin 2005). В § 4.2, эффекты филогенетического родства учтены в рамках модели Монте Карло из § 2.3 (Rouzine 2023). На очень больших временах необходим учет новых мутаций, что изучено другими авторами (Neher, et al. 2013).

§ 4.1 Эволюция с рекомбинацией и начальным генетическим разнообразием

С вероятностью r , геном имеет рекомбинацию с другим, случайно выбранным геномом, и рекомбинант заменяет одного из родителей (модель описанная в § 2.3). Главное предположение в данном разделе заключается в том, что, при заданном числе k невыгодных аллелей в геноме, они распределяются случайным образом среди L локусов, и что их расположение между разными геномами никак не коррелирует. Как показано в § 4.2, это предположение справедливо, если $5 \ln(Nr f_0) \gg \sqrt{L}$.

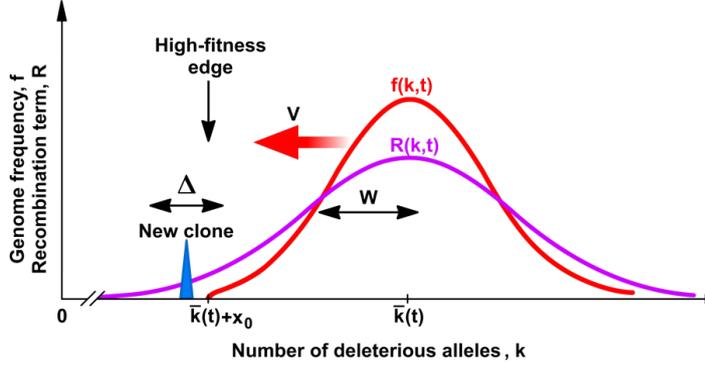


Рисунок 10. Схема движущейся уединенной волны. Красные и фиолетовые кривые: и тонкие линии, частота класса приспособленности, $f(k, t)$ и функция рекомбинационного генератора $R(k, t)$ соответственно. Синий пик: новый рекомбинантный клон, созданный за краем волны; Δ — интервал, в котором образуется наибольшее количество таких клонов; w и V — ширина и скорость (средняя частота замен) волны соответственно. По работе (Rouzine and Coffin 2005).

Пусть $f(k, t)$ — средняя доля геномов с k вредными аллелями с эффектом s . Динамическое уравнение имеет вид (Rouzine and Coffin 2005)

$$\partial f / \partial t = -s[k - \bar{k}(t)]f(k, t) + r[R(k, t) - f(k, t)], \quad \bar{k}(t) \equiv \int f(k, t)k dk \quad (21)$$

$$R(k, t) = \frac{1}{\sqrt{\pi k}} \int dk_1 \int dk_2 f(k_1, t)f(k_2, t)e^{-[k - \frac{k_1 + k_2}{2}]^2 / [\bar{k}(1 - \bar{k}/L)]} \quad (22)$$

Распределение приспособленности $f(k, t)$ и функция генерации рекомбинантов $R(k, t)$ показаны схематически на рисунок 10. Уравнения (21) и (22) имеют решение вида бегущей волны

$$f(k, t) = \phi[k - \bar{k}(t)], \quad R(k, t) = \rho[k - \bar{k}(t)] \quad (23)$$

Введем новые обозначения: относительное число аллелей $x \equiv k - \bar{k}(t)$, средняя частота замен выгодных мутаций $V \equiv -d\bar{k}/dt$, $b \equiv \frac{r}{s}$, $w^2 \equiv \frac{V}{s}$. Если общее число новых рекомбинантов в поколении велико, $Nr \gg 1$, то решение принимает асимптотически точный вид (Rouzine and Coffin 2005)

$$\phi(x) = \frac{1}{w\sqrt{2\pi}} e^{-\frac{(x+b)^2}{(2w^2)}}, \quad w^2 < \bar{k}(1 - \bar{k}/L), \quad x_0 < x < |x_0| \quad (24)$$

$$\rho(x) = \frac{1}{\sqrt{\pi(\bar{k}(1 - \bar{k}/L) + w^2)}} e^{-\frac{(x+b)^2}{\bar{k}(1 - \bar{k}/L) + w^2}} \quad (25)$$

$$x_0^2 \approx \bar{k}(1 - \bar{k}/L) \frac{2p(1+p)}{1-p} \log \left[\frac{s\sqrt{\bar{k}(1 - \bar{k}/L)(1-p)}}{r} \right], \quad r \ll s\sqrt{\bar{k}(1 - \bar{k}/L)(1-p)} \quad (26)$$

где лидер $x = x_0$ соответствует самому приспособленному геному, а p — скорость волны, нормализованная на свое значение в бесконечной популяции ($f \ll 1$, глава 2)

$$p \equiv \frac{V}{s\bar{k}(1 - \bar{k}/L)} = \frac{w^2}{\bar{k}(1 - \bar{k}/L)}.$$

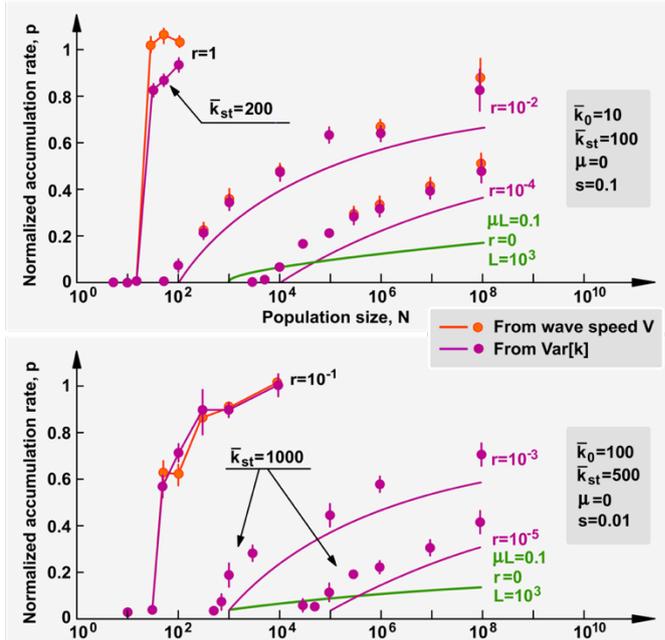
Стохастический край и скорость адаптации. Чтобы получить второе уравнение для $|x_0|$ и V , изучена стохастическая динамика фронта волны (рисунок 10). Миноритарный аллель — это новый клон, формирующийся вблизи края волны, который имеет эффективный коэффициент отбора $S = s|x_0|$. Вычисление средней скорости продвижения фронта по x и его сравнение со скоростью основной части волны дает второе уравнение для $|x_0|$

$$x_0^2 \approx \bar{k}(1 - \bar{k}/L)(1 + p) \log \frac{Nr}{p} \quad (27)$$

Из уравнений (26) и (27), окончательный результат для частоты замен V имеет вид

$$V = s\bar{k} \left(1 - \frac{\bar{k}}{L}\right) p, \quad p = \frac{\ln(Nr)}{\ln(Nr) + \ln\left(\frac{s^2\bar{k}(1 - \bar{k}/L)}{r^2}\right)}, \quad \frac{1}{N} \ll r \ll s\sqrt{\bar{k}(1 - \bar{k}/L)} \quad (28)$$

График этого результата как функции Nr показан на рисунок 11. Заметим, что в точке $Nr \sim 1$, адаптация полностью останавливается, а в пределе $\ln(Nr) \gg \ln\left(\frac{s^2 \bar{k}(1-\bar{k}/L)}{r^2}\right)$, скорость адаптации достигает предела независимых локусов, $V = s\bar{k}$. Эти результаты подтверждаются численно методом Монте Карло (рисунок 11). Алгоритм использовал то же приближение некоррелированных геномов, что и аналитическая часть.



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 (28). 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 (Rouzine and Coffin 2005).

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, за счет новых мутаций. В отличие от бесполой эволюции, скорость адаптации достигает половины значения полученного в одно-локусной модели при весьма умеренном размере популяции, $N \sim [\bar{k}(1 - \bar{k}/L)/r](s/r)^2$, что и показывает эволюционное преимущество полового размножения.

Рисунок 11. 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 $(d\bar{k}/dt)/[s\bar{k}(1 - \bar{k}/L)]$; Фиолетовые кружки:

§ 4.2 Эффекты общей родословной на эволюцию с рекомбинацией

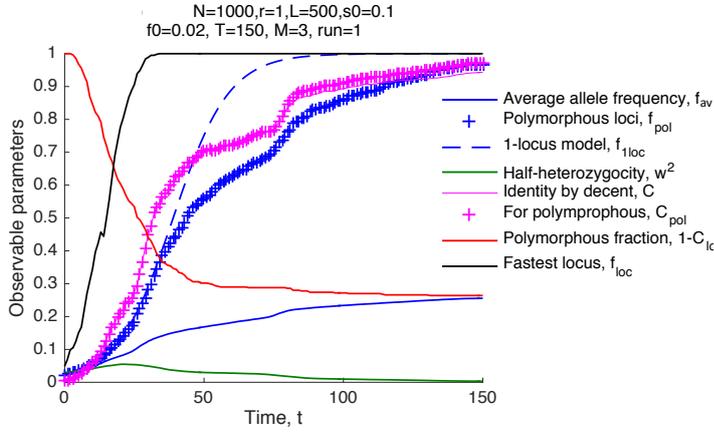
Предварительные работы (Gheorghiu-Svirshchevski, et al. 2007, Rouzine and Coffin 2007, Rouzine and Coffin 2010) показали, что межгеномные корреляции между гомологичными локусами нарастают во времени из-за общих предков геномов. Эти филогенетические эффекты снижают эффективность рекомбинации, вызывают исчезновение полезных аллелей на части локусов, а также приводит к сильным эффектам генетического сцепления. В работе обсуждаемой ниже (Rouzine 2023), та же модель изучена численно с учетом филогенетических эффектов. Получено условие применимости приближения некоррелированных геномов использованного § 4.1.

Модель Монте Карло. 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. Один из двух родителей заменяется рекомбинантом. Вычисления выполнены для $r = 1$, что соответствует обычному половому размножению, но все результаты легко обобщаются на случай любого r с помощью скейлинга во всех безразмерных выражениях: $(N, s) \rightarrow (Nr, s/r)$ (Rouzine and Coffin 2010, Neher, et al. 2010).

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 из-за эффекта клональной интерференции (рисунок 12, красная кривая). Доля локусов с невымершими аллелями, зависит в основном от одного составного параметра (рисунок 13А-С): $1 - C_{loss}(\infty) =$

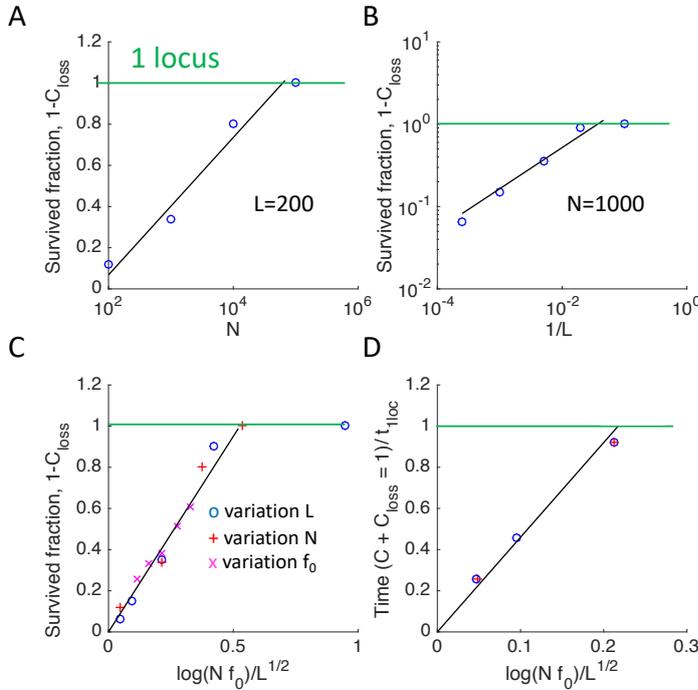
$2.0 \frac{\log(Nrf_0)}{\sqrt{L}}$, если $\log(Nrf_0) < 0.5\sqrt{L}$. Выше этой критической точки, $\log(Nrf_0) > 0.5\sqrt{L}$, получается $1 - C_{loss}(\infty) \approx 1$, то есть, вымиранием аллелей можно пренебречь.

Рисунок 12. 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 локусам, 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 локусов с общим исходным предком C , одинаковое значение для полиморфных локусов C_{pol} , доля полиморфных локусов $1 - C_{loss}$ и наибольшая из частот аллелей среди локусов $\max(f_{loc})$. Значения параметров показаны. По работе (Rouzine 2023).



и того же предка в момент времени t (Rouzine and Coffin 2010). 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) $C(T_2) \approx C_{loss}(T_2) \approx 0.5$ (рисунок 12). Время T_2 представляет приблизительную временную шкалу филогенетического дерева. Зависимость T_2 от параметров модели хорошо интерполируется формулой (рисунок 13D) $T_2 \approx t_{1loc} 5.0 \log(Nrf_0)/\sqrt{L}$, где $t_{1loc} = (1/s) \log(1/f_0)$ есть время полуадаптации в детерминистической однолокусной модели (§ 2.1).

Рисунок 13. Наблюдаемые зависят от одного составного параметра. А-С. 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 12] 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). Чувствительность к изменению коэффициента отбора s и числа кроссинговеров M мала (Rouzine 2023). Значения параметров по умолчанию: $N = 1000, L = 200, f_0 = 0.02$, если не указано иное. Остальные параметры как на рисунке 12. По работе (Rouzine 2023).

Когда отношение времен T_2/t_{1loc} велико, что соответствует неравенству $\ln(Nrf_0) \gg 0.2\sqrt{L}$, 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.

Сильная вариация скорости адаптации среди локусов. Показано, что адаптация полиморфных локусов происходит с очень разной скоростью (Rouzine 2023). В некоторых локусах

аллели накапливаются гораздо медленнее, а в некоторых гораздо быстрее, чем предсказывает однолокусная модель (рисунок 12, черная кривая). Сверхбыстрая эволюция некоторых локусов происходит из-за того, что сегменты генома, содержащие эти локусы, имеют необычно большое количество полезных аллелей (эффект генетического фона). Популяция имеет сложную родословную структуру, которая варьируется в зависимости от локуса. Размеры генетических линий растут со временем, с разной скоростью для разных локусов, и распределены в очень широком диапазоне (Rouzine 2023).

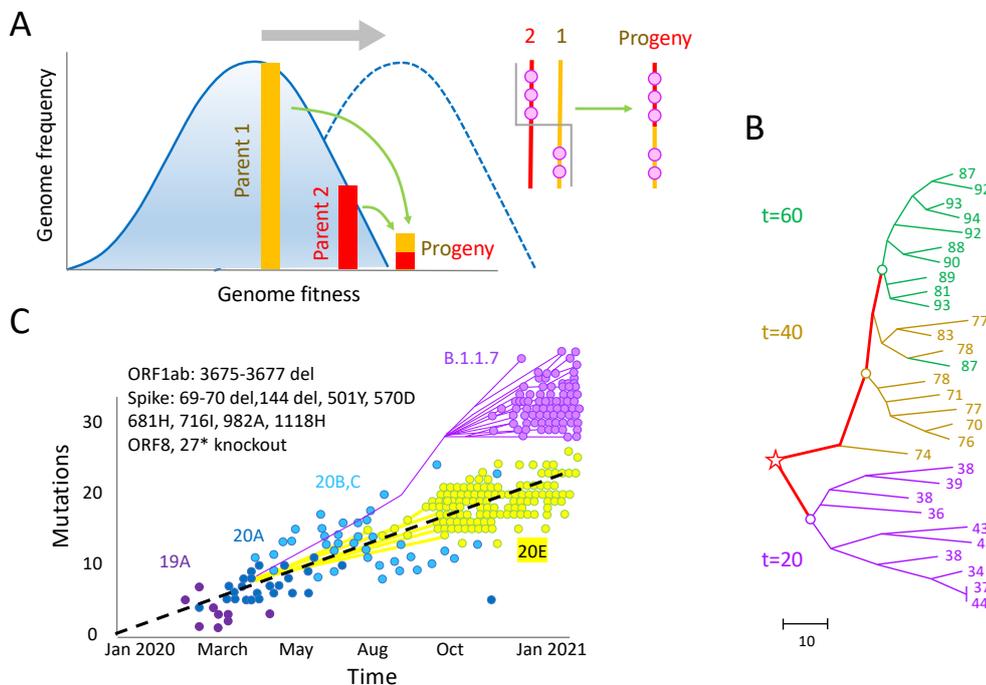


Рисунок 14. Модель образования варианта SARS-CoV-2 «вызывающего беспокойство». А. Формирование высокоприспособленного генома рекомбинацией. В. Филогенетическое дерево симулированных геномов имеет полифилиетическую структуру с общим стволом (получена в MEGA11). Количество полезных аллелей в геноме показано на листьях. С. Число мутаций как функция времени и филогенетика штамма Альфа SARS-CoV-2 (Corey, et al. 2021). По работе (Rouzine 2023).

Эффекты генетического сцепления не включают «неравновесия сцепления». Таким образом, несмотря на половое размножение $r = 1$ и при любом числе кроссинговеров M , в длинном геноме рекомбинация не может компенсировать сильные эффекты генетического сцепления, такие как клональная интерференция и генетический фон. Тем не менее, доли аллелей в удаленных локусах не коррелируют (Rouzine 2023), как и должно быть в соответствии с законом Моргана.

Применение к вариантам вызывающим беспокойство. Результаты, полученные в настоящем исследовании, могут быть потенциально актуальны для вирусов с частой рекомбинацией, таких как ВИЧ, полиомиелит или SARS-CoV-2. Загадкой, важной для разработки будущих стратегий вакцинации, является происхождение вызывающих беспокойство вариантов (VOC), вызываемых большими группами новых мутаций, которые возникают одновременно. Популярные гипотезы появления VOC (Otto, et al. 2021) включают обратный зооноз, эволюцию у пациентов с ослабленным иммунитетом (Corey, et al. 2021, Kemp, et al. 2021) и эволюцию в популяционных очагах, не охваченных генетическим надзором. На основании настоящего исследования можно добавить еще одно возможное объяснение. SARS-CoV-2 имеет наблюдаемую перекрестную рекомбинацию (Ignatieva, et al. 2022, Jackson, et al. 2021, Yi 2020). Следовательно, большие пакеты мутаций могут возникать благодаря комбинированным эффектам рекомбинации и естественного отбора и представляют собой последовательности, содержащие самые быстрые локусы (рисунок 14).

Глава 5. Определение адаптивного ландшафта из геномных данных

Адаптивный ландшафт — это частный случай отображения генотипа на фенотип: зависимость приспособленности, определяемой как среднее число потомков, от геномной последовательности. Он включает как эффекты отдельных мутаций, коэффициенты отбора, так и их взаимодействие, эпистаз. 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 устойчивые к вездесущим эффектам генетического сцепления.

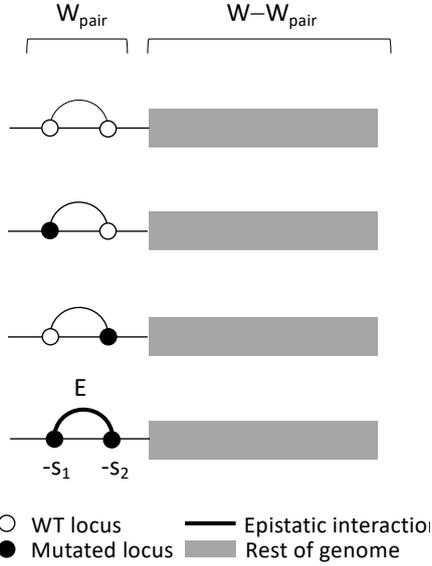


Рисунок 15. Пара взаимодействующих локусов в длинном геноме. Белые и черные кружки: дикий тип 0 и мутантный аллель 1. Толстая дуга: существующее взаимодействие. Тонкая дуга: потенциальное взаимодействие. Серый прямоугольник: остальная часть генома. Относительная сила эпистаза E определена формулой (29). По работе (Pedruzzi, et al. 2018).

§ 5.1 Универсальный эволюционный след эпистаза

До недавнего времени, метод прямой оценки силы эпистаза без запутывания и подгонки других параметров популяции отсутствовал. В работе (Pedruzzi, et al. 2018) обсуждаемой ниже, такой метод разработан и протестирован. Главный результат — это существование квазиравновесия между эпистазом и беспорядком, характерном для режима многолокусной адаптации (Rouzine 2020). Этот результат показывает как второе начало термодинамики работает для эволюции и не зависит от наличия рекомбинации.

Модель стохастической эволюции с эпистазом. Модель Монте-Карло та же что и в § 2.3, рекомбинации нет. Однако, логарифм среднего числа потомков генома теперь равен

$$W = \sum_{i=1}^L s_i K_i + \sum_{i < j} s_{ij} K_i K_j, \quad s_{ij} \equiv E_{ij} (|s_i| + |s_j|) T_{ij} \quad (29)$$

что обобщает формулу (3). Первый член в формуле (37) соответствует эффектам отдельных мутаций. Второй, новый член описывает парные взаимодействия локусов с величинами s_{ij} . Взаимодействия более высокого порядка пренебрегаются. Фактор E_{ij} - относительная величина взаимодействия между локусами i и j , в то время как матрица T_{ij} указывает взаимодействующие, $T_{ij} = 1$, и невзаимодействующие пары, $T_{ij} = 0$. Альтернативная модель глобального эпистаза исследована другими авторами (Good and Desai 2015).

След эпистаза для взаимодействующей пары локусов. Пусть взаимодействующая пара оснований с невыгодными аллелями $-s_1, -s_2$ и величиной эпистаза E является частью длинного генома с логарифмом приспособленности W (рисунок 15). Эпистатическое взаимодействие с другими локусами генома отсутствует.

Как показано в главах 3 и 4, распределение приспособленности W при многолокусной адаптации представляет собой узкий пик с $W = W_{\text{max}}(t)$, который медленно движется во времени. Гипотеза квазиравновесия состоит в том, что энтропия, т.е., логарифм числа конфигураций, вычисляемый по ансамблю возможных последовательностей, в каждый момент времени, близок к своему максимуму, при том ограничении, что приспособленность фиксирована на $W_{\text{max}}(t)$. При этом предположении значение S зависит только от $W(t)$, а не от времени в явном виде, то есть $S(t) = S[W(t)]$. Показано, что доли четырех возможных гаплотипов данной пары локусов связаны соотношением

$$E = 1 - \frac{\log(f_{11}/f_{00})}{\log(f_{01}f_{10}/f_{00}^2)} \quad (30)$$

где аргументы логарифмов конечны и не ноль. Формула (30) может быть использована для оценки силы взаимодействия E из набора последовательностей ДНК, РНК или белка. Результат (30) проверен в режиме адаптации с помощью моделирования Монте-Карло в широком диапазоне параметров, для длинного генома, состоящего из эпистатических пар. Он не работает на очень больших временах, когда система уже близка к стационарному состоянию и важны новые невыгодные мутации.

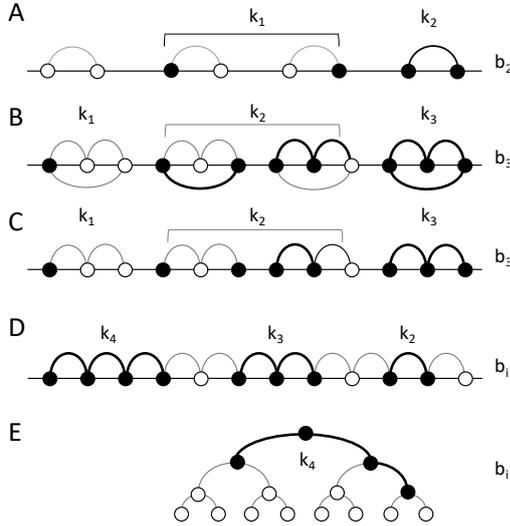


Рисунок 16. Примеры эпистатической сети. Черные и белые кружки обозначают вредные и полезные аллели, соответственно. Существующие и потенциальные эпистатические связи между локусами показаны толстыми и тонкими дугами соответственно. Толстые дуги соединяют кластеры вредных аллелей. k_i — число кластеров 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, Д) длинной цепи, где каждый локус имеет две связи. (Е) Двоичное дерево. (А-Е) показывают топологию связей, но не расположение локусов вдоль генома. По работе (Pedruzzi, et al. 2018).

Влияние топологии сети. Рассмотрены несколько примеров более сложной топологии сети (рисунок 16b-e). Коэффициенты отбора и эпистаз везде заданы одинаковыми, $s_i = -s_0 < 0$, $E_{ij} = E > 0$. В

этом случае, геном можно полностью охарактеризовать набором чисел кластеров аллелей k_i разного размера i с b_i связями (рисунок 16). Приспособленность (4.1) можно представить в виде суммы по кластерам разного размера

$$W \equiv -s_0 f_0 L = -s_0 \sum_{i=1}^{i_{max}} k_i (i - 2E b_i) \quad (31)$$

Обозначение f_0 имеет смысл эффективной доли вредных аллелей при отсутствии эпистаза с той же приспособленностью генома W . Число связей b_i для размера кластера $i > 2$ зависит от конкретной топологии (рисунок 16); для любой топологии $b_1 = 0, b_2 = 1$. Значения k_i определяются условием, что энтропия S определяемая формулой

$$e^S = \prod_{i=1}^{i_{max}} C_{L_i}^{k_i} (n_i)^{k_i} \quad (32)$$

максимальна с тем ограничением, что приспособленность W фиксирована формулой (31). По определению, энтропия есть логарифм числа конфигураций. 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. Доли гаплотипов могут быть выражены через количество кластеров k_i , $f_{11} = \frac{1}{L_{pair}} \sum_i k_i b_i$, $f_{10} = f_{01} = f - f_{11}$, где $L_{pair} = \sum_{ij} T_{ij}$ — общее количество взаимодействующих пар в геноме.

Аналитические результаты для всех топологий, полученные из этих формул, показаны на рисунке 17. Показано, что в интервале $E < E_{UFE} = \min_{i>2} \left[\frac{i-2}{2(b_i-1)} \right]$, формула (30) является точной оценкой E . В этом случае, кластерами размером три локуса и выше можно пренебречь. Точка полной компенсации, где теряется генетическая стабильность популяции, определяется из формулы $E_c = \min_i [i/(2b_i)]$, где b_i — среднее количество связей в кластере размера i (Pedruzzi, et al. 2018).

Заключение. С помощью анализа и моделирования Монте-Карло бесполой популяции, получена оценка эпистаза через частоты гаплотипов пар локусов f_{11}, f_{10}, f_{00} найденных из геномных данных, независимая от других параметров системы. Для более сложных сетей взаимодействующих локусов, модель предсказывает поправки к $UFE = E$, зависящие от топологии сети. Граница генетической устойчивости по E получена как функция топологических параметров, что можно использовать для предсказания порогов генетической стабильности предраковых клеток или устойчивости патогенов к лекарствам.

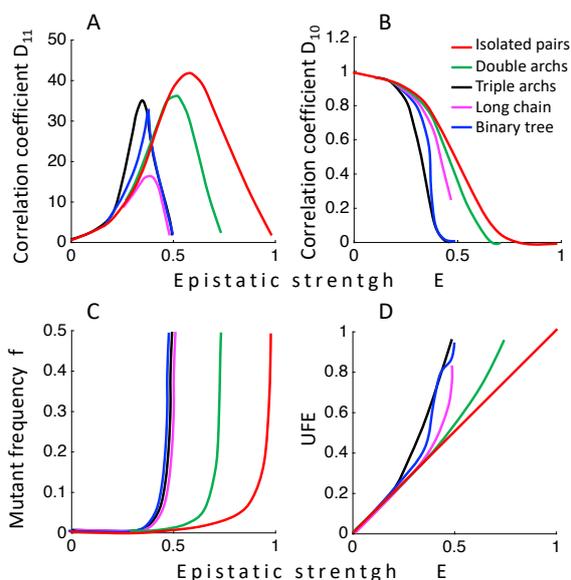


Рисунок 17. Универсальный след UFE сохраняет точность для различных топологий при умеренной величине эпистаза и требует топологически-зависимой поправки при большей. Зависимости от E показаны для пяти топологий на рисунок 16. (А, В) Коэффициенты корреляции D_{11}, D_{01} . (С) Доля вредных аллелей f . (D) UFE по формуле (38). $f_0 = 1/100$. Таким образом, UFE является точным для изолированных пар и завышает E при больших E для других топологий при $E > E_{UFE}$. Система неустойчива за пределами точки полной компенсации E_c . По работе (Pedruzzi, et al. 2018).

§ 5.2 Обнаружение эпистатических пар локусов в одной популяции: миссия невыполнима

Важным ограничением работы (Pedruzzi, et al. 2018) описанной в § 5.1 является то, что эпистатические пары локусов считаются заранее известны. В реальных геномных данных, это не так. Эффекты генетического сцепления также вызывают аллельные корреляции, и отличить эффект генетического сцепления от эпистатического взаимодействия в данных очень нелегко (Wei, et al. 2014). Проблема обнаружения эпистатических пар рассмотрена ниже (Pedruzzi and Rouzine 2019).

След эпистатических пар неотличим от генетического сцепления. Одним из широко используемых показателей корреляции является коэффициент Пирсона (Balding 2006)

$$r_{\alpha\beta}^{ij} = \frac{f_{\alpha\beta}^{ij} - f_{\alpha}^i f_{\beta}^j}{\sqrt{f_{\alpha}^i(1-f_{\alpha}^i) f_{\beta}^j(1-f_{\beta}^j)}} \quad (33)$$

который меняется между -1 до 1 , где $f_{\alpha\beta}^{ij}$ — средняя частота выбранного гаплотипа $\alpha\beta$, для локусов i и j , а f_{α}^i — доля двух возможных аллелей на локусе i .

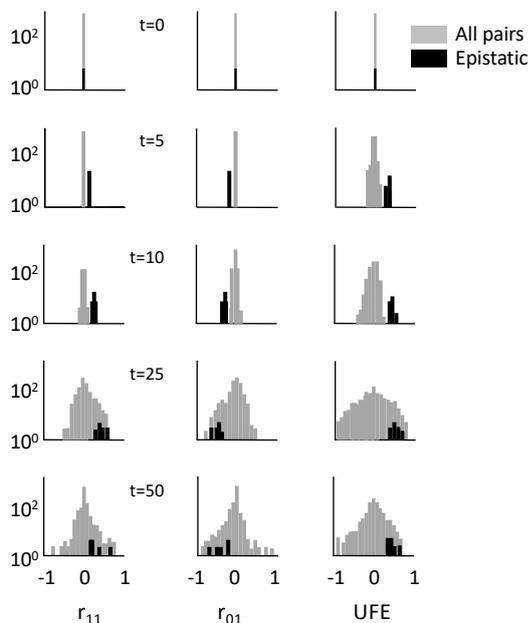


Рисунок 18. Показатели корреляции позволяют обнаружить эпистаз только в узком временном окне. В столбцах показан профиль распределения r_{11}, r_{01} , формула (33), и UFE, формула (30). Разные строки соответствуют разным моментам времени: $t = 1, t = 5, t = 10, t = 25$ и $t = 50$. Разные оттенки соответствуют 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 ((Pedruzzi and Rouzine 2019), рисунок S1). По работе (Pedruzzi and Rouzine 2019).

Другой коррелятор — это введенный выше «универсальный след эпистаза» UFE, формула (30). В оригинальной работе использовались еще два коррелятора (Pedruzzi and Rouzine 2019), с похожими результатами. Общим для всех корреляторов является то, что все они обнуляются, если $f_{\alpha\beta}^{ij} = f_{\alpha}^i f_{\beta}^j$, что означает отсутствие аллельных корреляций. Другие методы детекции эпистаза основанны на филогенетическом дереве (Kryazhimskiy, et al. 2011) и приближении «равновесия квази-сцепления»

QLE (Kimura 1965, Neher and Shraiman 2011). Как станет ясно, окончательные выводы от метода обнаружения не зависят.

Гистограммы корреляторов r_{11}, r_{01}, UFE для всевозможных пар локусов были рассчитаны, в разные моменты времени, для одной популяции полученной методом Монте-Карло (§ 2.3). Гистограммы этих значений, как для заданных эпистатических пар, так и для всех возможных пар локусов, показаны на рисунок 18. Площадь под каждым распределением — это общее количество таких пар. В процессе эволюции, распределение неэпистатических пар расширяется и поглощает распределение эпистатических пар (рисунок 18, $t = 25, t = 50$). Таким образом, за исключением узкого временного окна, отличить эпистатические пары от остальных, в одной популяции, невозможно. Использование комбинаций нескольких корреляторов не помогает (Pedruzzi and Rouzine 2019). Только рекомбинация улучшает обнаружение, в некоем промежуточном интервале r (Pedruzzi and Rouzine 2019).

Разбегание популяций ответственно за сильные эффекты сцепления. Построено филогенетическое дерево при $r = 0$ (рисунок 19), для трех независимо эволюционировавших популяций. Популяции расходятся. Генетическая дистанция между их недавними общими

предками линейно увеличивается во времени, что и приводит к сильным корреляциям аллелей со случайным знаком.

Заключение. Аллельные корреляции для неэпистатических пар локусов имеют случайный знак, и их нельзя отличить от корреляций из-за эпистаза. Это ограничение не может быть устранено никаким методом обнаружения эпистатических пар. Например, популярный метод основанный на приближении квазиравновесия сцепления (QLE), пренебрегает эффектами сцепления, полагая, что рекомбинация и эпистаз являются единственными противодействующими силами, а популяция очень велика (Kimura 1965, Neher and Shraiman 2011).

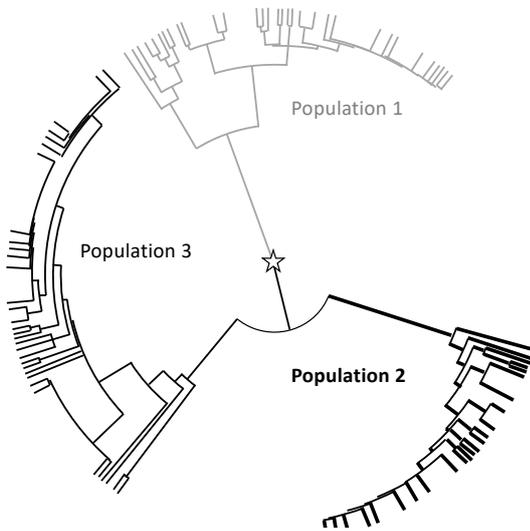


Рисунок 19. Филогенетические деревья независимо эволюционирующих популяций растут в разные стороны. Филогенетическое дерево реконструировано для трех популяций, независимо эволюционирующих из одного и того же исходного состояния, в течение 30 поколений. $N = 20\,000$ геномов, начальная аллельная частота $f_0 = 0.40$, остальные параметры как на рисунке 18. По работе (Pedruzzi and Rouzine 2019).

§ 5.3 Обнаружение эпистатических пар локусов методом трехлокусной корреляции

Единственный способ помочь обнаружению эпистаза — усреднить частоты гаплотипов по многим независимым популяциям, эволюционирующим в сходных условиях. Однако, оказывается, даже в коротком геноме из 40 локусов, двухсот популяций недостаточно. Метод устранения шума при скромном числе независимых популяций разработан в работе (Pedruzzi and Rouzine 2021).

Компьютерная модель для создания последовательностей для тестирования метода. Модель популяции как в § 5.1. Некоторые пары локусов отобраны для взаимодействия с относительной положительной амплитудой $E = 0.75$. Для демонстрации точности, выбрана простая эпистатическая сеть двойных дуг (рисунок 20B). Эта сеть производит три типа межлокусных аллельных корреляций: (i) эпистатические взаимодействия, которые заданы в программе, (ii) непрямые взаимодействия через третий локус и (iii) генетическое сцепление (§ 5.2).

Задача — обнаружить все потенциальные эпистатические взаимодействия с помощью корреляционного анализа и сравнить с заданными в программе. Для каждой пары оснований (i, j) , мера корреляции, введенная в § 5.1, имеет вид формулы (30) где $f_{00}, f_{10}, f_{01}, f_{11}$ — частоты гаплотипов, усредненные по результатам моделирования (рисунок 20A). Рассматриваются пары

оснований, где $f_{ij} > f_{cut}$, где $f_{cut} = 0.05$, и $UFE_{ij} > 0.5$. В соответствии с результатами в § 5.2, граф корреляций содержит сотни ложных взаимодействий с $UFE_{ij} > 0.5$, которые прячут 24 истинные взаимодействия (рисунок 20C). При усреднении f_{ij} по 200 прогонам (независимым популяциям), количество ложных связей уменьшается на порядок (рисунок 20D).

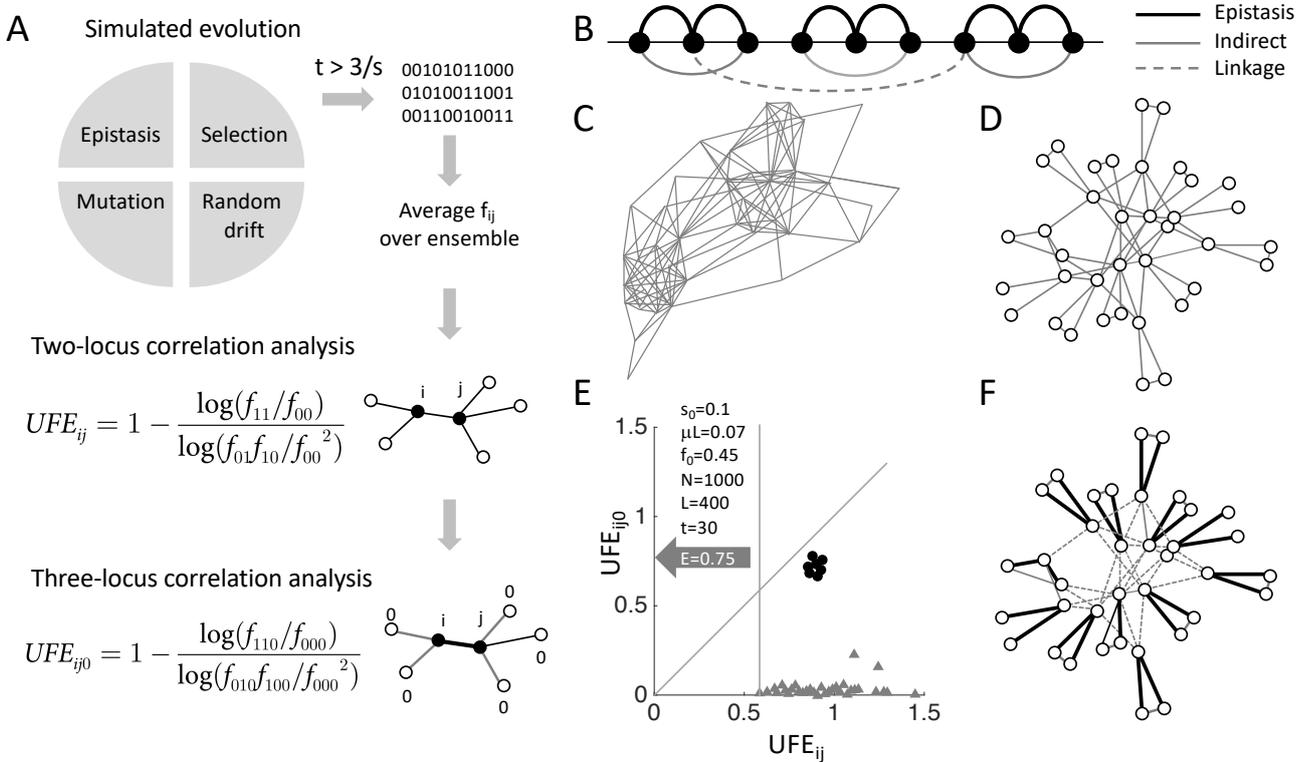


Рисунок 20. Схематическая диаграмма метода и его тестирование на смоделированных данных последовательности. А. Компьютерная модель бесполой эволюции включает факторы случайной мутации, отбора, эпистаза и случайного генетического дрейфа. Частоты парных гаплотипов f_{ij} усредняются по результатам моделирования (независимые популяции). Мера парной корреляции UFE_{ij} рассчитывается как показано. Косвенные связи и остаточная связь обнаруживаются и отфильтровываются с помощью показателя тройной корреляции UFE_{ij0} . 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 (C) 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 (Pedruzzi and Rouzine 2021).

Трехлокусный коррелятор. Остаточные ложные связи вызваны недостаточным усреднением f_{ij} по ограниченному ансамблю. Предложенный ниже протокол перерезает важные окольные цепи аллелей вокруг пары локусов, которые создают ложноположительные связи. Для этого коррелятор UFE_{ij0} снова вычисляется для каждой пары локусов, но используются только геномы, в которых соседний локус пары имеет аллель 0 (рисунок 20A). Затем вычисляется минимальное значение по всем 0-узлам, $\min(UFE_{ij0})/UFE_{ij}$. Ложные связи определяются как связи с низким отношением $\min(UFE_{ij0})/UFE_{ij}$. Точечная диаграмма подтверждает, что для ложных пар, $\min(UFE_{ij0})$ в несколько раз меньше, чем UFE_{ij} (рисунок 20E). Для прямых эпистатических взаимодействий, как и ожидалось, эти два показателя почти одинаковы. В итоге, восстановленная эпистатическая сеть в точности совпадает с сетью, заданной при моделировании (рисунок 20B). В дополнение, использование UFE дает неплохие оценки эпистатической силы, $UFE \approx E$ с 15% точностью (рисунок 20F).

Аналитическая проверка. Исключение косвенных взаимодействий было исследовано также аналитически, в предположении что доли гаплотипов хорошо усреднены по популяциям (Pedruzzi and Rouzine 2021). Было использовано приближение адаптивного квазиравновесия (Pedruzzi, et al. 2018), введенное в § 5.1. Результат подтверждает симуляцию выше. В общем случае более сложной эпистатической сети с петлями, количество дополнительных нулей, необходимых для исключения ложных пар, должно быть равно топологической связности сети.

Применение к вирусу гриппа А. Следующая задача была в том, чтобы выявить конкретные мутации и их взаимодействия, вызвавшие появление пандемического штамма 2009 года. Около 8000 последовательностей нейраминидазы вируса гриппа А H1N1 за период 2000-2010 гг были загружены из базы данных <https://www.fludb.org>.

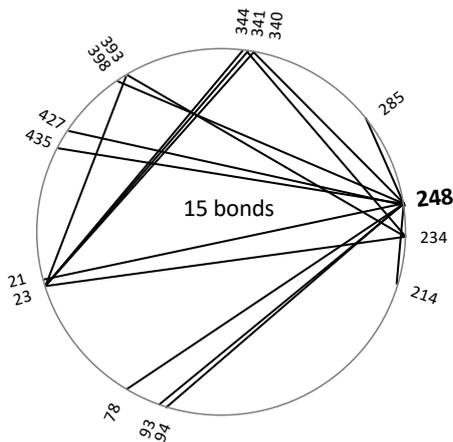


Рисунок 21. Эпистатическая сеть (долина приспособленности), ответственная за пандемию 2009 года и предсказанная из последовательностей белка нейраминидазы гриппа А H1N1, полученных в период с 2005 по 2010 год. По работе (Pedruzzi and Rouzine 2021).

Примерно одинаковое количество последовательностей было отобрано из нового и старого штаммов, несколько сотен раз, для усреднения. Пандемический штаммы до и после 2009 г. различались примерно в 100 аминокислотах данного белка. Результат — это сеть с 16 эпистатическими взаимодействиями (рисунок 21). Локус 248 в белке NA является носителем первичной мутации, связанным с множественными компенсаторными мутациями, что подразумевает прохождение вирусом через долину

приспособленности (рисунок 21). Тот же вид сети наблюдается, например, для ВИЧ в связи с выработкой устойчивости к лекарствам (Nijhuis, et al. 1999) или ускользанием от иммунного ответа (см. ниже § 7.2). Предполагаемая первичная мутация 248 повышает инфекционность вируса через расширения диапазона допустимой кислотности среды (Takahashi, et al. 2013). Хотя эмпирические доказательства за или против предполагаемой сети пока неизвестны, для проверки этого предсказания подошел бы метод, так называемого, «глубокого сканирования» (Lee, et al. 2018).

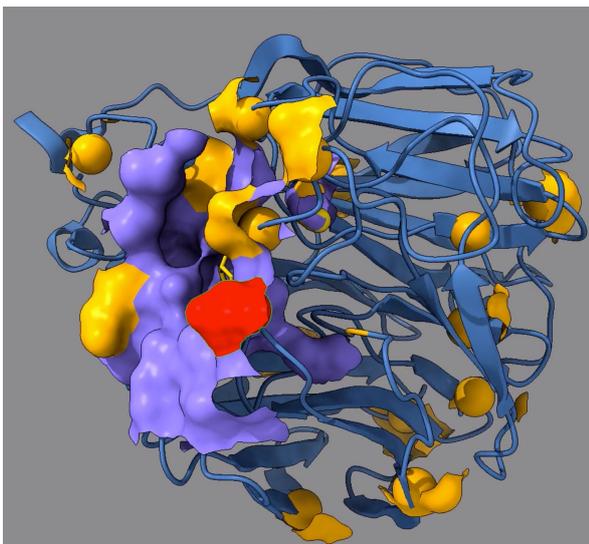


Рисунок 22: Структурное расположение предсказанной сети эпистаза для появления нового штамма нейраминидазы вируса гриппа в 2009 году. На рисунке показана трехмерная структура нейраминидазы гриппа А H1N1 (идентификационный код PDB 4QVZ). Цветные сферы представляют предсказанные эпистатические узлы из рисунка 21. Красная сфера: предсказанная первичная мутация (остаток 248). Оранжевые сферы: компенсационные узлы с рисунка 21. По работе (Pedruzzi and Rouzine 2021).

Структурная интерпретация. Чтобы понять биохимическую природу предполагаемых компенсаторных участков в белке NA (рисунок 21), они были помечены на трехмерной структуре белка (рисунок 22). Предсказанные компенсаторные мутации расположены на поверхности белка внутри -спиралей, соединяющих и определяющих

пространственную ориентацию β -листов. Компенсирующие мутации меняют ориентацию между -листами. Это случай аллостерического взаимодействия.

Заключение. Разработан и протестирован метод обнаружения взаимодействующих пар локусов в геномных данных, на фоне сильного шума из-за генетического сцепления и непрямых взаимодействий. Этот метод использован для выявления первичных и вторичных мутаций, ответственных за пандемию «свиного гриппа» 2009 года.

§ 5.4 Оценка коэффициентов отбора отдельных локусов из геномных данных

§§ 5.1 - 5.3 посвящены эпистатическим взаимодействиям. Адаптивный ландшафт включает также эффекты отдельных мутаций. В предлагаемой работе (Barlukova and Rouzine 2021) разработан метод измерения коэффициентов отбора для отдельных локусов, применимый при сильных эффектах генетического сцепления. Он также применим, когда популяция пока еще в процессе адаптации и далека от стационарного состояния.

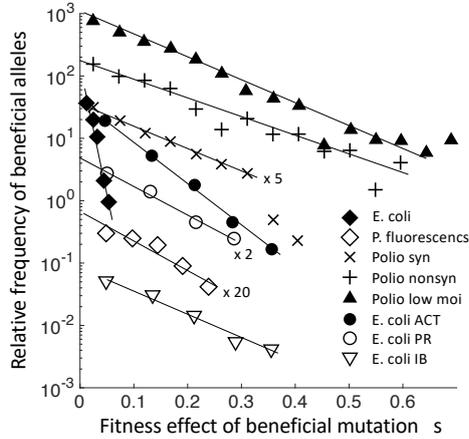


Рисунок 23. Эксперименты над вирусами и бактериями растущими в клеточной культуре, показывающие экспоненциальное распределение оснований генома по коэффициенту отбора s . Ордината: доля полезных аллелей $DFE(s, t)\varphi(s)$, определенная в формуле (42). Абсцисса: увеличение приспособленности из-за полезной мутации (коэффициент отбора). Символы представляют результаты, полученные для разных участков генома в экспериментах с *Escherichia coli* (Imhof and Schlotterer 2001), *Pseudomonas fluorescens* (Kassen and Bataillon 2006), синонимичными мутациями полиовируса, несинонимичными мутациями полиовируса (Acevedo, et al. 2014), полиовирусом с низким MOI (Stern, et al. 2014), ацетамидом *E. coli* (ACT), пропионамидом (PR) и изобутирамидом (IB) (Wrenbeck, et al. 2017). По работе (Barlukova and Rouzine 2021).

Экспериментальное распределение коэффициентов отбора имеет универсальный вид. 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 (Imhof and Schlotterer 2001, Kassen and Bataillon 2006, Acevedo, et al. 2014, Stern, et al. 2014, Wrenbeck, et al. 2017) (рисунок 23).

Полезные мутации могут возникать только на локусах, занятых невыгодными аллелями. Средняя частота невыгодных аллелей в локусе с коэффициентом отбора s , обозначаемая $f(s, t)$, представляет собой частоту локусов, доступных для полезных мутаций. Предположим, на менее приспособленном локусе, случилась мутация с преимуществом s . Эксперимент (рисунок 23) измеряет мутации с наблюдаемой плотностью распределения вида

$$\begin{aligned} DFE_{\text{obs}}(s, t) &= DFE(s, t)\varphi(s) \\ DFE(s, t) &\equiv f(s, t)g(s) \end{aligned} \quad (34)$$

где $\varphi(s)$ — вероятность невымирания аллеля, а функция $g(s)$ — «внутренняя плотность распределения» локусов, которая является свойством вируса и клеточной культуры и не зависит от времени. Она может быть измерена в эксперименте путем проведения мутагенеза для каждого основания и измерения изменения в приспособленности (Lee, et al. 2018).

Аналитический вывод универсального DFE. Экспоненциальная форма усредненной по ансамблю $\langle f(s, t) \rangle$ выведена аналитически в (Barlukova and Rouzine 2021) тремя независимыми способами. Вероятность вредного аллеля на локусе $\langle f(s, t) \rangle$ имеет вид

$$\langle f(s, t) \rangle = [1 - \langle f(s, t) \rangle] e^{-\beta(t)s} \quad (35)$$

где $\beta \equiv \partial S / \partial W$ и линейно растет со временем. Таким образом, роль генетического сцепления, после усреднения по ансамблю, сводится к замедлению времени, $t \rightarrow \beta(t)$. Компьютерное моделирование подтверждает этот аналитический результат (Barlukova and Rouzine 2021). 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, что и объясняет экспериментальный результат на рисунке 23.

Оценка коэффициентов выбора из набора последовательностей. Теоретический результат, формула (35), имеет важное практическое применение (Likhachev and Rouzine 2023), а именно, он дает 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 (Likhachev and Rouzine 2023). Выполняются ли они, становится самоочевидно в конце процедуры, то есть, она самотестирующаяся.

Заключение. Показано, аналитически и численно, что экспоненциальное распределение коэффициентов отбора полезных мутаций, часто наблюдаемое в эксперименте, есть следствие медленной адаптации и локального максимума энтропии. Разработан метод оценки коэффициентов отбора на каждом полиморфном локусе на основе наборов геномных последовательностей взятых из большого числа адаптирующихся популяций, в три момента времени.

Глава 6. Эволюционная роль признака

Свойства организмов развились в ходе длительного процесса, когда естественный отбор действовал на фоне стохастических факторов, таких как мутация, генетический дрейф, и рекомбинация (Rouzine 2020). 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, на самом деле дает важное преимущество.

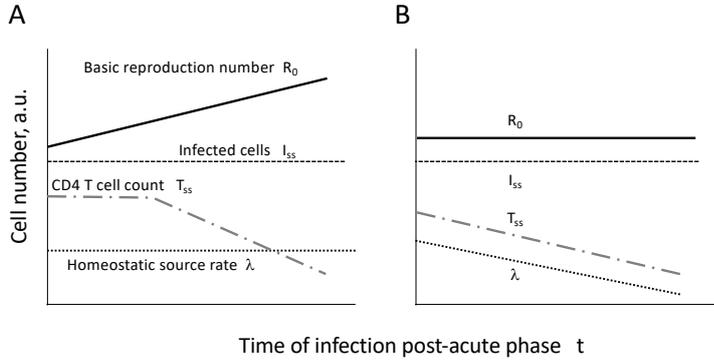


Рисунок 24: Две альтернативные гипотезы о природе постепенного снижения числа CD4 Т-лимфоцитов, вызывающего СПИД. Штрих-пунктирная линия показывает квазистационарный уровень CD4 Т-клеток T_{ss} , предсказанный формулой (45), если (А) базовое число передачи R_0 изменяется вследствие постепенной вирусной адаптации, (В) R_0 остается постоянным, но гомеостатический источник CD4 Т-клеток λ со временем ослабевает. По работе (Rouzine 2020).

§ 6.1 СПИД, как следствие адаптации ВИЧ к хозяину

Существование идеального стационарного состояния предсказывается простыми математическими моделями типа «охотник-жертва» (Rouzine and Coffin 1999, De Boer and Perelson 1998, Nowak and May 2000). Однако причина медленного снижения количества CD4 клеток на долгих временных масштабах, приводящему к СПИД, остаются неясными. Этот раздел посвящен работе (Rouzine 2020), которая предлагает модель адаптации вируса к индивидуальному хозяину, как объяснение этого эффекта (рисунок 24А).

Модель вирусной динамики. Простейшая модель вирусной динамики с цитотоксическим иммунным ответом имеет вид (De Boer and Perelson 1998)

$$\frac{dT}{dt} = \lambda - (d_T + kI)T, \quad \frac{dI}{dt} = kIT - d_I I \left(1 + \frac{E}{E_0}\right), \quad \frac{dE}{dt} = aE \frac{I}{I + I_0} - d_E E \quad (36)$$

Здесь λ — линейный источник незараженных CD4 Т клеток (T), d_T — их смертность, k — эффективность их заражения вирусом, d_I — смертность зараженных клеток (I), a — максимальная частота пролиферации иммунных цитотоксических клеток (E), I_0 — характерное число I , при котором достигается полумаксимальная экспансия иммунных клеток, d_E — смертность эффекторных клеток, E_0 — число иммунных клеток, которое увеличивает скорость гибели зараженных клеток в два раза. Полагая правые части уравнений (44) нулю, получаются стационарные значения

$$T_{ss} = \frac{T_0}{1 + \frac{R_0}{R^*}}, \quad I_{ss} = \frac{I_0 d_E}{a - d_E}, \quad E_{ss} = E_0 \left[\frac{R_0}{1 + \frac{kd_E I_0}{d_T(a - d_E)}} - 1 \right] \quad (37)$$

где введены композитные параметры

$$T_0 \equiv \frac{\lambda}{d_T}, \quad R_0 \equiv \frac{k\lambda}{d_I d_T} > 1, \quad R^* \equiv \frac{T_0 d_T}{I_{SS} d_I} \gg 1 \quad (38)$$

Модель не включает латентно-инфицированные клетки, которые важны в начальной слизистой фазе инфекции (см. ниже § 6.2), в пациентах с низкой вирусемией, и при долгой антиретровирусной терапии (Perelson, et al. 1997, Rouzine, et al. 2015). В этих случаях, динамика существенно стохастична. Модель также не включает хелперные Т-клетки, важные в пациентах с низкой вирусемией (Vingert, et al. 2010).

Скорость адаптации ВИЧ к хозяину. В данной модели, стационарный уровень незараженных клеток T_{SS} , формула (45), медленно убывает из-за увеличения R_0 со временем (рисунок 24А). Скорость адаптации V определяется как скорость, с которой логарифм приспособленности вируса, т.е., базового числа репродукции R_0 , формула (38), изменяется в единицу времени со скоростью адаптации $V = d \log R_0 / dt$. Скорость адаптации V может быть выражена через параметры эволюционной модели (главы 2-4). Для пациентов с вирусной нагрузкой $< 10^5$ копий РНК/мл плазмы, рекомбинацией можно пренебречь (Batorsky, et al. 2011), и скорость адаптации равна (Rouzine, et al. 2008)

$$V = 2s^2 \frac{\log(I_{SS} \sqrt{s} U_b)}{\log^2 \frac{s \log I_{SS}}{U_b}} \quad (39)$$

где U_b — частота полезных мутаций на геном на поколение инфицированных клеток (1 день), s — увеличение $\log R_0$ на одну мутацию, а I_{SS} — число инфицированных клеток при хронической инфекции, формула (37). Здесь значение s есть среднее эффективное значение, которое оценивается из геномных данных (§ 2.3). У пациентов с высокой вирусемией, $> 10^5$ копий РНК/мл плазмы, для оценки частоты замен V необходимо использовать модели, включающие рекомбинацию (глава 4).

Прогнозирование скорости прогрессирования СПИДа. Стационарное количество CD4 Т-клеток T_{SS} , формула (37), уменьшается с базовым числом репродукции R_0 , которое медленно увеличивается во времени в процессе адаптации, $V \equiv \frac{d \log R_0}{dt}$. Средний исходный уровень CD4 Т-клеток в хронической ВИЧ-инфекции составляет $T_{SS}(0) = 500$ клеток/мкл крови. Ранние симптомы СПИДа начинаются когда $T_{SS}(t_{AIDS}) < 200$ клеток/мкл крови. Время до симптомов СПИДа

$$t_{AIDS} \approx \frac{\log R_0(t_{AIDS}) - \log R_0(0)}{V} = \frac{1}{V} \left\{ \psi + \log \left[1 + \frac{R^*}{R_0(0)} (1 - e^{-\psi}) \right] \right\} \quad (40)$$

где V и R^* выражены через I_{SS} из (39) и (38), соответственно. Это предсказание модели может быть использовано для прогнозирования времени до наступления СПИД в индивидуальных пациентов. Если воспользоваться оценкой s найденной в §2.3, формула (40) дает $t_{AIDS} \sim 10$ лет, что клинически верно.

Отрицательная корреляция между временем до СПИДа и вирусной нагрузкой подтверждается наблюдениями. Формулу (48) можно переписать как функцию вирусной нагрузки v

$$t_{AIDS} = \frac{A}{\log(Bv)} \left[1 + \log \left(1 + \frac{C}{v} \right) \right] \quad (41)$$

где A, B, C выражаются через параметры модели. Используя A и B как подгоночные параметры, формула (41) подогнана к данным наблюдений в пациентах (Arnaut, et al. 1999) (рисунок 24). Причиной хорошего согласия модели адаптации с данными является медленная логарифмическая зависимость скорости эволюции от вирусной нагрузки. Эта особенность является универсальным свойством большого класса моделей эволюции с множественными эволюционирующими локусами (главы 3 и 4) (Rouzine 2020).

Альтернативные модели нарушения гомеостаза (Okoye and Picker 2013). Первый способ ввести в модель нарушение гомеостаза — постулировать, что максимальное число делений конечно из-за теломеров, $\lambda(T_{SS}) t_{AIDS} = A$, и источник незараженных клеток регулируется гомеостазом, как $\lambda(T) = \lambda_0 (1 - T/T_0)$. Главная проблема этой модел - это отсутствие предсказания постепенного спада числа CD4 Т-клеток. По этой модели, стационарное состояние просто резко обрывается во времени, что не соответствует клиническим наблюдениям. Далее, прогрессирование СПИДа также может быть вызвано постепенной деградацией гомеостатической среды, из-за накопления вирусных

продуктов (Okoeye and Picker 2013), откуда можно постулировать, что $d\lambda/dt = -aI_{SS}\lambda$, $R_0 = const$ и повторить вывод для $T(t)$. Однако, конечный результат для t_{AIDS} слишком резко уменьшается с вирусной нагрузкой по сравнению с фактической корреляцией (рисунок 25), так что и эта модель не годится.

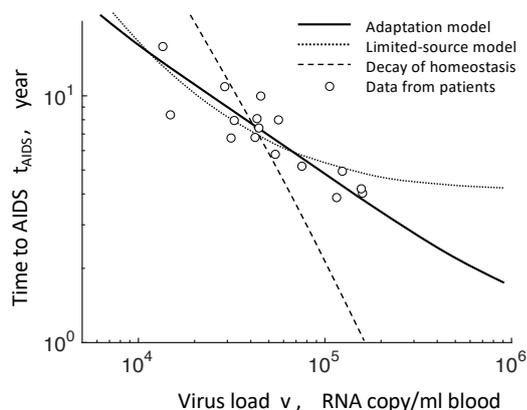


Рисунок 25. Время до СПИДа, предсказанное тремя моделями с двумя составными параметрами подбора, подогнанное к данным по ВИЧ+ мужчинам, сероконверсия которых произошла во время исследования (Agnout, et al. 1999). Из 16 пациентов, 11 не получали лечения, а 5 получали неоптимальную антиретровирусную терапию в той или иной форме (не тройные коктейли); на результат это лечение не повлияло. Кружки: время выживания в зависимости от вирусной нагрузки в данных пациентов. Сплошная толстая кривая: наилучшее предсказание модели адаптации, формула (41). Пунктирная кривая: предсказание модели ограниченного гомеостатического источника. Пунктирная линия: предсказание модели распада гомеостатической среды, вызванного вирусом. По работе (Rouzine 2020).

Заключение. Прогрессирование к СПИДу можно объяснить эволюцией ВИЧ, который со временем становится более приспособленным к хозяину и, тем самым, снижает количество клеток-мишеней, находящееся в стационарном состоянии.

§ 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 (Chun, et al. 1997, Finzi, et al. 1997). Размер латентной клеточной популяции в хронических пациентах относительно невелик, около одной клетки на миллион CD4 T-клеток (Pierson, et al. 2000). Однако латентный резервуар важен, потому что он заставляет вирус сохраняться во время длительной антиретровирусной терапии и является основным препятствием для лечения ВИЧ (Richman, et al. 2009). Однако, его эволюционная польза в течение миллионов лет естественной эволюции лентивирусов в отсутствие терапии была неясна. В работе (Rouzine, et al. 2015), рассмотренной ниже, исследована гипотеза о том, что латентность обеспечивает преимущество в приспособленности на стадии передачи между людьми или животными (рисунок 26). Эксперименты на приматах свидетельствуют о том, что слизистая оболочка изначально неблагоприятна для роста вируса иммунодефицита обезьян, $R_0^{muc} < 0.1$ (Miller, et al. 2005). 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 (рисунок 26).

Модель передачи ВИЧ. В отсутствие иммунного ответа, который включен позже, и для больших клеточных популяций, базовая модель (рисунок 26) описывается обыкновенными дифференциальными уравнениями

$$\frac{dT}{dt} = b - d_T T - kVT, \quad \frac{dI}{dt} = (1 - p_{lat})kVT - d_I I, \quad \frac{dL}{dt} = p_{lat}kVT - d_L L - rL, \quad \frac{dV}{dt} = nd_I I - cV \quad (42)$$

От модели в § 6.1, эта модель отличается присутствием латентных клеток и явным включением вирусных частиц. С вероятностью p_{lat} , клетки T 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. Уравнения (42) решались численно с помощью программы ode15s.m в пакете MATLAB™.

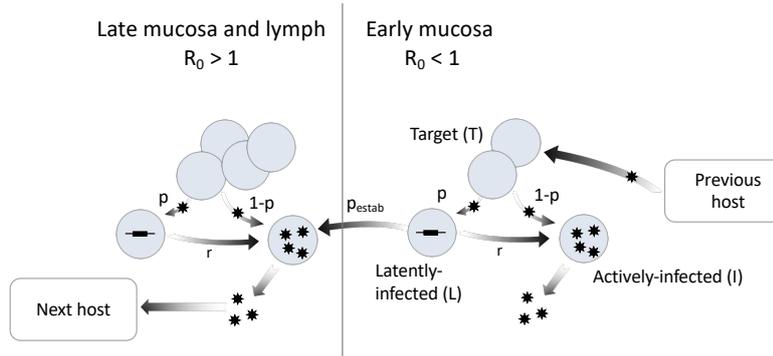


Рисунок 26. Латентность ВИЧ — это стратегия троянского коня для оптимизации передачи вируса. Схема двухкамерной модели. Справа налево: ВИЧ проникает в хозяина через слизистую оболочку, но из-за небольшого количества перmissive клеток-мишеней в ранней фазе инфекции слизистой оболочки (до 6-го дня) $R_0 < 1$. Чтобы успешно установить системную инфекцию, вирус должен сохраниться до момента, когда $R_0 > 1$, ожидая в латентно инфицированных клетках. Каждый из двух компартов описывается уравнениями (42) или их стохастическим аналогом. По работе (Rouzine, et al. 2015).

Латентность влияет как на слизистую, так и на системную инфекцию. Меньше 1% незащищенных половых контактов приводит к системному заражению ВИЧ (Fraser, et al. 2007). Вероятность передачи вируса между партнерами во время полового контакта мала и пропорциональна произведению числа инфицированных клеток в слизистой, I_0 , и вероятности того, что инфицированная клетка установит системную инфекцию, p_{estab}

$$p_{trans}(p_{lat}) = p_{estab}(p_{lat}) I_0(p_{lat}) \ll 1 \quad (43)$$

Вероятность p_{estab} пропорциональна p_{lat} . Средняя инфекционная доза I_0 , переданная новому человеку, грубо пропорциональна средней вирусной нагрузке при системной инфекции (Fraser, et al. 2007). Аналитический результат в модели с уравнениями (42) имеет вид (Rouzine, et al. 2015)

$$I_0 \approx \text{const}(p_{lat}) \cdot [(1 - p_{lat})R_0^{LT} - 1] \quad (44)$$

что монотонно уменьшается с p_{lat} (рисунок 27B). Из уравнений (43) и (44), вероятность передачи

$$p_{trans} \approx \text{const}(p_{lat}) \cdot p_{lat} \cdot [(1 - p_{lat})R_0^{LT} - 1] \quad (45)$$

что имеет максимум (рисунок 27C) при

$$p_{lat}^{opt} \approx \frac{1}{2} - \frac{1}{2R_0^{LT}} \quad (46)$$

Эксперименты на клеточной культуре подтверждают предсказания модели. Предсказанная латентная вероятность 50% действительно наблюдалась в нескольких культурах клеток (Dahabieh, et al. 2013) и в макаках-резусах (Whitney, et al. 2014). Однако это значение отличается от очень низкой частоты латентных клеток, наблюдаемой у хронически инфицированных пациентов, получающих анти-ретровирусное лечение, $p_{lat} \sim 10^{-5} - 10^{-4}$ (Sedaghat, et al. 2007).

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 CD8 Т-клетками, могут активировать клетки и активировать транскрипцию ВИЧ (Calvanese, et al. 2013). Кроме того, CD8 Т-клетки убивают зараженные клетки и подавляют размножение вируса (§ 6.1). 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 (Razooku, et al. 2015), что и объясняет пониженный уровень латентности у хронически инфицированных пациентов (рисунок 28).

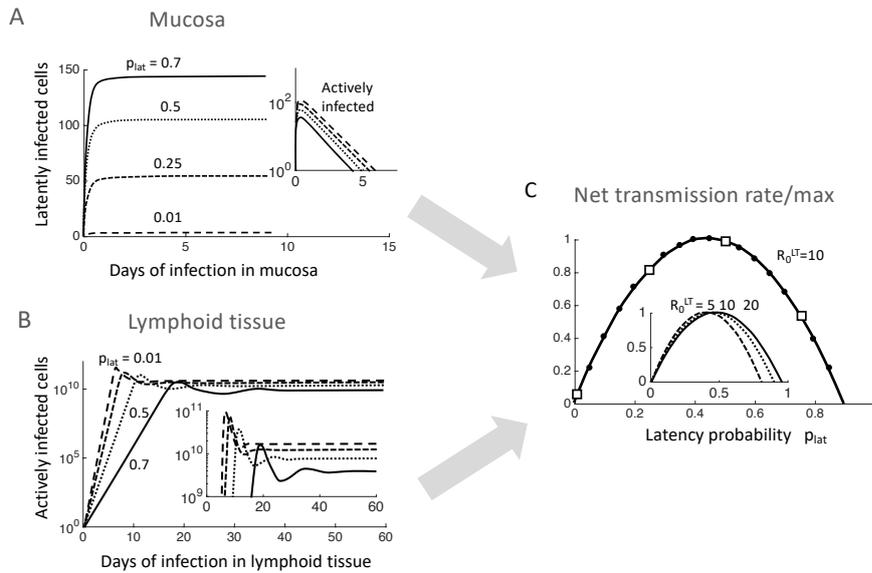


Рисунок 27: Вероятность оптимальной латентности оптимизирует увеличение скорости установления инфекции по сравнению с уменьшением заражения между хозяевами. (А) Динамика латентно инфицированных клеток при ранней инфекции слизистой оболочки ($R_0^{muc} = 0.25$). Количество выживших латентно инфицированных 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 численным решением уравнений (50) при $b_{LT} = 2 \cdot 10^{11} d_T, d_T = 0.1, d_I = 1, r = 0, d_L = 0$. (С) Нормализованная вероятность передачи вируса, p_{trans} , как функция p_{lat} , полученная аналитически из формулы (53) (сплошная линия) и численно с использованием уровней плато для активно инфицированных клеток I и латентно инфицированных клеток L из (А, В) (точки). По работе (Rouzine, et al. 2015).

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 28В).

$$p(E) = p(0) \frac{E_{0L}}{E_{0L} + E}, \quad r(E) = r(0) + d_I \frac{E}{E_{0L} + E} \quad (47)$$

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 $r(0) = 10^{-4} - 10^{-3}$ соответствуют отсутствию ВИЧ-специфического иммунного ответа. Расширенная модель, объединяющая уравнения (42) с иммунным откликом, решена численно, и результаты легко подгоняются к данным в пациентах по всем компартментам, используя четыре подгоночных параметра (Rouzine, et al. 2015).

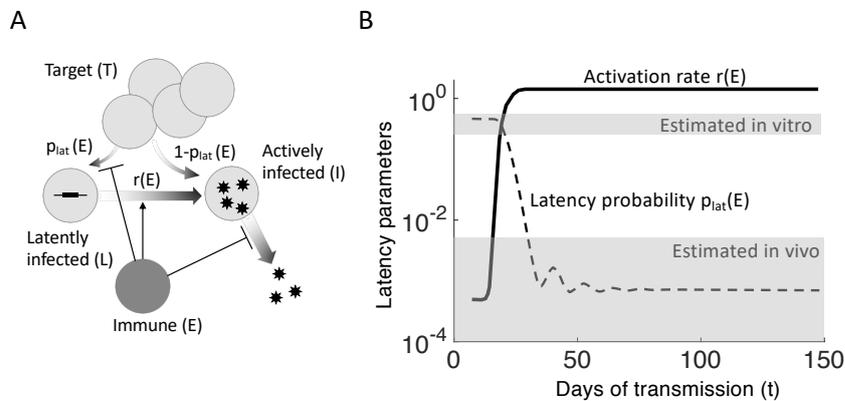


Рисунок 28: Включение иммунного ответа объясняет разницу в частоте латентных клеток ВИЧ между моделями клеточных культур и хронически инфицированными людьми. (А) Расширенная модель включает уравнения (42), а также дополнительные члены и уравнения для CD8 Т-клеток, E , которые убивают активно инфицированные клетки и активируют латентно инфицированные клетки (Rouzine, et al. 2015). (В) Вероятность латентности p_{lat} и скорость реактивации r сильно изменяются после пика виремии из-за активации сторонних цитокинов иммунными клетками, формулы (47). По работе (Rouzine, et al. 2015).

Заключение. Латентность ВИЧ служит троянским конем, который позволяет вирусу проникать со слизистой внутрь организма привлекая CD4 клетки и вызывать системную инфекцию.

§ 6.3 Роль рекомбинации в адаптации и оптимальная частота мутаций в системе вирус полиомиелита - мышшь

Модели многолокусной эволюции, в том числе результаты *главы 3*, предсказывают существование оптимальной частоты мутаций для скорости адаптации (Rouzine, et al. 2003, Gerrish, et al. 2013), из-за баланса между выгодными и невыгодными мутациями. Далее, эволюционная роль рекомбинации заключается в снижении негативных последствий генетического сцепления (*глава 4*). Обсуждаемая ниже работа (Xiao, et al. 2016) проверила эти два предсказания экспериментально и с помощью моделирования.

Эксперименты на штаммах с измененной частотой рекомбинации и мутации. Используя серийные пассажи, экспериментаторы (Xiao, et al. 2016) искусственно селектировали мутантный вирус с необычно низкой скоростью рекомбинации, в 10 раз меньше, который содержал единственную мутацию аспартата на гистидин на номере 79, обозначенную D79H. Чтобы изучить эффект одновременного изменения частоты рекомбинации и мутации D79H, был объединен с ранее обнаруженным высокоточным мутантом G64S (т.е., с низкой частотой мутаций) или низкоточным мутантом H273R (Vignuzzi, et al. 2006).

Двойные мутанты нарушают адаптацию вируса у мышей. Чтобы убить мышшь, вирус должен быстро адаптироваться и проникнуть в ЦНС, пока его не вычистил иммунный ответ (Whitton, et al. 2005). Поэтому патогенность вируса является мерой его скорости эволюции. Пять мутантов полиовируса описанные выше и дикий тип использовались для заражения мышей (рисунок 29А). После внутривенного заражения, вирус дикого типа (WT) быстро размножился в организме животного, вторгнулся в ЦНС и убивал всех мышей в течение 6 дней (рисунок 29В). Использование двойных мутантов с измененной частотой мутации, в любую сторону, и дефективных по рекомбинации резко снижало патогенность (рисунок 29В и С).

Математическое моделирование. При наличии начальной генетической вариации, рекомбинация и естественный отбор действуют независимо от мутаций *de novo* (*глава 4*) (Rouzine and Coffin 2005, Gheorghiu-Svirschevski, et al. 2007, Rouzine and Coffin 2007, Rouzine and Coffin 2010, Dutta, et al. 2008). В долгосрочной эволюции или при отсутствии исходного полиморфизма, два процесса действуют синергически (Neher, et al. 2013). Следовательно, следующей задачей было проверить, на основе эксперимента, (А) действительно ли адаптация имеет оптимум по вероятности мутаций, и (Б) зависит ли оптимальная вероятность мутаций от рекомбинации. Чтобы выяснить, какая гипотеза соответствует экспериментальной системе, для коротких промежутков времени, была использован программа Монте Карло (Batorsky, et al. 2011) из § 2.3.

Без рекомбинации, моделирование подтвердило существование оптимальной скорости адаптации при $\mu L \approx 0.3$ (рисунок 30А). Это оптимальное значение приближенно равно частоте мутаций полиомиелита дикого типа. Результаты показывали, что оптимальная частота мутаций μL не изменяется из-за рекомбинации, но все скорости адаптации увеличиваются равномерно для всех скоростей мутаций (рисунок 30А).

Полу-гетерозиготность $d = 1.5$ была подогнана, чтобы объяснить данные по вирулентности на рисунке 29. Это подогнанное значение находится в пределах диапазона, наблюдаемого для полиовируса в клеточной культуре (Xiao, et al. 2016), и недалеко от диапазона оцененного для вируса везикулярного стоматита, $d = 0.4 - 0.7$ (Dutta, et al. 2008). Второй и последний параметр модели, подогнанный к данным — это доля оснований с вредными аллелями, $\alpha = 0.08$. Только эти основания могут иметь полезную мутацию. Остальные параметры ($N = 10^3 - 10^4$, $T = 20$, $s = 0.2$) определены прямо из эксперимента.

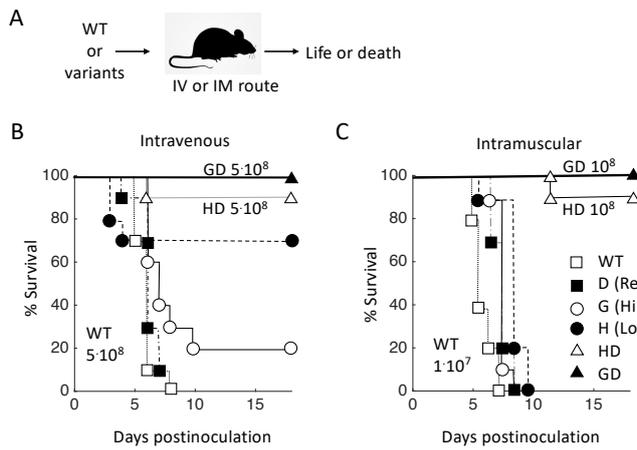


Рисунок 29: Дефект рекомбинации и изменение частоты мутаций значительно ослабляют вирулентность полиовируса. (А) Выживаемость мышей, инфицированных WT или мутантами двумя путями. (В) Процентная выживаемость восприимчивых мышей, инфицированных через хвостовую вену $5 \cdot 10^8$ инфекционных единицами 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 (Xiao, et al. 2016).

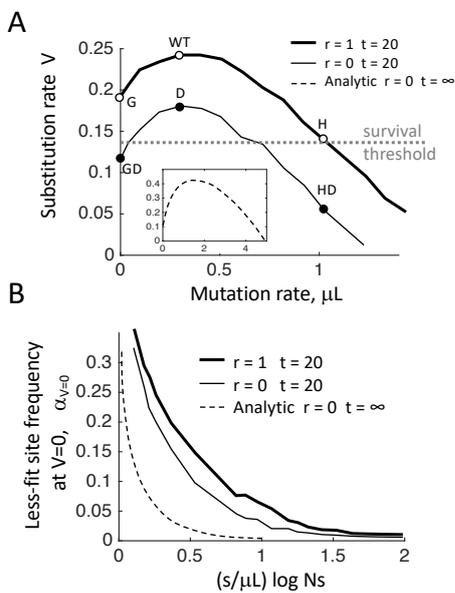


Рисунок 30. Скорость эволюции, предсказанная моделью, имеет максимум при оптимальной частоте мутаций, что объясняет различия в смертности мышей между штаммами вируса. (А) Кривые: полная скорость адаптации V , определяемая как количество полезных минус вредных мутаций на геном за цикл репликации, показанная как функция скорости геномных мутаций μL для половых ($r = 1$, толстая кривая) и бесполовых ($r = 0$, 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 (Batorsky, et al. 2011). Кружки: экспериментальные штаммы вируса, способные вторгаться в ЦНС (над пунктирной линией), неспособные к вторжению (ниже пунктирной линии) или находятся на пороге (пунктирная линия), подогнанном для объяснения двоякого поведения сильно мутирующего варианта (H), которое зависит от пути заражения (рисунок 29). Белые и черные кружки соответствуют рекомбинационно-компетентным и рекомбинационно-дефицитным штаммам, соответственно. Вставка: Скорость адаптации в бесполовых популяциях после длительного времени, формула (8). (В) Частота вредных аллелей, при которой скорость адаптации V равна нулю, 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, формулы (8) и (3) при $v = 0$ [также (Goyal, et al. 2012), приложение, их формула (13)]. Фиксированные параметры: (А) $\alpha = 0.008$; (А, В) $N = 1000, s = 0.2, d = 1.5, M = 1, T = 20$. По работе (Xiao, et al. 2016).

Математическое моделирование соответствует данным о выживаемости мышей. Вирус проходит в мозг, если скорость адаптации V превышает некий порог. При таком разумном предположении, модель предсказывает правильный порядок патогенеза для всех штаммов (рисунок 30А).

Заключение. Как теоретически предсказано ранее (Rouzine 2003), вероятность мутаций вируса полиомиелита эволюционировала, чтобы оптимизировать адаптацию вируса в организме хозяина и проникновение в мозг. Рекомбинация на коротких временах действует независимо от мутации. Этот результат объясняет узкий диапазон вероятности мутации на геном, наблюдаемый для многих вирусов. Он отличается от парадигмы «катастрофы ошибок», предсказывающей «плавление» генома выше некоего порога вероятности мутаций (Steinhauer and Holland 1987), основанной на детерминистских моделях эволюции.

Глава 7. Наперегонки с враждебным биологическим видом

В экологических системах виды вынуждены адаптироваться к изменяющейся среде, в которой обитают им враждебные и эволюционирующие виды (Clarke, et al. 1994). 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.

§ 7.1 Эволюция респираторного вируса в популяции под давлением гуморального иммунитета

Вирусы генетически изменяются, чтобы избежать иммунной памяти в ранее инфицированных людях. Например, глобальная персистенция вируса гриппа из-за повторного заражения ранее инфицированных людей вызывается быстрой эволюцией областей связывания антител в белке гемагглютинаина, служащим вирусным антирецептором (Smith, et al. 2004). В ответ, иммунная система следует по пятам за эволюцией вируса, производя лимфоциты, которые распознают эти новые антигенные варианты. Чтобы аналитически выразить скорость эволюции через параметры системы, обсуждаемая ниже работа (Rouzine and Rozhnova 2018) объединила стандартный эпидемиологический подход «инфицированный-выздоровевший» (Lin, et al. 2003) с иммунобиологией (Murphy 2011) и теорией стохастической многолокусной эволюции (глава 3). Применение данной модели к SARS-CoV-2 и разработке вакцин обсуждается в (Rouzine and Rozhnova 2023).

Модель передачи гриппа в популяции. Приведенный ниже анализ основан на ряде нетривиальных приближений, проверенных в оригинальной работе ((Rouzine and Rozhnova 2018), приложение). Штамм вируса помечен одномерной антигенной координатой x , определяемой как антигенное расстояние, с точки зрения связывания антител, от исходного штамма 1918 года. Предполагается, что иммунная система каждого выздоровевшего помнит лишь о самом последнем заражении, каковое приближение обосновано в исходной работе. Многомерные варианты этой эпидемиологической модели рассмотрены численно (Rouzine and Rozhnova 2018). Рассматривается усредненная эволюция на больших временах, сезонные колебания пренебрегаются. Модель использует тот факт, что эпитопы нейтрализующих антител у всех людей расположены одинаково. Стоимость мутаций эпитопов предполагается незначительной, что справедливо только для рецепторов вирусов, вызывающих респираторные инфекции.

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

$$\begin{aligned} \frac{dr(x,t)}{dt} &= -r(x)R_0 \int_x^\infty dy K(x-y)i(y,t) + i(x,t) \\ \frac{di(x,t)}{dt} &= i(x,t)[R_0 \int_{-\infty}^x dy K(y-x)r(y,t) - 1] + (\text{mutation}) \end{aligned} \quad (48)$$

где $\int_{-\infty}^\infty dx [r(x, t) + i(x, t)] = 1$. Время в уравнениях (551) измеряется в единицах времени выздоровления, t_{rec} . Максимальная скорость передачи равна базовому числу воспроизводства в наивной популяции R_0 . Люди, выздоровевшие от штамма x , могут быть инфицированы штаммом u с вероятностью, пропорциональной функции перекрестного иммунитета $K(x - u)$, которая зависит от генетического (более точно, антигенного) расстояния между штаммами $u = x - u$, так что $K(u) > 0, u < 0$; $K(u) \equiv 0, u > 0$; $K(-\infty) = 1$. Характерное антигенное расстояние $a = 1/|K'(0)| \gg 1$ называется расстоянием перекрестного иммунитета. Зависимость передаваемой доза также входит в $K(u)$.

Вариант x иногда подвергается случайной мутации, $x \rightarrow x + \Delta x$, что помогает ему снизить распознавание иммунной памятью. Средняя вероятность мутации в эпитопе при передаче вируса равна U_b , и Δx выбирается из распределения

$$\rho(\Delta x) = \frac{e^{-(\Delta x)^\beta}}{\Gamma(1+\frac{1}{\beta})}, \quad \Delta x > 0 \quad (49)$$

где β — фиксированный параметр.

Двухкомпонентная бегущая волна. Уравнения (48) имеют решение 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, $[di/du]_{u=0} = 0$. Подставляя его в (48), имеем

$$i(u) = Ac\delta(u), \quad r(u) \approx A \exp\left(-AR_0 \int_u^0 dv K(v)\right), \quad u < 0 \quad (50)$$

где константа A определяется из нормировки, $r(u > 0) = 0$, $\delta(u)$ — дельта-функция Дирака. Доля инфицированных

$$\frac{N_{\text{inf}}}{N} \equiv \int_{-\infty}^{\infty} du i(u) = Ac = \frac{2R_0c}{\pi a} \ll 1 \quad (50)$$

Моделирование подтверждает этот результат и демонстрирует бегущую волну с двумя компонентами (рисунок 31), где пик плотности зараженных $i(u)$ гораздо меньше и уже пика выздоровевших $r(u)$.

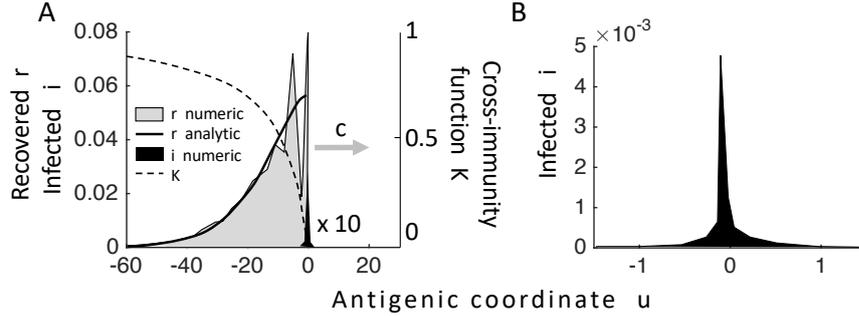


Рисунок 31: Эпидемиологическая модель предсказывает устойчивую волну, бегущую вдоль оси приспособленности. (А) Частота выздоровевших и инфицированных в популяции в системе отсчета, движущейся вместе с волной. Здесь ось абсцисс представляет собой антигенную координату в этой системе отсчета, $u = x - ct$. Сплошная линия показывает аналитический предсказание для $r(u)$, формула (50). Небольшой черный пик — это распределение инфицированных $i(u)$. Серая область показывает результат стохастического моделирования эпидемиологической модели по уравнениям (48) плюс случайные мутации с распределением (49). Штриховая линия — функция перекрестного иммунитета $K(u)$. (В) Плотность инфицированных $i(u)$. Параметры: $R_0 = 2$, $a = 9$, $U_b = 5.8 \cdot 10^{-6}$ на все эпитопы, $N = 10^8$. По работе (Rouzine and Rozhnova 2018).

Движущийся адаптивный ландшафт. Логарифм приспособленности штамма $w(x, t)$ равен экспоненциальной скорости роста плотности инфицированных $i(x, t)$ на период передачи вируса. Из (48) и анзаца бегущей волны, получается

$$w(u) = R_0 \int_{-\infty}^u dv K(v - u)r(v) - 1 \quad (51)$$

При $|u| \ll a$, где сосредоточены все зараженные люди, можно линеаризовать $w(u) \approx \sigma u$, где

$$\sigma \equiv R_0 \int_{-\infty}^0 du r(u) \left| \frac{dK(u)}{du} \right| \quad (52)$$

— средний коэффициент отбора, обратно пропорциональный расстоянию перекрестного иммунитета a и увеличивающийся с базовым числом репродукции R_0 .

Связь с теорией эволюции. Таким образом, задача об эволюции респираторного вируса сведена к раннее решенной задаче (глава 3) о многолокусной эволюции с распределенным коэффициентом отбора $s = \sigma \Delta x$ (Good, et al. 2012). Если антигенное расстояние Δx выбирается из распределения (49) с большим β , средняя частота замен c имеет вид (Rouzine and Rozhnova 2018)

$$c = \frac{\sigma^2}{s^*} \frac{2 \log(N_{\text{inf}} \sigma)}{\log\left(\frac{\beta \sigma}{U_b}\right)}, \quad s^* \equiv \sigma \left(\frac{\sqrt{2}}{\beta} \log \frac{\sigma}{U_b}\right)^{\frac{1}{\beta-1}} \quad (53)$$

где $s^* \gg \sigma$ — наиболее вероятный коэффициент отбора фиксированной мутации (Good, et al. 2012), σ подставляется из формулы (52), а размер зараженной популяции N_{inf} — из (50). Формула (61) для $\beta = 2$ и отдельный результат для $\beta = 1$ подтверждены численно, методом Монте-Карло, в широком диапазоне значений N и U_b (Rouzine and Rozhnova 2018).

Другой важной наблюдаемой величиной является время до появления самого последнего общего предка двух сосуществующих вирусов (Brunet, et al. 2007, Walczak, et al. 2012, Desai, et al. 2013, Neher and Hallatschek 2013)

$$T_{\text{MRCA2}} = z \sqrt{\frac{2 \log(N\sigma)}{v}} \quad (54)$$

где, для $\beta = 2, z = 3$. Значение $\beta = 2$ использовано для подгонки данных.

Сравнение с данными по гриппу А. 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 (Carrat, et al. 2008, Biggerstaff, et al. 2014). Две величины, предсказанные моделью, T_{MRCA2} и скорость замен c , формулы (53) и (54), также известны. Неизвестные параметры модели a и U_b , подогнаны для двух штаммов гриппа, H2N3 и H1N1, и $K(u) = |u|/(|u| + a), u < 0$ (Lin, et al. 2003). Полученная результате подгонки, для обоих штаммов, оценка $a = 14$ примерно подтверждается независимыми данными по гриппу в лошадях (Park, et al. 2009). Прогнозируемая годовая заболеваемость (4–7)% и вероятность мутаций $U_b = 3 \cdot 10^{-4}$ также находятся в пределах экспериментально наблюдаемого диапазона (ссылки в (Bedford, et al. 2015)).

Важно, что приведенные выше результаты (Rouzine and Rozhnova 2018) объясняют обратную корреляцию, наблюдаемую между T_{MRCA2} и частотой замен c в штаммах H2N3, H1N1 и вирусе гриппа В (Bedford, et al. 2015). Причина корреляции заключается в том, что предсказанная частота замен c линейно пропорциональна коэффициенту отбора σ , а T_{MRCA2} обратно пропорциональна σ .

Заключение. Рассмотрена модель эволюции респираторного вируса. Показано, что 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 аналитические результаты и оценки согласуются с предыдущим численным моделированием (Bedford, et al. 2012) и примерно подтверждены в более поздних работах основанных на более грубом приближении для коллективного иммунитета (Yan, et al. 2019, Marchi, et al. 2021).

§ 7.2 Эволюция ВИЧ в хозяине под давлением клеточного иммунитета

При репликации ВИЧ внутри хозяина, основным сдерживающим фактором являются CD8 T клетки, цитотоксические лимфоциты (CTL), которые убивают инфицированные клетки и снижают репликацию вируса с помощью секреции цитокинов. Вирус избегает полной вычистки, в частности, за счет появления мутаций в антигенно-важных областях, эпитопах, которые его делают частично устойчивым к иммунному ответу. Знание факторов, контролирующих порядок мутированных локусов, полезно для разработки эффективной стратегии вакцинации, которая до сих пор отсутствует именно из-за быстрой эволюции этого вируса. С этой целью, разработана математическая модель вирусной эволюции, управляемой давлением отбора динамически меняющихся клонов CTL (Batorsky, et al. 2014).

Модель динамики ВИЧ в присутствии множественных эпитопов. Модель включает динамику нескольких вирусных штаммов в присутствии нескольких клонов CTL (рисунок 32А) и формализуется системой

$$\frac{dT}{dt} = \lambda - d_T T - \beta T \sum_i f_i I_i, \quad \frac{dI_i}{dt} = \beta T f_i I_i - d_I I_i - \kappa I_i \sum_j r_{ij} E_j \quad (55)$$

$$\frac{dE_j}{dt} = \sigma + c E_j \sum_i \frac{(r_{ij} I_i)}{n_j + \sum_i r_{ij} I_i} - d_E E_j \quad (56)$$

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 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.

Инфицированная клетка маркируется интегрированным провирусом i , содержащим n эпитопов: $g^i = [e_1^i, e_2^i, \dots, e_n^i]$, где эпитоп $e_j^i = [a_{j1}^i, a_{j2}^i, \dots, a_{jm}^i]$, 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 уравнениях (6.17) фактором

$$f_i \equiv \exp(-\sum_j s_{jk} a_{jk}^i) \quad (57)$$

предполагающим отсутствие эпистаза (глава 5). По определению, штамм со всеми $s_{jk} = 0$ является переданным начальным штаммом с приспособленностью 1.

Предполагается, что клон активированных CTL, число которых обозначено E_j , пополняется со скоростью σ , размножается с частотой, пропорциональной количеству инфицированных клеток, экспрессирующих родственной эпитоп j , своей avidности $1/h_j$, но не выше предельной частоты c (56). Мутация в эпитопе j в основании k снижает логарифм узнавания r_{ij} в (56) соответствующим клоном CTL на относительную величину α_{jk}

$$r_{ij} \equiv \exp(-\sum_k \alpha_{jk} a_{jk}^i) \quad (58)$$

Для простоты предполагалось, что avidности разных CTL клонов равны, $h_j \equiv h$. Потери приспособленности и узнавания ниже переобозначаются как $\Delta r_{ij} \equiv 1 - r_{ij}$ и $\Delta f_i \equiv 1 - f_i$.

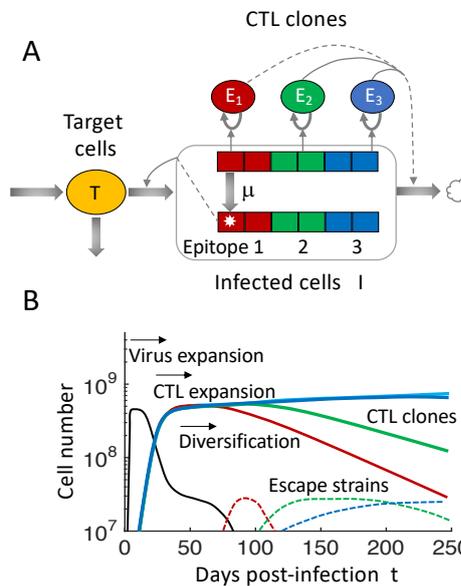


Рисунок 32: Компьютерная модель взаимодействия между ВИЧ и несколькими клонами CTL. (А) Модель, уравнения (55)-(58), включает три взаимодействующих клеточных компартмента: клетки-мишени T , инфицированные клетки I и множественные клоны CTL, E . Вирусные геномы содержат несколько эпитопов, которые могут 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, предсказанные 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, $\Delta f_i = \Delta r_i = 0.1$. По работе (Batorsky, et al. 2014).

К данной модели, добавляются случайные мутации с вероятностью $\mu = 3 \cdot 10^{-5}$ на основание на поколение инфицированных клеток (временной период $1/d_I$) (Mansky and Temin 1995). 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 в (Batorsky, et al. 2014).

Упрощенная модель для изучения порядка ускользящих мутаций. В дополнение к основной модели, удобно ввести упрощенную модель, которая предсказывает порядок фиксации мутаций ускользания на фазовой диаграмме $(\Delta f, \Delta r)$. Когда CTL клон уменьшается в ответ на ускользание локуса в эпитопе, иммунное давление на все другие локусы в этом эпитопе тоже снижается. Сокращение числа CTL, описываемое формулой (56), можно приблизительно представить как уменьшение Δr для всех оснований в эпитопе

$$\Delta r(n+1) = \Delta r(n) \exp[-d'_E \sum_i \Delta r_i(0)] \quad (59)$$

где i – основания в эпитопе, которые уже ускользнули. Параметр d'_E определяется как сокращение на одну мутацию, в отличие от d_E в основной модели, который за день.

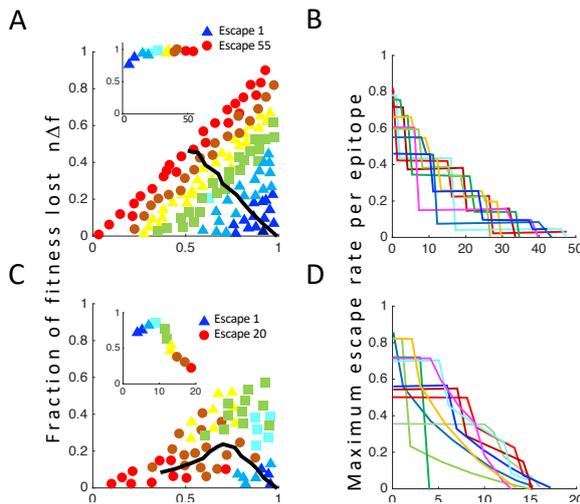
Моделирование динамики ускользания вируса. В начале симуляции, клетки-мишени T находятся на стационарном уровне неинфицированного пациента, имеется небольшое начальное количество CTL-клеток и небольшое количество вируса. Результаты численного решения уравнений (55)-(58) предсказывают три фазы ВИЧ-инфекции (рисунок 32B): 1) Передаваемый штамм ВИЧ размножается и истощает клетки-мишени. 2) CTL его распознающие размножаются и уменьшают количество инфицированных клеток до тех пор, пока не установится устойчивое состояние (хроническая инфекция). 3) Появляются и размножаются ускользающие мутанты вируса. Часть клонов CTL сжимается, но общее количество вируса остается примерно прежним (Batorsky, et al. 2014). Предсказанная трехфазная динамика наблюдалась у ВИЧ-инфицированных, включая снижение клонов CTL родственным мутированным эпитопам (Goonetilleke, et al. 2009, Liu, et al. 2011, Henn, et al. 2012, Fischer, et al. 2010).

Детерминанты скорости ускользания мутантного штамма. Присутствие CTL создает положительный отбор ускользающих мутаций (фаза 3 на рисунок 32B). Начальное состояние до мутации является стационарным, в котором правые части уравнений (55), (56) равны нулю. В эпитопе 1, мутантный штамм растет как $I_i(t) = I_i(0) \exp(\epsilon_i t)$, где

$$\epsilon_i = \frac{1}{I_i} \frac{dI_i}{dt} = kE_{tot} \left(\frac{\Delta r_i}{n_1} - \Delta f_i \right) \quad (60)$$

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, $\epsilon_i > 0$. По данным в пациентах, частота ускользания ϵ_i резко замедляется в течение первых 100 дней после заражения (Ganusov, et al. 2011). Предсказания модели согласуются с этими выводами.

Траектория ускользающих мутаций в плоскости затраты-выгода. Во время хронической инфекции, ВИЧ активирует 10-100 поддающихся обнаружению CTL-клонов против различных эпитопов, в которых зафиксировано 5-30 ускользающих мутаций (Goonetilleke, et al. 2009). Чтобы исследовать типичную траекторию ускользающих мутаций по эпитопам и их локусам в плоскости «затрат и выгод» (Δf , Δr), была использована упрощенная модель, описанная выше. 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 33). Локусы нумеровались в порядке убывания частоты ускользания ϵ_i , формула (60).



которые ускользнули, с $d'_E = 1/m = 0.1$. По работе (Batorsky, et al. 2014).

После генерации большого количества случайных наборов, выяснилось, что значения Δr и Δf локуса каждого ранга обратно коррелируют, что и наблюдается в эксперименте (Mostowu, et al. 2012). Типичная траектория в плоскости «затрат и выгод» проходит от высоких выгод (Δr) и низких затрат (Δf) к низким потерям и высоким затратам (рисунок 33A). Скорость ускользания на эпитоп уменьшается с количеством совершенных ускользаний (рисунок 33B). В отсутствие уменьшения числа CTL, траектория остается прямой до тех пор, затраты не превысят выгоды

Рисунок 33: Траектория ухода в плоскости затрат-выгод с течением времени изгибается из-за распада CTL. (A) Для каждого ускользнувшего локуса, показана относительная стоимость приспособленности, Δf , умноженная на количество эпитопов, $n = 10$, и относительная потеря распознавания, Δr . Цвета показывают ранг ускользающих мутаций, от ранних ускользающих мутаций (синий) до поздних ускользающих мутаций (красный). Врезка: Подгонка наклона траектории для каждого ранга. (B) 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 ускользнувших локусов (черная линия) показана без уменьшения числа CTL (A, B) или с уменьшением числа CTL (C, D) для 1000 запусков моделирования. Уменьшение числа CTL моделируется формулой (59), суммируя по всем i основаниям в эпитопе,

(рисунок 33А). Когда СТЛ клоны уменьшаются, средняя траектория изгибается по направлению к оси абсцисс (рисунок 33С). Также уменьшается число ускользящих мутаций на эпитоп (рисунок 33В, D).

Порядок антигенного ускользания внутри эпитопа с двумя локусами. Порядок появления ускользящих гаплотипов изменяется между эпитопами (Goonetilleke, et al. 2009). «Простой» порядок (наблюдаемый нечасто) представлен одним ускользящим гаплотипом, 00 → 10. «Вложенный» порядок добавляет новую мутацию к единственному мутанту, 00 → 10 → 11. «Чехарда» означает переключением доминантной мутации: 00 → 10 → 01. Эффект “чехарды” можно рассматривать, как вариант клональной интерференции (глава 3), когда первая ускользящая мутация испытывает более слабое давление отбора, чем вторая мутация, возникающая позже и вытесняющая первый клон (Gerrish and Lenski 1998). Порядок мутаций ускользания внутри двухлокусного эпитопа был определен аналитически для различных наборов $\Delta r_1, \Delta r_2, \Delta f_1, \Delta f_2$ и показан на фазовой диаграмме (Batorsky, et al. 2014).

Закключение. Результаты (Batorsky, et al. 2014) показывают, что процесс антигенного ускользания регулируется компромиссом между ценой и выгодой ускользящих мутаций. Данная модель также объясняет положительную корреляцию между потерей узнавания и стоимостью мутации, наблюдаемую в ВИЧ (Mostowy, et al. 2012). Результаты подчеркивают необходимость прямых измерений потерь узнавания различных ускользящих мутаций. Подход траектории на рисунок 33 может помочь с прогнозированием консервативных эпитопов для использования в вакцинах. Приведенный выше анализ имеет ряд упрощений, уместность которых обсуждается в оригинальной работе (Batorsky, et al. 2014).

§ 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) (Holland 1990). Реплицируясь за счет дикого типа, DIP препятствует 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. В работе, обсуждаемой в данном разделе (Rouzine and Weinberger 2013), исследуется возможность появления вариантов ускользания ВИЧ. Цель состоит в том, чтобы предсказать параметры конструкции DIP, необходимые для стабильного подавления ВИЧ.

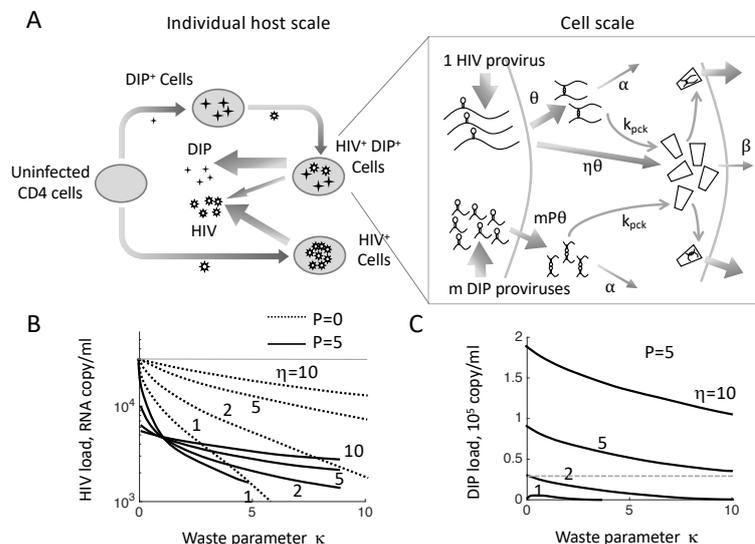


Рисунок 34. DIP, которые крадут капсид, стабильно подавляют стационарное количество ВИЧ в широком диапазоне параметров. (А) Схема модели, включающей две уровня биологической организации. Уровень индивидуального хозяина представляет собой простейшую модель репликации ВИЧ, расширенную за счет включения DIP, уравнения (70) Неинфицированные клетки могут быть инфицированы либо ВИЧ, либо DIP. DIP+ клетки могут быть суперинфицированы ВИЧ. Модель на уровне одной клетки описывается уравнениями (61). 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. (В) Стационарная нагрузка ВИЧ и (С) стационарная нагрузка DIP при различных значениях двух параметров одиночных клеток: «параметр отходов» κ и отношение образования капсида к геному η . Горизонтальная серая линия на (В) 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 (В) и (С): отношение экспрессии DIP к ВИЧ $P = 5$ и базовое число репродукции $R_0 = 10$. 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 $\eta \gg 1$ [(Rouzine and Weinberger 2013), рисунок S2]. По работе (Rouzine and Weinberger 2013).

Модель. Модель (рисунок 34) объединяет два биоуровня: одноклеточную модель, описывающей конкуренцию за белки (в модели, капсиды), и динамику вируса на уровне организме хозяина. Чтобы сделать анализ прозрачным, одноклеточная модель (рисунок 34А, 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} . Биологические процессы с параметрами показаны на рисунок 34А, справа, и описаны в его подписи. Уравнения модели имеют вид

$$\frac{dG}{dt} = \theta - k_{pck}GC - \alpha G, \quad \frac{dC}{dt} = \eta\theta - k_{pck}(G + G_{DIP})C - \beta C, \quad \frac{dG_{DIP}}{dt} = mP\theta - k_{pck}G_{DIP}C - \alpha G_{DIP} \quad (61)$$

Стационарные числа G и G_{DIP} в клетке используются в качестве входных данных для модели вирусной динамики в организме хозяина (рисунок 34А, слева).

Уровень организма. По сравнению с простейшей моделью динамики вируса (§ 6.1 или 6.2), 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 рисунок 34А, слева, описывается такими уравнениями

$$\begin{aligned} \frac{dT}{dt} &= b - (d + kV + kV_{DIP})T, & \frac{dI}{dt} &= kVT - \delta I \\ \frac{dT_{DIP\ m}}{dt} &= kV_{DIP}T_{DIP\ m-1} - (d + kV + kV_{DIP})T_{DIP\ m}, & m &= 1, 2, 3, \dots \\ \frac{dI_{D\ m}}{dt} &= kVT_{DIP\ m} - \delta I_{D\ m}, & m &= 1, 2, 3, \dots, & \frac{dV}{dt} &= n\delta I + n\delta \sum_{m=1}^{\infty} \psi_m I_{D\ m} - cV \\ \frac{dV_{DIP}}{dt} &= n\delta \sum_{m=1}^{\infty} \rho_m \psi_m I_{D\ m} - cV_{DIP} \end{aligned} \quad (62)$$

Параметры модели. Таблицы 6.3 и 6.4 данные в тексте диссертации содержат все параметры модели на обоих биологических уровнях. Параметры ρ_m, ψ_m связанные с DIP

$$n = \frac{k_{pck}[GC]_{P=0}}{\delta}, \quad \psi_m n = \frac{k_{pck}GC}{\delta}, \quad \rho_m \psi_m n = \frac{k_{pck}G_{DIP}C}{\delta} \quad (63)$$

Используя стационарные значения G, C, G_{DIP} найденные из (61), параметры ψ_m и ρ_m можно выразить через параметры уровня клетки η, mP и κ , что было сделано в исходной работе (Rouzine and Weinberger 2013). Результаты выражаются через четыре безразмерных параметра: (i) Параметр отходов, $\kappa = \alpha\beta/\theta k_{pck}$, (ii) Отношение продукции капсидов к димеризованным геномам ВИЧ, η . (iii) Базовый число репродукции, R_0 (iv) Отношение экспрессии генома DIP к ВИЧ, P . Лентивирусные DIP были сконструированы ранее в диапазоне $P = 8 - 10$ (D'Costa, et al. 2009).

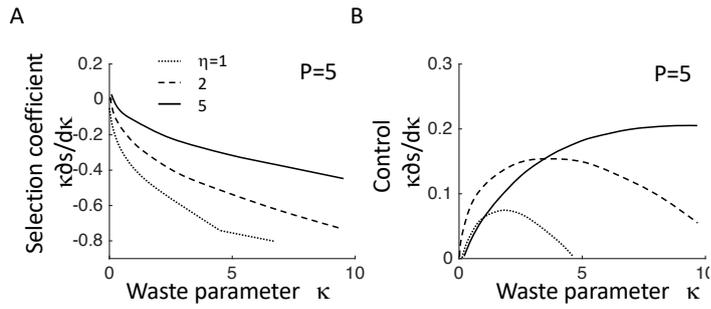


Рисунок 35: ВИЧ не может ускользнуть от DIP, увеличивая растрату ресурсов. (А) Нормализованный коэффициент отбора $\partial s/(\partial \kappa/\kappa)$ как функция κ , для различных отношений скорости продукции капсидов и геномов η , формула (64). (В) Отрицательный контроль для $\partial s/(\partial \kappa/\kappa)$ 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 (Rouzine and Weinberger 2013).

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 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 (Rouzine and Weinberger 2013). Присутствие DIP подавляет ВИЧ в широком диапазоне параметров (рисунок 34В). Степень подавления ВИЧ и нагрузка DIP увеличиваются с асимметрией экспрессии DIP и ВИЧ, $P > 1$. Условие динамической устойчивости DIP в популяции задается условием, которое при малом параметре отходов $\kappa \ll 1$ имеет вид (Rouzine and Weinberger 2013), $\eta > \eta_c = \frac{1+P}{P} \frac{R_0}{R_0-1}$. Этот порог на η понижается примерно в R_0 раз, если в модель включено деление клеток зараженных DIP, но не ВИЧ, что биологически правильно (см. § 7.4 ниже).

Подавление ВИЧ с помощью DIP является эволюционно стабильным. В принципе, ВИЧ мог бы мутировать, чтобы увеличить параметр κ . Снижение эффективности упаковки k_{pack} позволило бы деградировать больше капсидов, чтобы осталось меньше для DIP и тот вымер. Однако эта мутация также имеет свою цену, потому что увеличение отходов продуктов снизит выход вирионов ВИЧ (рисунок 34В). Коэффициент отбора ∂s , определяемый как малое увеличение приспособленности мутантного штамма ВИЧ из-за малого изменения параметра отходов $\partial \kappa$, выражен аналитически через параметры модели (Rouzine and Weinberger 2013). Ответ имеет общий вид

$$\partial s = \left[\frac{\partial n_{av}}{n_{av}} \right]_{E[m]=const}, \quad n_{av} \equiv n \frac{1 + \sum_{m=1}^{\infty} \psi_m E[m]^m}{1 + \sum_{m=1}^{\infty} E[m]^m} \quad (64)$$

где n_{av} — число вирионов ВИЧ на среднюю клетку, $E[m]$ — число ДНК провирусов DIP на клетку, а ψ_m и n выражены аналитически через η, mP, κ (Rouzine and Weinberger 2013). Мутация, вызывающая изменение параметра отходов κ , изменяет параметры ψ_m и n , которые зависят от κ .

Окончательный результат для $\partial s/(\partial \kappa/\kappa)$ (рисунок 35А) 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 34В 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 (64). 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 35В). Направление эволюционного отбора устойчиво в широком диапазоне $P > 2$ ((Rouzine and Weinberger 2013), рисунок 3D,Е).

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 (64) показывает что ВИЧ всегда развивается в сторону высоких значений η (Rouzine and Weinberger 2013), и ускользание невозможно.

Заключение. Интерференция путем кражи белков является эволюционно стабильной при условии, что она стабильна динамически. Модель имеет то важное ограничение, что она не учитывает эпистаз, т.е., компенсирующие мутации (глава 5).

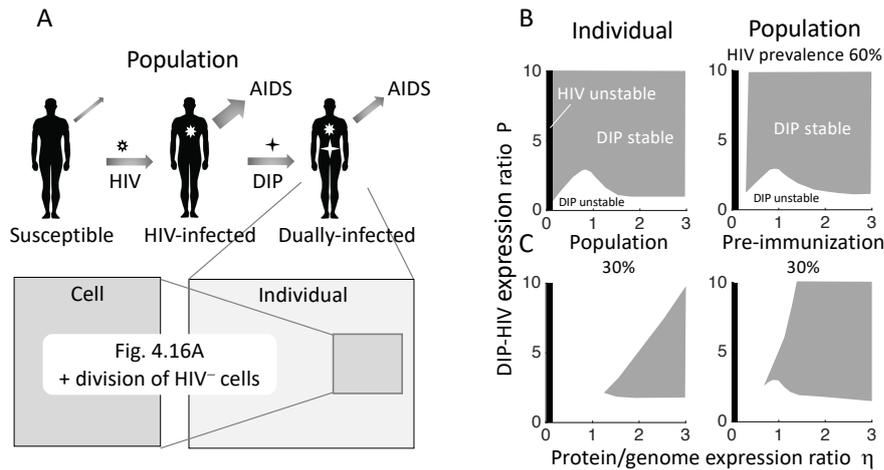


Рисунок 36. DIP и ВИЧ динамически костабильны на трех биологических уровнях. (А) Схема многомасштабной математической модели, которая отслеживает уровни ВИЧ и DIP на трех биологических уровнях: отдельных клеток, индивидуального хозяина, и населения. Модель клетки показана на рисунке 34А в § 7.3. Выходные 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 34А, 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 (65). Частоты смертности и коэффициент передачи вируса рассчитываются на основе нагрузки ВИЧ и DIP у носителей по работе (Fraser, et al. 2007). (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 (Rast, et al. 2016).

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

Как показано в § 7.3, DIP-устойчивый 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 обработанном DIPом. Действительно, из-за более высокой вирусной нагрузки, он имеет более высокую вероятность передачи (Fraser, et al. 2007), чем дикий тип, и, в конечном итоге, мог бы его вытеснить. Работа (Rast, et al. 2016) расширяет модель (Rouzine and Weinberger 2013) до уровня населения, чтобы понять, могут ли устойчивые к DIP мутанты ВИЧ распространяться в группах высокого риска.

Модель динамики ВИЧ и DIP на трех биологических уровнях. Каждый уровень в модели (рисунок 36А) представлен системой обыкновенных дифференциальных уравнений, описывающих репликацию и передачу ВИЧ и DIP.

Уравнения динамики уровня хозяина приведенные в исходной работе (Rast, et al. 2016) имеют вид (70), 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, который зависит от общего количества всех клеток и уменьшается, когда оно приближается к нормальному уровню неинфицированного хозяина [(Rast, et al. 2016), текст S1, уравнения (17), (18), (51), (52)].

Модель на уровне населения (рисунок 36А) основана на стандартном эпидемиологическом подходе «восприимчивые-зараженные» (Anderson, et al. 1992), включая DIP, как описано в (Metzger, et al. 2011), но только в одной демографической группе, группе высокого риска.

$$\begin{aligned} \frac{1}{\delta_I} \frac{dS}{dt} &= \frac{1}{B} (1 - S) - R_0^{\text{pop}} (S I + \mu S I_D) \\ \frac{1}{\delta_I} \frac{dI}{dt} &= R_0^{\text{pop}} (S I + \mu S I_D - \phi I I_D) - I, \quad \frac{1}{\delta_I} \frac{dI_D}{dt} = R_0^{\text{pop}} \phi I I_D - \tau I_D \end{aligned} \quad (65)$$

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. Параметры: R_0^{pop} — базовое число репродукции в населении, μ — относительная скорость передачи ВИЧ из ВИЧ+DIP+ людей по сравнению с ВИЧ+DIP- людьми, ϕ — относительное сокращение продолжительности жизни из-за неподавленной ВИЧ-инфекции, B — относительное сокращение продолжительности жизни из-за неподавленной ВИЧ-инфекции, τ — относительное снижение смертности от ВИЧ вследствие суперинфекции DIP, δ_I — смертность ВИЧ+DIP- людей

Уравнения (65) 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 DIP или без и перейти в компартмент I . Там, они могут быть коинфицированы DIP и перейти в 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 36B, C).

Три биологических уровня (клетка, хозяин и популяция хозяина) связываются друг с другом, 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. Результатом верхнего уровня является доля ВИЧ-инфицированных и ВИЧ-DIP-коинфицированных лиц в популяции, рассчитанная как функция клеточных параметров η и P , которая может служить для оценки терапевтической ценности DIP как самораспространяющегося агента.

Условия сосуществования ВИЧ и DIP в группах высокого риска. Фазовая диаграмма стабильности в переменных η и P на уровне хозяина и популяции показана на рисунке 36B. Включение деления DIP+ВИЧ-клеток в модель хозяина ((Rast, et al. 2016), текст S1, раздел B) значительно расширяет область стабильности DIP в хозяине по сравнению со старым условием стабильности (§ 7.3), и разрешает $\eta < 1$. В то же время, для стабильности DIP требуется достаточно большая асимметрия экспрессии провируса DIP по отношению к ВИЧ, P (рисунок 36B, слева), которая зависит от η . Максимальное значение P , при котором DIP нестабилен, составляет $P_{crit} \approx 3$, что соответствует $\eta \approx 1$.

Для популяции высокого риска, точные границы области стабильности зависят от распространенности ВИЧ до того, как в нее была введена DIP (рисунок 36B, справа), и от того, могут ли DIP преинфицировать людей до того, как они были инфицированы ВИЧ (рисунок 36C) (Metzger, et al. 2011). Если DIP не может заражать людей до инфекции ВИЧ, условие распространения и устойчивости DIP имеет вид

$$R_{\text{eff}} = R_0^{\text{pop}} - \frac{B\tau}{\phi} > 1 \quad (66)$$

Если же DIP могут преинфицировать индивидуумов и оставаться латентными до последующего инфицирования ВИЧ (Metzger, et al. 2011), то распространенность ВИЧ гораздо меньше влияет на область стабильности DIP (рисунок 36C).

Эволюционные конфликты останавливают появление DIP-устойчивых мутантов ВИЧ. Модель предсказывает два эволюционных конфликта: между адаптивными ландшафтами на уровне хозяина и популяции и внутриуровневый конфликт между передачей ВИЧ от лиц с DIP и без DIP. Чтобы определить, какие мутанты ВИЧ могли бы распространяться среди населения модель популяционного уровня обобщается с включением двух штаммов ВИЧ: исходный штамм в стационарном состоянии и потенциально DIP-резистивный штамм (рисунок 37A). Когда два штамма заражают одного и того же хозяина, более приспособленный штамм быстро фиксируется (вертикальные стрелки на рисунок 37A). Конкурентное исключение внутри хозяина (§ 2.1) происходит очень быстро по сравнению с динамикой распространенности вируса в популяции.

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

(Rast, et al. 2016) (рисунок 37В). Для определения 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 37С). Приближения данного анализа были проверены в исходной работе (Rast, et al. 2016).

Заключение. 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. Несмотря на многообещающие сдвиги с разработкой терапевтического DIP (Vignuzzi and Lopez 2019, Shirogane, et al. 2021, Xiao, et al. 2021, Chaturvedi, et al. 2021), прогресс идет медленно, и причиной может быть эпистаз, который позволяет ускользание вируса от DIP через долины приспособленности.

Заклучение

В результате работ, вошедших в данную диссертацию, было ускорено соединение генетики популяций и вирусологии, причем в качестве «клея» послужили математические методы статистической физики. Разработан общий аналитический подход к многолокусной эволюции, как в половых, так и бесполой 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 (Coffin 1995), 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.

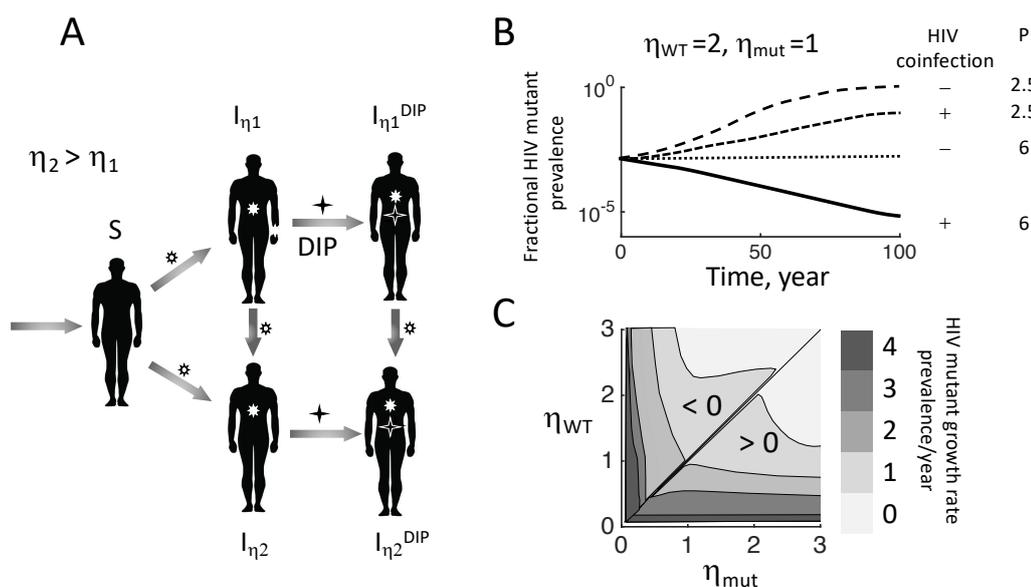


Рисунок 37: Из-за межуровневого и внутриуровневого эволюционного конфликта, DIP-резистивные мутанты ВИЧ не могут размножаться в популяции хозяев. (А) Схема эволюционной модели с конфликтом. Два мутанта ВИЧ с разными значениями η конкурируют за передачу в популяции, получающей DIP. Когда два штамма ВИЧ заражают одного и того же человека, штамм с более высоким значением η замещает другой штамм благодаря своей более высокой

приспособленности. (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. (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 $P > 3$. По работе (Rast, et al. 2016).

Выводы

1. Эффективный размер популяции у среднего нелеченного ВИЧ-инфицированного составляет не менее миллиона активно-инфицированных клеток, и невыгодные мутации придающие устойчивость к монотерапии существуют в популяции вируса уже до терапии.
2. Высокое разнообразие и скорость эволюции ВИЧ 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».
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 и приводит к потере выгодных аллелей на многих локусах. 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. Метод трех локусов позволяет исключить этот эффект для геномов длиной меньше 100 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 базовое число передачи в наивном населении и перекрестную реактивность антител.
8. В отсутствие эпистаза, вирус не может развить устойчивость к своей дефектной интерференционной частице.

Список основных публикаций по теме диссертации

Монографии

1. 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.
2. 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.

3. 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.
4. 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.
5. 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.
6. 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.
7. Rouzine I. M., Coffin J. M. Evolution of human immunodeficiency virus under selection and weak recombination // *Genetics*. – 2005. – V. 170, № 1. – P. 7-18.
8. 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.
9. 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.
10. 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.
11. Brunet E., Rouzine I. M., Wilke C. O. The stochastic edge in adaptive evolution // *Genetics*. – 2008. – V. 179, № 1. – P. 603-20.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.

22. 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.
23. Pedruzzi G., Barlukova A., Rouzine I. M. Evolutionary footprint of epistasis // *PLoS Comput Biol.* – 2018. – V. 14, № 9. – P. e1006426.
24. Rouzine I. M., Rozhnova G. Antigenic evolution of viruses in host populations // *PLoS Pathog.* – 2018. – V. 14, № 9. – P. e1007291.
25. Pedruzzi G., Rouzine I. M. Epistasis detectably alters correlations between genomic sites in a narrow parameter window // *PLoS ONE.* – 2019. – V. 14. – P. e0214036.
26. Rouzine I. M. An Evolutionary Model of Progression to AIDS // *Microorganisms.* – 2020. – V. 8, № 11. – P. 1714.
27. 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.
28. 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.
29. 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.
30. 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.
31. Rouzine I. M. Long-range linkage effects in adapting sexual populations // *Sci Rep.* – 2023. – V. 13, № 1. – P. 12492.
32. 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.
33. 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.
34. 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.
35. 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.
36. Rouzine I. M., Razoogy 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.