

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КубГУ»)

Факультет компьютерных технологий и прикладной математики
Кафедра математического моделирования

Допустить к защите
Заведующий кафедрой
Д-р физ.-мат. наук, акад. РАН, проф.
_____ В.А. Бабешко
_____ 2018 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
(БАКАЛАВРСКАЯ РАБОТА)

МАТЕМАТИЧЕСКИЕ ТЕХНОЛОГИИ МОДЕЛИРОВАНИЯ
ВИРУСНОЙ ДИНАМИКИ

Работу выполнила _____ А.А. Вербицкая

Направление подготовки 01.03.02 Прикладная математика и информатика

Направленность (профиль) Математическое моделирование и вычислительная математика: Математическое моделирование

Научный руководитель
д-р физ.-мат. наук, проф. _____ А.В. Павлова

Нормоконтролер
канд. физ.-мат. наук, доц. _____ С.Е. Рубцов

Краснодар
2018

РЕФЕРАТ

Выпускная квалификационная работа 57 с., 25 рис., 2 табл., 4 приложения, 22 источника.

ГЕНЕТИЧЕСКАЯ ЭВОЛЮЦИЯ ВИРУСОВ, ДИНАМИКА ВИРУСНЫХ ПОПУЛЯЦИЙ, МОДЕЛЬ КВАЗИВИДОВ, РЕКОМБИНАЦИЯ, МУТАЦИЯ, ОТБОР, ГЕНЕТИЧЕСКОЕ РАЗНООБРАЗИЕ, РАССТОЯНИЕ ХЭММИНГА, ДИФФЕРЕНЦИАЛЬНАЯ МОДЕЛЬ, СИСТЕМА ОДУ

Объектом исследования являются закономерности развития популяции вируса иммунодефицита человека (ВИЧ) в организме.

Целью работы является построение моделей динамики популяции квазивидов ВИЧ на основе различных подходов и исследование с их помощью разнообразия данной популяции.

В результате выполнения выпускной квалификационной работы построена стохастическая компьютерная модель, учитывающая точечные мутации, рекомбинации вирусных геномов и отбор потомков по величине функции приспособленности. Проведено исследование генетического разнообразия популяции с использованием расстояния Хэмминга. Построена также адаптированная дифференциальная модель на основе системы обыкновенных дифференциальных уравнений (ОДУ). Проведено исследование влияния концентрации противовирусного препарата на динамику развития популяции вируса в организме.

СОДЕРЖАНИЕ

Введение.....	4
1 Моделирование генетической эволюции вирусов. Обзор математических моделей.....	6
2 Стохастическая модель квазивидов	10
2.1 Рекомбинация	12
2.2 Мутация и размножение.....	13
2.3 Отбор	15
3 Результаты работы стохастической модели.....	16
4 Исследование генетического разнообразия популяции.....	20
5 Дифференциальная модель квазивидов.....	25
6 Адаптация дифференциальной модели	28
7 Результаты работы дифференциальной модели	32
7.1 Результаты работы модели без введения препарата	32
7.2 Результаты работы модели для случая применения противовирусной терапии	33
8 Описание приложения	38
8.1 Среда реализации и структура приложения.....	38
8.2 Графический интерфейс	38
Заключение	41
Список использованных источников	43
Приложение А. Словарь новых понятий	46
Приложение Б. Результаты вычислительных экспериментов на базе дифференциальной модели	47
Приложение В. Программа, реализующая стохастическую модель.....	49
Приложение Г. Программа, реализующая дифференциальную модель	56

ВВЕДЕНИЕ

Многообразие вариантов течения и исхода инфекционных заболеваний зависит от динамики взаимодействия возбудителей инфекции и иммунной системы организма [1]. Объем клинических и экспериментальных данных в иммунологии растет очень быстро вследствие значительных достижений в технологиях количественного анализа разнообразных аспектов иммунных реакций. Но увеличение объема данных само по себе не является достаточным условием для понимания механизмов, определяющих изменчивость вирусов, и прогнозирования динамики инфекционных агентов иммунных реакций. Математическое моделирование и использование современных вычислительных технологий может составить основу для анализа и интерпретации динамики иммунного ответа и эволюции вирусных популяций.

Существуют различные модели, объясняющие разнообразные явления, фиксируемые в ходе инфекции [2–4], но ни одна из них не учитывает всех эффектов, известных в клинической практике. Одна из центральных проблем понимания динамики вирусных инфекций человека связана с быстрой изменчивостью вирусов и отбором более приспособленных и устойчивых мутантов в контексте применяемых противовирусных препаратов.

Проблема резистентности вируса иммунодефицита человека (ВИЧ) имеет большую значимость, прежде всего, в силу влияния на эффективность противовирусной терапии [5]. Быстрое развитие лекарственной устойчивости остается основным препятствием для терапии ВИЧ. Вирус иммунодефицита человека (ВИЧ) обладает уникальной изменчивостью, отличаясь чрезвычайно высоким уровнем мутаций.

Для разработки эффективных подходов к терапии инфекции, вызванной ВИЧ, необходимо понять закономерности развития популяции вируса в организме.

Исследование механизмов, лежащих в основе способности ВИЧ к диверсификации генома и формированию лекарственной устойчивости, способствует созданию эффективных препаратов и вакцин против вируса иммунодефицита человека и разработке оптимальных стратегий лечения.

Группы исследователей работают со стохастическими моделями ВИЧ-инфекции, рассматривающими популяции клеток как взаимодействующие дискретные вероятностные ряды. Для описания динамики ВИЧ применяются также так называемые популяционные модели, построенные на основе определенных гипотез о взаимодействии рассматриваемых популяций.

Целью выпускной квалификационной работы является построение стохастической и дифференциальной моделей динамики квазивидов ВИЧ, учитывающих процессы, протекающие внутри вирусной популяции, а также зависимость параметров приспособленности вирусных геномов от концентрации препарата AZT, блокирующего обратную транскрипцию вирусной РНК в ДНК.

1 Моделирование генетической эволюции вирусов. Обзор математических моделей

Непрерывная генетическая и антигенная изменчивость вируса в человеческой популяции приводит к появлению новых вариантов вирионов ВИЧ, что резко усложняет проблему получения вакцины и затрудняет проведение специальной профилактики СПИД.

Высокая генетическая изменчивость, обуславливающая способность популяции адаптироваться к действию иммунного ответа и противовирусной терапии, является принципиальной особенностью инфекции ВИЧ [5]. Отбор наиболее приспособленных мутантов дает возможность популяции становиться резистентной к антивирусным препаратам. Теоретической основой для исследования этого явления может служить концепция квазивидов (или мутационно-селекционного баланса).

Одной из принципиальных особенностей популяции ВИЧ является ее высокая генетическая изменчивость, лежащая в основе способности популяции адаптироваться к действию противовирусной терапии. Быстрое развитие лекарственной устойчивости остается на сегодняшний день главным препятствием для лечения ВИЧ-инфекции. ВИЧ отличается чрезвычайно высоким уровнем мутаций. Понимание механизмов, лежащих в основе способности ВИЧ к диверсификации генома и формированию лекарственной устойчивости, способствует созданию эффективных препаратов и вакцин против вируса иммунодефицита человека и разработке оптимальных стратегий лечения.

В работе [5] используется компьютерная модель, имитирующая геномную диверсификацию ВИЧ в инфицированном организме, основные стадии работы предлагаемой модели: инициализация популяции вирионов, инфицирование клеток, обратная транскрипция (рекомбинация и мутация вирионов), определение разнообразия вирионов, трансляция, перемешивание и продуцирование вирусов, выбор устойчивых. В работе исследуется влияние

рекомбинации на процесс адаптации вирусной популяции к лекарственным препаратам. По мнению авторов, рекомбинация может способствовать накоплению мутаций и ускорить формирование устойчивости к проводимой антиретровирусной терапии ВИЧ [5].

В работе [6] сделан вывод о том, что способность вируса к рекомбинации, вопреки широко распространенному мнению, не всегда способствует развитию лекарственной устойчивости. Влияние рекомбинации на развитие лекарственной устойчивости зависит от интенсивности отбора, а также от численности вирусной популяции. Рекомбинация значительно ограничивает развитие лекарственной устойчивости для больших эффективных размеров вирусной популяции.

В численных моделях, реализованных в работах ряда исследователей, рассматривалась постоянная эффективная численность популяции, в то время как размер популяции ВИЧ изменяется в течение инфекции, и особенно при высокоактивной антиретровирусной терапии, когда численность может упасть более, чем на порядок.

В работе [6], предлагается генетическая модель репликации ВИЧ, использующая алфавит из четырех символов A, a, B, b , с помощью которых кодируется двухлокусное двухаллельное представление вириона и исследуется резистентность к лекарственному препарату. В модели учитываются повторные инфицирования (4 типа однократного и 10 типов двукратного инфицирования), рассчитываются частоты появления гомозиготных и гетерозиготных вирионов, определяются частоты провирусов будущего поколения на основе информации о вирусах текущей популяции с учетом мутаций и рекомбинаций. Используя результаты детерминированной и стохастической моделей для четырех типов провирусов, исследуется влияние рекомбинации на развитие лекарственной устойчивости. При этом авторы считают, что эволюция ВИЧ лучше описывается случайным процессом при переменной численности вирусной популяции.

Последние десятилетия характеризуются появлением лекарственных препаратов, создающих генетические барьеры на пути распространения ВИЧ-инфекции. В [2] построена математическая модель динамики ВИЧ под воздействием антиретровирусных препаратов для прогнозирования времени, необходимого для формирования мутировавших геномов ВИЧ (имеющих полный набор резистентных мутаций), способных преодолеть генетический барьер препарата.

В работе [7] с целью прогноза возникновения лекарственной устойчивости у пациентов, перенесших антиретровирусную терапию, изучаются уже существующие накопления полезных аллелей в популяции гаплоидных геномов.

Авторы [3, 4] Burke, Ramsey, De Jong, Grefenstette считают, что большинство современных генетических алгоритмов, не вполне соответствуют правдоподобным моделям с точки зрения биологии. Ими предлагается уточненная вычислительная модель вирусной эволюции на основе генетических алгоритмов. Чтобы приблизить модели к исследуемым биологическим объектам предлагается выполнить следующие условия: геном имеет переменную длину, биологически гены не зависят от позиции; геномы могут содержать не кодирующие участки; геномы могут содержать дублирующие или соперничающие гены; геномы имеют перекрывающиеся участки. Для идентификации геномов в [3, 4], кроме символов A, C, T, G, алфавит содержит коды START и STOP. Работа модели включает три стадии: сканирование генома; расчет функций совместимости генов с целью выявления активных; определение устойчивости генома, исходя из совместимости активных генов.

По мнению авторов [4], большой интерес представляют эволюционные алгоритмы для геномов переменной длины. Исследования виртуальных вирусов различной длины генома с помощью численной модели, основанной на генетических алгоритмах, позволили выявить интересные результаты,

например, длина особей в популяции самостоятельно адаптируется непосредственно в ответ на мутации.

2 Стохастическая модель квазивидов

Термин «стохастический» описывает процесс или параметр, определяемые случайным образом. Впервые термин был введен для описания математических объектов, называемых стохастическими процессами, но сейчас в математике термины «стохастический процесс» и «случайный процесс» считаются взаимозаменяемыми. Слово первоначально произошло от древнегреческого «στόχος», что означает «цель, предположение».

В теории вероятностей и смежных областях стохастический или случайный процесс является математическим объектом, обычно определяемым как совокупность случайных величин. Примерами таких систем могут являться: рост популяции бактерий, электрический ток, колеблющийся из-за теплового шума, или движение молекулы газа. Стохастические процессы широко используются в качестве математических моделей систем и явлений, которые, как представляется, изменяются случайным образом. Они находят применение в различных областях знаний, включая такие как биология, химия, экология и физика, а также технологии и инженерные области (изображений, теория информации, криптография, телекоммуникации и др.).

В данной работе реализована стохастическая модель динамики квазивидов ВИЧ с использованием генетических алгоритмов.

Генетический алгоритм (ГА) – это эвристический алгоритм, который используется для решения задач оптимизации и моделирования путем комбинирования, случайного подбора и вариации искоемых параметров и использует механизмы, аналогичные естественному отбору в природе. ГА является разновидностью эволюционных вычислений, с помощью которых решаются оптимизационные задачи с использованием методов естественной эволюции, таких как наследование, рекомбинации, мутации и отбор. Отличительной особенностью генетического алгоритма является акцент на использование оператора «скрещивания», который производит операцию

рекомбинации, роль которой аналогична роли скрещивания в живой природе [8].

Будем называть виртуальным геномом последовательность символов, которая кодирует некоторый участок РНК вируса иммунодефицита человека [9]. Далее будет использован четырехбуквенный алфавит, каждая буква которого соответствует нуклеотиду: А – аденин, G – гуанин, Т – тимин, С – цитозин. В модели виртуальный ген является аналогом нити вирусной РНК, а виртуальная популяция (вектор пар виртуальных генов) – аналог вирусных частиц [10,11]. Общая схема генетического алгоритма приведена на рисунке 1.

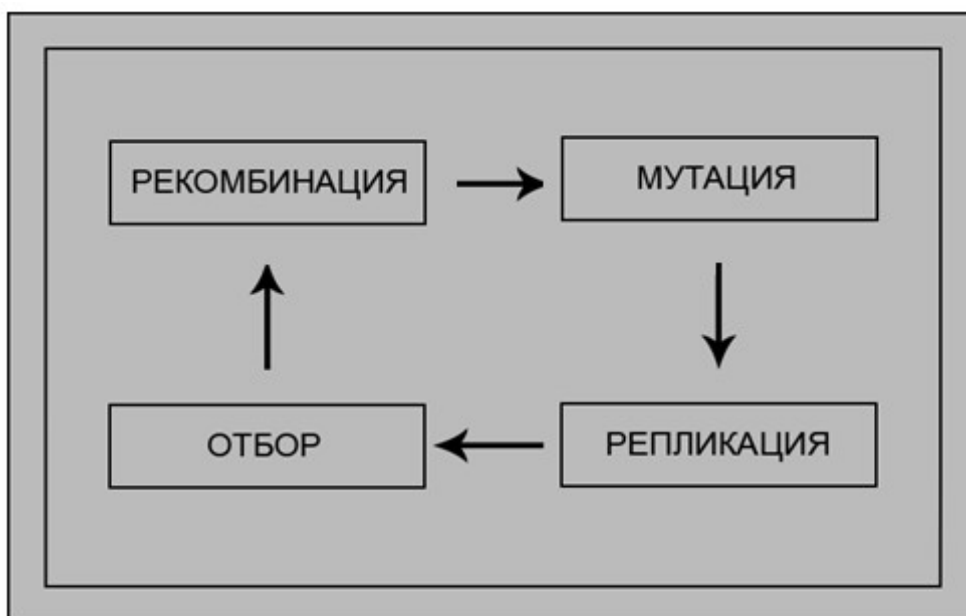


Рисунок 1 – Схема генетического алгоритма

Рассматриваемая популяция подразделяется на восемь классов [9–11]:

- WT (дикий тип – в 41 и 215 позициях находятся метионин (ATG) и треонин (ACC) соответственно, т.е. в позициях 121–123 содержится комбинация символов ATG, а в позициях 643–645 – комбинация символов ACC);
- M41L (в 41 позиции последовательность TTG);
- T215N (в 215 позиции последовательность AAC);
- T215S (в 215 позиции последовательность TCC);

- T215Y (в 215 позиции последовательность TAC);
- M41L/T215N (в 41 позиции последовательность TTG, в 215 позиции последовательность AAC);
- M41L/T215S (в 41 позиции последовательность TTG, в 215 позиции последовательность TCC);
- M41L/T215Y (в 41 позиции последовательность TTG, в 215 позиции последовательность TAC).

Размер популяции – 1000 вирусных цепочек, не меняется в ходе программы. Длина генома – 1800.

Начальная популяция состоит из вирусных мутантов дикого типа. При инициализации популяции символы в позициях вирусных цепочек, отличных от 121–123 и 643–645, генерируются случайным образом. Начальная популяция состоит из одинаковых вирусных цепочек.

2.1 Рекомбинация

В генетических алгоритмах и эволюционных вычислениях кроссовер, также называемый рекомбинацией, является генетическим оператором, используемым для объединения генетической информации двух родителей с целью генерации нового потомства. Это один из способов стохастически генерировать новые геномы в существующей популяции. Он аналогичен кроссоверу, который происходит во время полового размножения в биологии [8].

В данной модели рекомбинация представляется следующим образом: в процессе обратной транскрипции происходит перенос генов с одной цепочки РНК на другую крест-накрест (рисунок 2). При этом цепочки-родители, участвовавшие в рекомбинации, заменяются своими потомками.

В рассматриваемой модели вероятность рекомбинации двух вирусных цепочек принята равной 0,5. Для каждой пары цепочек берется случайное

число от 0 до 1, и если оно меньше или равно 0,5, то происходит рекомбинация.

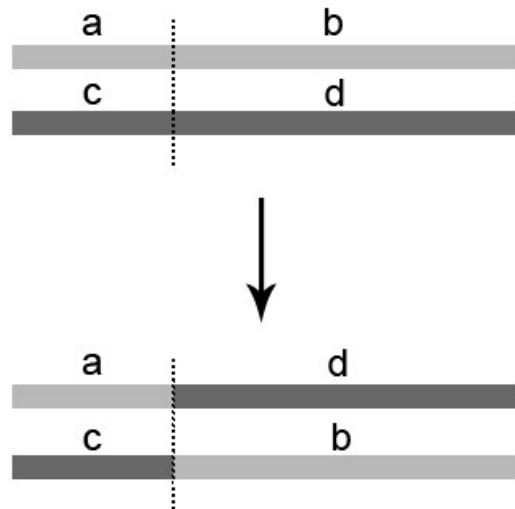


Рисунок 2 – Схема рекомбинации

Точка разрыва выбирается случайным образом. На текущем этапе моделируется разрыв только в одной позиции. Такая реализация соответствует однократному инфицированию.

2.2 Мутация и размножение

Следующий этап генетического алгоритма – мутация.

Мутация – это генетический оператор, используемый для сохранения генетического разнообразия от одного поколения популяции до следующего и является аналогом биологической мутации. Мутация изменяет значения одного или нескольких генов в геноме и происходит в соответствии с определяемой в модели мутационной вероятностью. Эта вероятность должна быть установлена на низком уровне.

В данной модели вероятность мутации принята равной 0,2. Как и в случае рекомбинации, мутация будет происходить, если случайное число от 0 до 1 оказывается меньше или равно 0,2. Позиция, в которой будет происходить мутация, выбирается случайным образом.

Для моделирования вероятности мутации использованы оригинальные данные, полученные в ходе биологических экспериментов и использованные в работе [11]. Скорости мутаций приведены в таблице 1.

Таблица 1. Скорости мутаций

A → C	0,543
A → G	2,277
A → T	0,319
C → A	1,858
C → G	0,228
C → T	4,601
G → A	7,361
G → C	0,262
G → T	0,655
T → A	0,81
T → C	2,809
T → G	0,58

Для реализации мутации генома в модели был адаптирован метод рулетки. Скорости мутации для каждого основания были расположены на колесе рулетки так, что размер сектора, соответствующего каждому основанию, пропорционален скорости мутации в данное основание. На рисунке 3 изображены рулетки для расчета вероятностей мутаций каждого из рассматриваемых в модели типов оснований.

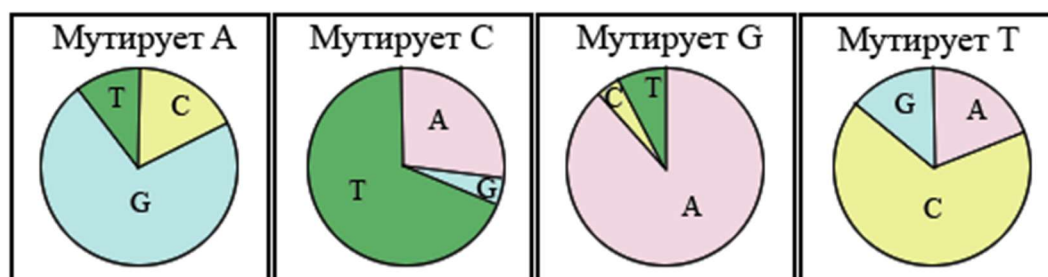


Рисунок 3 – Рулетка для расчета вероятности мутации оснований

Процесс размножения популяции на данном этапе реализуется как умножение на некоторое число (20) количества генов всех типов, присутствующих в популяции.

2.3 Отбор

Отбор – это этап генетического алгоритма, в котором отдельные геномы выбираются из популяции для следующей итерации.

Отбор генов в новую популяцию производился по известным значениям функции приспособленности, приведенным в таблице 2, с помощью метода рулетки. Экспериментальные данные были взяты из работы [12].

Таблица 2. Значения функции приспособленности

Концентрация AZT μ M	дикий тип WT	Мутация вида M41L	Мутация вида T215N	Мутация вида T215S	Мутация вида T215Y	Мутация вида M41L+ T215N	Мутация вида M41L+ T215S	Мутация вида M41L+ T215Y
0	1	0,604	0,203	0,253	0,704	0,291	0,493	0,782
0,03	0,56	0,509	0,096	0,1	0,606	0,124	0,393	0,617
0,3	0,162	0,166	0,0119	0,016	0,293	0,038	0,116	0,385

Расчеты проводились для двух различных значений концентрации противовирусного препарата AZT μ M равной 0 и 0,3.

Размер сектора рулетки для каждого квазивида выбирался пропорциональным его устойчивости к AZT μ M, т.е. вероятность цепочки быть отобранной в новую популяцию тем больше, чем выше ее резистентность к AZT μ M (или приспособленность – при отсутствии противовирусной терапии).

3 Результаты работы стохастической модели

Для стохастической модели расчеты были проведены для заданного количества поколений – 1000. В каждом поколении информация о том, геномы какого вида и в каком количестве присутствуют в популяции, сохранялась в массиве. Для иллюстрации полученных данных были построены графики функций, описывающих изменение численности геномов от поколения к поколению.

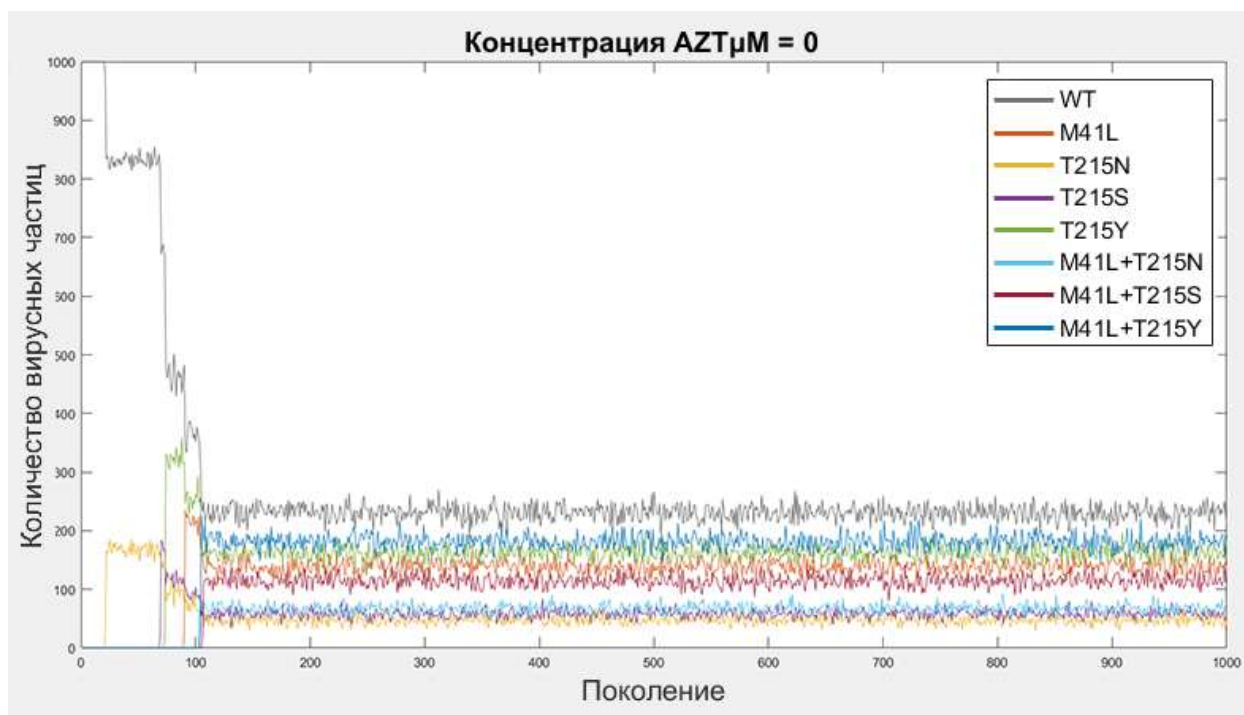


Рисунок 4 – Количество вирусных частиц разных квазивидов при концентрации AZT μ M = 0

График, приведенный на рисунке 4, иллюстрирует изменения, происходящие в составе популяции со сменой поколений.

Можно видеть, как в определенных поколениях в результате мутации или рекомбинации появляются вирусные цепочки квазивидов, которые ранее не присутствовали в популяции. После появления всех рассматриваемых квазивидов их соотношение в популяции с течением времени изменяется незначительно. Изменения происходят за счет рекомбинации и мутации, которые могут изменить принадлежность вирусных частиц к тому или иному

типу, и элемента случайности при отборе в новую популяцию. Доминирующим остается дикий тип, имеющий наибольшее значение функции приспособленности.

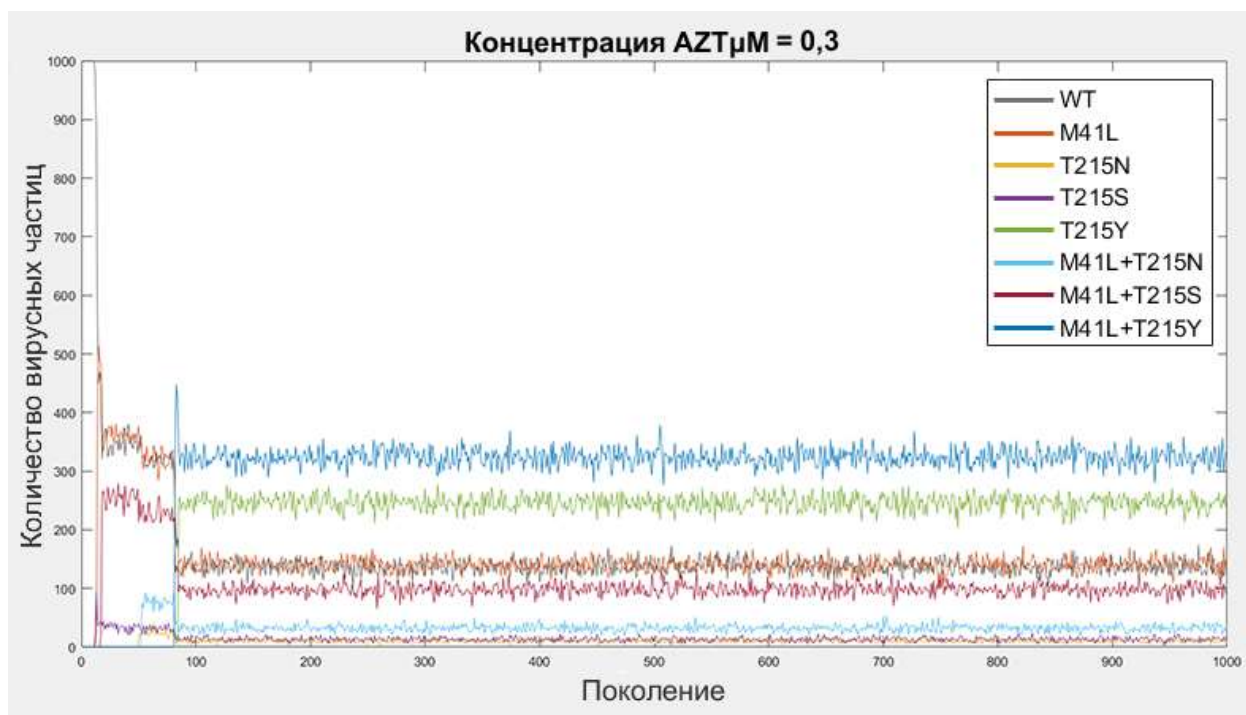


Рисунок 5 – Количество вирусных частиц разных квазивидов при концентрации $AZT_{\mu M} = 0,3$

Как и в случае отсутствия противовирусной терапии (концентрация препарата $AZT_{\mu M}=0$), наблюдается зависимость количества вирусных цепочек каждого рассматриваемого вида в популяции от его сопротивляемости $AZT_{\mu M}$. Как и в отсутствие лекарства, результатом эволюции является разнородная вирусная популяция.

В силу случайной природы происходящих в популяции процессов, графики, показанные на рисунках 4 и 5, имеют скачкообразный вид. Чтобы сгладить их, алгоритм повторялся многократно, после чего результирующие значения были усреднены. Среднее рассчитывалось по 100 проходам. Полученные результаты отражены на рисунках 6 и 7.

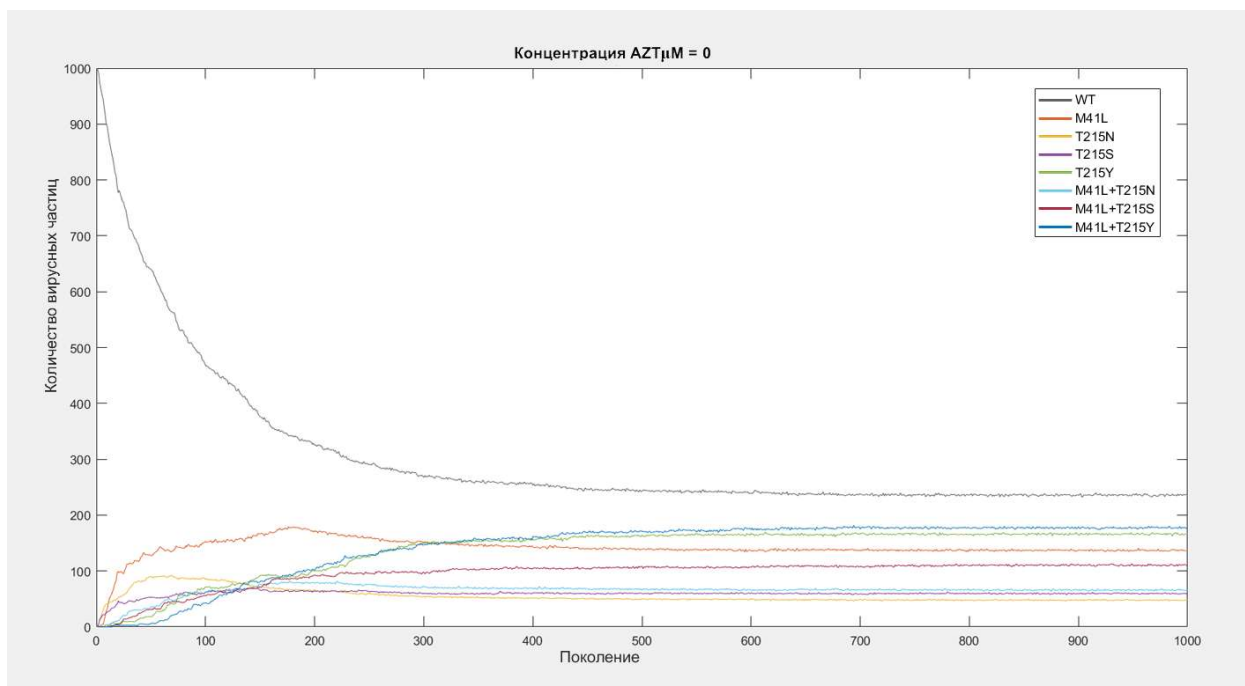


Рисунок 6 – Осредненные значения при концентрации $AZT_{\mu M} = 0$

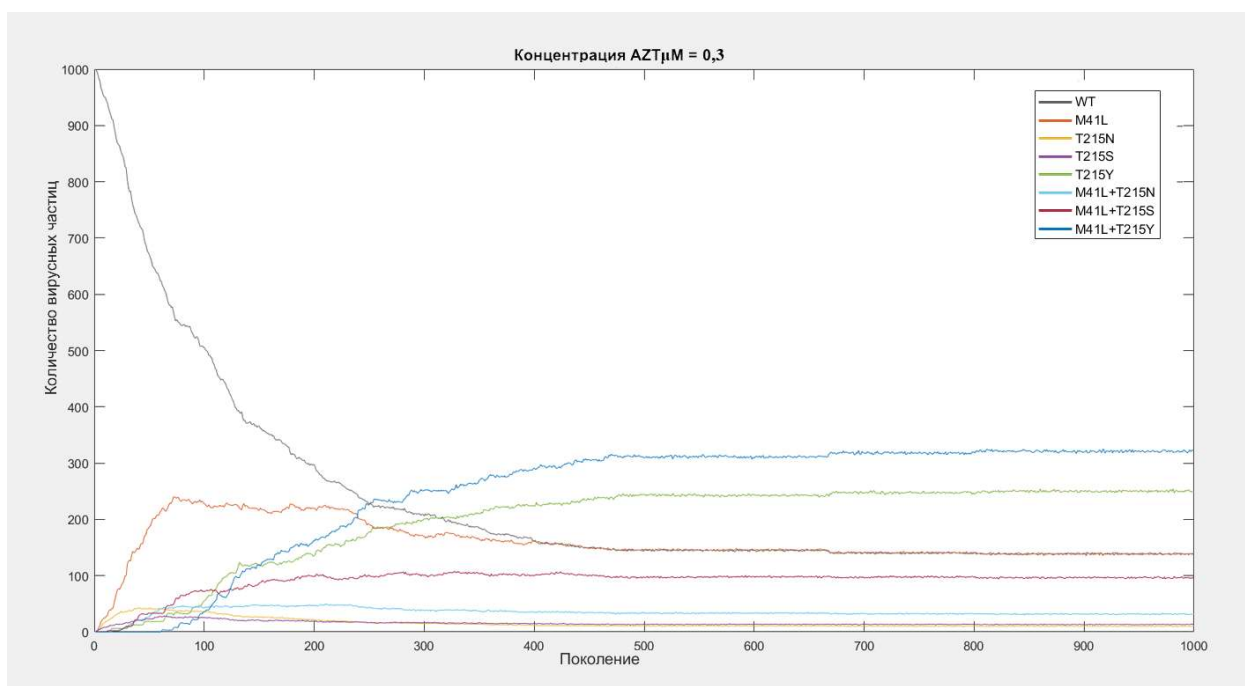


Рисунок 7 – Усредненные значения при концентрации $AZT_{\mu M} = 0,3$

Без введения противовирусной терапии численность видов мутантов устанавливается примерно за 450 циклов. При этом доминирующим оказывается исходный WT (дикий тип), который характеризуется

наибольшим значением функции приспособленности в данных условиях.

При введении противовирусной терапии численность квазивидов в популяции устанавливается только около 500 цикла, доминирующим уже является квазивид M41L+T215Y, за ним следует T215Y. Наименьшую численность имеют мутанты T215N и T215S, которым соответствуют наименьшие значения функции приспособленности для данной концентрации препарата.

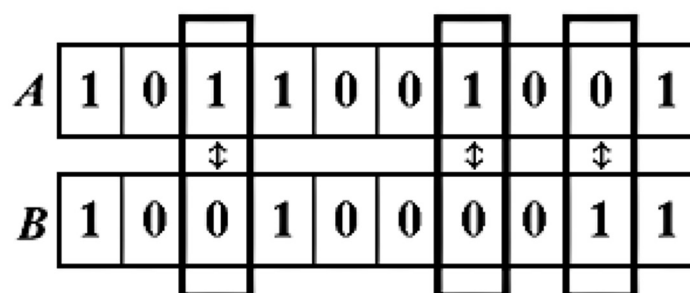
Как было сказано выше, и при введении лекарства, и в его отсутствие результатом эволюции является разнородная вирусная популяция, меняется лишь соотношение между числом цепочек различных типов в пользу наиболее приспособленных (резистентных) мутантов.

4 Исследование генетического разнообразия популяции

Для определения степени разнообразия популяции использована характеристика, называемая расстоянием Хэмминга.

В теории информации расстояние Хэмминга между двумя строками равной длины – это число позиций, в которых соответствующие символы различны. Другими словами, оно измеряет минимальное количество подстановок, необходимых для преобразования одной строки в другую, или минимальное количество ошибок, которые могли бы преобразовать одну строку в другую. В более общем контексте расстояние Хэмминга является одной из нескольких строковых метрик для измерения расстояния редактирования между двумя последовательностями [13].

Принцип подсчета отражен на рисунке 8.



Расстояние Хэмминга = 3

Рисунок 8 – Расстояние Хэмминга

Таким образом, для двух оснований расстояние Хэмминга определяется следующим образом:

$$d(b_l, b_m) = \begin{cases} 0, & b_l = b_m, \\ 1, & b_l \neq b_m. \end{cases}$$

Для двух геномов

$$g_i = (b_i^1, \dots, b_i^L),$$

$$g_j = (b_j^1, \dots, b_j^L),$$

$$D(g_i, g_j) = \sum_{l=1}^L d(b_i^l, b_j^l).$$

Среднее расстояние для популяции:

$$D = \frac{\sum_{j=1}^N \sum_{i=j+1}^N D(g_i, g_j)}{\frac{N(N-1)}{2}},$$

где L – длина генома, N – численность популяции.

Полный перебор популяции не рационален и очень затратен по времени и ресурсам, поэтому было произведено усовершенствование алгоритма, что позволило значительно увеличить скорость выполнения программы.

На первом шаге сравним все первые гены цепочек во всей популяции. За один проход просуммируем количества различных нуклеотидов, причем:

$$N = \sum N_i,$$

где N_i – количество i -го нуклеотида в первых позициях цепочек, N – численность популяции.

Затем посчитаем количество пар геномов с одинаковыми нуклеотидами, отдельно для каждого из видов нуклеотидов, используя формулу для числа сочетаний из N_i по 2:

$$C_{N_i}^2 = \frac{(N_i)!}{2(N_i - 2)!},$$

где N_i – количество i -го нуклеотида.

Полученные числа просуммируем и вычтем из общего количества возможных пар геномов, полученного по формуле:

$$C_N^2 = \frac{(N)!}{2(N - 2)!},$$

где N – численность популяции.

Полученное число будет количеством пар геномов с различными нуклеотидами в паре.

Проведем эти операции для всех 1800 рядов и просуммируем на каждом этапе полученные количества пар. В итоге получим количество различных пар геномов во всей популяции.

Данный алгоритм значительно быстрее, т.к. в результате мы совершаем всего лишь один проход по всей популяции, а не полный перебор всех вариантов.

В каждом поколении было посчитано среднее по популяции расстояние Хэмминга, эта информация сохранялась в массиве. Так как начальная популяция состоит из одинаковых вирусных цепочек дикого типа, для нее это значение было равно нулю. Полученные данные отображены на рисунках 9 и 10.

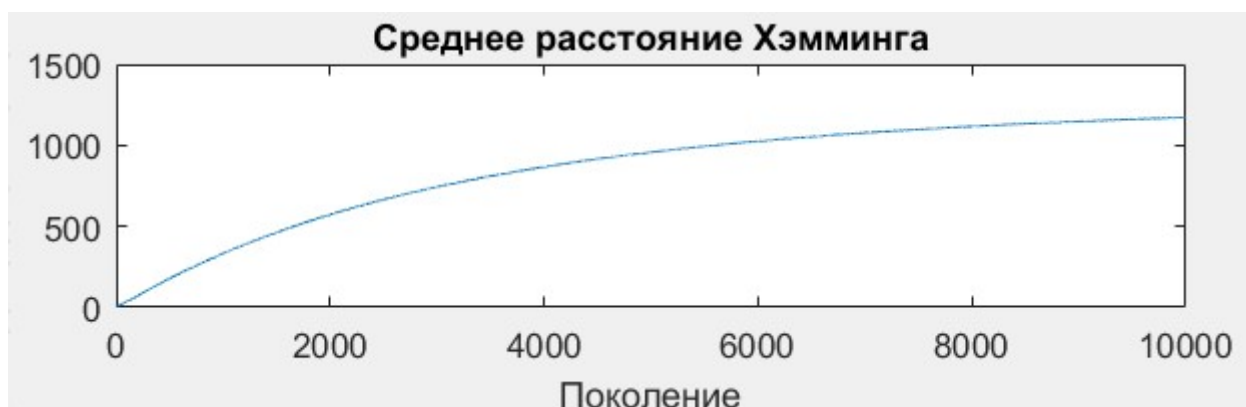


Рисунок 9 – Расстояние Хэмминга при концентрации $AZT_{\mu M} = 0$

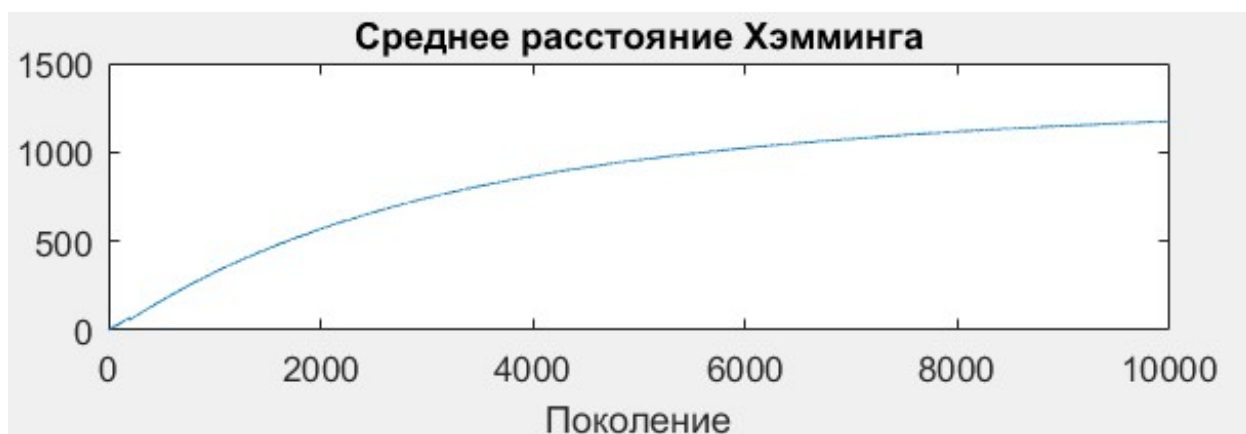


Рисунок 10 – Расстояние Хэмминга при концентрации $AZT_{\mu M} = 0,3$

Таким образом, можно наблюдать наиболее явный рост графика в поколениях 0–4000, после чего он замедляется и в поколениях 6000–10000 меняется слабо.

Также был проведен подсчет среднего расстояния Хэмминга по популяции, учитывающий только те позиции в геноме, которые влияют на определение квазивида цепочек. Полученные данные отображены на рисунках 11 и 12.

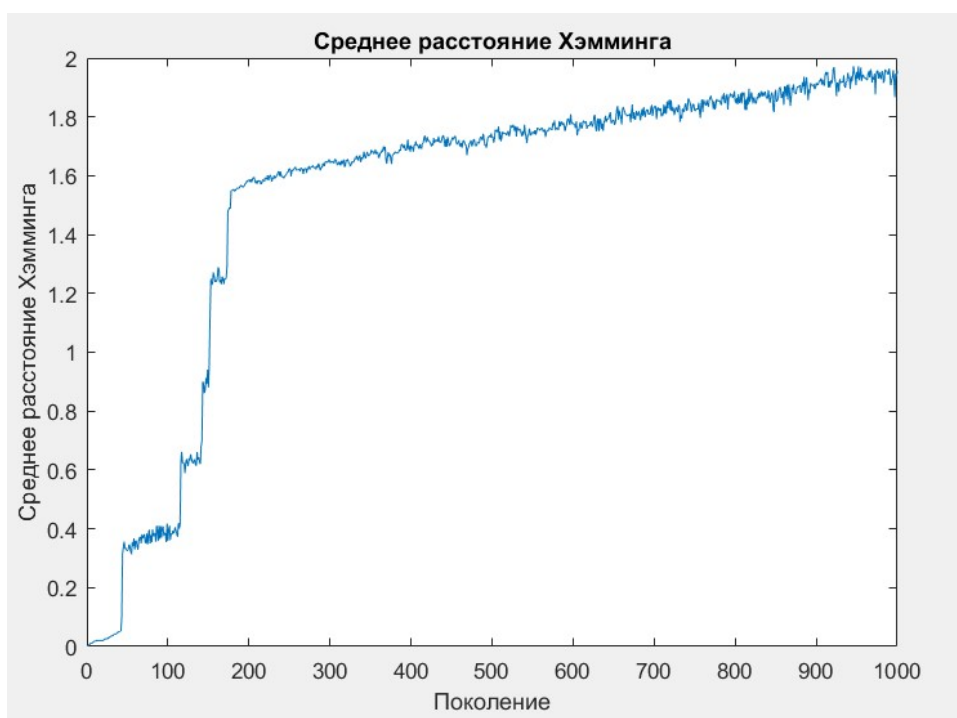


Рисунок 11 – Расстояние Хэмминга при концентрации AZT μ M = 0

Количественное описание степени генетической неоднородности вирусной популяции, оцениваемое с помощью рассчитанного расстояния Хемминга, соответствует выводу о разнородности эволюционирующей популяции, а не абсолютному вытеснению доминирующим типом всех остальных.

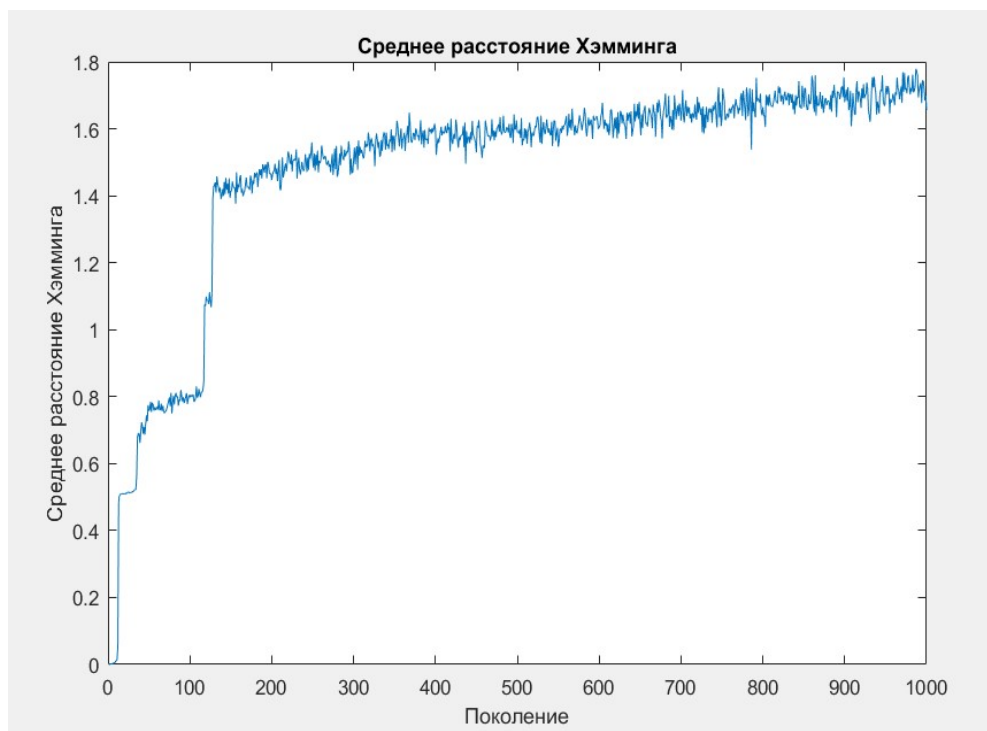


Рисунок 12 – Расстояние Хэмминга при концентрации AZT μ M = 0,3

На данном этапе построения модели не наблюдается существенной зависимости изменения среднего расстояния Хэмминга от концентрации препарата AZT μ M.

5 Дифференциальная модель квазивидов

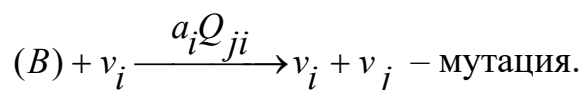
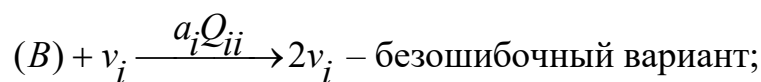
Одна из наиболее известных моделей квазивидов – модель, описывающая достаточно простую эволюцию полинуклеотидных последовательностей. Характер эволюции существенно зависит от численности популяции. Если последняя велика, эволюция может рассматриваться как детерминированный процесс. В этом случае эволюционная динамика популяции может быть описана системой обыкновенных дифференциальных уравнений (ОДУ).

Предположим, что существует n различных последовательностей нуклеиновых кислот v_1, v_2, \dots, v_n . Каждая разновидность (вариант) характеризуется определенной последовательностью нуклеотидов, которая определяет скорость репликации данной разновидности.

Обозначим скорости репликации разновидностей v_1, v_2, \dots, v_n как A_1, A_2, \dots, A_n соответственно. В отсутствие мутации разновидность с самой высокой скоростью репликации будет расти быстрее и достигнет размера популяции. Результатом отбора (без ошибок) является однородная популяция, состоящая из генотипа, обладающего наибольшей скоростью репликации.

Но репликация не лишена ошибок. Таким образом, необходимо определить вероятности Q_{ij} того, что на очередном шаге репликации цепочка РНК вида v_j в результате мутации станет цепочкой вида v_i . Вероятности Q_{ij} для $i=1, 2, \dots, n$ и $j=1, 2, \dots, n$ формируют так называемую «матрицу мутации».

Репликации квазивидов могут быть представлены следующими реакциями [14]:



Символ B обозначает материалы с низким молекулярным весом (это четыре нуклеотида), которые требуются для синтеза РНК или ДНК. Предполагается, что доступное количество (нужный объём) B является постоянным и, следовательно, не будет использован в кинетических дифференциальных уравнениях. Без ошибок репликации и мутации – это параллельные реакции одного и того же механизма. Вероятности мутации Q_{ij} влияют на продукт репликации. Коэффициент a_i описывает скорость репликации генома с номером i .

Система обыкновенных дифференциальных уравнений, описывающая развитие (эволюцию) во времени популяции этих последовательностей нуклеиновых кислот, может быть записана в виде [14]

$$\dot{v}_1 = a_1 Q_{11} v_1 + a_2 Q_{12} v_2 + \dots + a_n Q_{1n} v_n. \quad (1)$$

Точкой в (1) и далее обозначена производная по времени. Новые частицы разновидности (варианта) v_1 могут быть образованы в результате безошибочной репликации v_1 . Это происходит при скорости a_1 и вероятности Q_{11} , поэтому общая скорость воспроизводства цепочек вида v_1 – $a_1 Q_{11} v_1$. Ошибочная репликация любых других мутантов v_2, \dots, v_n может также привести к появлению новых вирусных частиц v_1 . Это представлено ростом темпов в уравнении (1). Таким же образом можно записать скорость производства любых других разновидностей и получить систему дифференциальных уравнений:

$$\dot{v}_i = \sum_{j=1}^n a_j Q_{ij} v_j, \quad i = 1, \dots, n. \quad (2)$$

Вероятность мутации определяется по формуле

$$Q_{ij} = p_{ij}^{H_{ij}} (1 - p)^{(1-H_{ij})}.$$

Здесь p – скорость мутации в расчёте на одно основание, H_{ij} – это расстояние Хэмминга между штаммами i и j , то есть количество оснований, в котором две линии отличаются. Расстояние Хемминга определяется формулами, приведенными в предыдущем разделе.

Без ошибок репликация задаётся формулой

$$Q_{ii} = (1-p)^i.$$

Популяция не будет состоять только из наиболее быстро растущей последовательности, напротив, она будет состоять из целого ансамбля мутантов с различными скоростями репликации. Это ансамбль мутантов – квазивидов.

Частота появления того или иного варианта внутри квазивидов зависит не только от его имплекативного значения, но также и от вероятности, с которой он производится ошибочной репликацией других шаблонов и их частоты в распределении квазивидов. Это важно для понимания структурной организации квазивидов.

Квазивиды сами по себе являются целью отбора в процессе воспроизводства популяции. Этот факт имеет большое значение. Эволюция, как правило, рассматривается как взаимодействие между процессами мутации и отбора. Отбор – фактор, который способствует преимуществам мутантов, которые были собраны по чистой случайности.

В рамках теории квазивидов, использующей аппарат обыкновенных дифференциальных уравнений, можно эффективно исследовать влияние мутаций на генетическую эволюцию вирусных популяций [14, 15].

6 Адаптация дифференциальной модели

Характер популяционной эволюции существенно зависит от численности популяции n . Если n очень велико, то численности особей каждого вида можно рассматривать как большие числа. В этом случае эволюционная динамика популяции может быть описана системой обыкновенных дифференциальных уравнений.

Будем называть виртуальным геномом последовательность символов, которая кодирует некоторый участок РНК вируса иммунодефицита человека [9]. Далее будет использован четырехбуквенный алфавит, каждая буква которого соответствует нуклеотиду: А – аденин, Г – гуанин, Т – тимин, С – цитозин. В модели виртуальный ген является аналогом нити вирусной РНК, а виртуальная популяция (вектор пар виртуальных генов) – аналог вирусных частиц.

Как и в работе [9], рассматриваются вирусные мутанты, резистентные к действию препарата (AZT), блокирующего обратную транскрипцию вирусной РНК в ДНК [16], ограниченного спектра: в 41 и 215 позициях аминокислотной последовательности находятся метионин (ATG) и треонин (ACC) соответственно, т.е. вирусы начальной популяции позициях 121–123 содержат комбинацию символов ATG, а в позициях 643–645 – комбинацию символов ACC.

Рассматриваемая популяция подразделяется на восемь классов:

WT (дикий тип – в 41 и 215 позициях находятся метионин (ATG) и треонин (ACC) соответственно, т.е. в позициях 121–123 содержится комбинация символов ATG, а в позициях 643–645 – комбинация символов ACC);

M41L (в 41 позиции последовательность TTG);

T215N (в 215 позиции последовательность AAC);

T215S (в 215 позиции последовательность TCC);

T215Y (в 215 позиции последовательность TAC);

M41L/T215N (в 41 позиции последовательность TTG, в 215 позиции последовательность AAC);

M41L/T215S (в 41 позиции последовательность TTG, в 215 позиции последовательность TCC);

M41L/T215Y (в 41 позиции последовательность TTG, в 215 позиции последовательность TAC).

Рассмотрим модель следующего вида, предложенную в работе [14]:

Уравнения теории квазивидов имеют вид:

$$\frac{d\mathbf{v}}{dt} = \mathbf{W}\mathbf{v} - d(\mathbf{v})\mathbf{v}. \quad (3)$$

Вектор \mathbf{v} содержит в качестве компонентов численности генотипов, составляющих популяцию $\mathbf{v} = \{v_1, v_1, \dots, v_n\}$, матрица \mathbf{W} имеет вид

$$\mathbf{W} = \begin{pmatrix} a_1 Q_{11} & a_2 Q_{12} & \cdots & a_n Q_{1n} \\ a_1 Q_{21} & a_2 Q_{22} & \cdots & a_n Q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_1 Q_{n1} & a_2 Q_{n2} & \cdots & a_n Q_{nn} \end{pmatrix}. \quad (4)$$

Здесь $v_i(t)$ – численность i -го квазивида в вирусной популяции, где $i = 1$ соответствует дикому типу (WT), $i = 2$ – мутанту M41L, $i = 3$ – T215N, $i = 4$ – T215S, $i = 5$ – T215Y, $i = 6$ – M41L+T215N, $i = 7$ – M41L+T215S, $i = 8$ – M41L+T215Y; a_i – скорость репликации i -го квазивида; Q_{ij} – вероятность того, что мутант j -го типа станет мутантом i -го типа.

Второе слагаемое в правой части (3) вводится для сохранения постоянства численности популяции

$$d(\mathbf{v}) = \frac{\sum_{i=1}^n a_i v_i}{\sum_{i=1}^n v_i}. \quad (5)$$

Для различной численности популяции система решалась при следующих начальных условиях:

$$v_1(0) = 10000, v_i = 0, i = \overline{2,8}, \quad (6)$$

что соответствует в начальный момент времени однородной популяции, состоящей из дикого типа;

$$v_i(0) = 1250, i = \overline{1,8}, \quad (7)$$

что соответствует разнородной популяции, содержащей в начальный момент времени одинаковое количество геномов каждого типа.

В качестве значений скоростей репликации геномов a_j ($j = \overline{1,8}$) могут быть взяты экспериментальные данные из работы [12], приведенные в таблице 2.

Вероятности переходов от одного типа мутанта к другому Q_{ij} оцениваются следующим образом. Вероятность замены одного основания на другое в конкретной позиции $p = 2 \cdot 10^{-5}$ на один нуклеотид [9, 11]. Так как используется четырехбуквенный код (A, C, G, T), то каждое основание может быть заменено одним из трех других. И вероятность того, что мутация произойдет в заданной позиции, равна $3p$. Вероятность того, что ни в одной из позиций (41-й или 215-й) ни одно из кодирующих оснований не изменится, равна $Q_{ii} = (1 - 3p)^6$. Вероятность конкретной мутации в одном из шести рассматриваемых оснований равна $Q_{ij} = p(1 - 3p)^5$. Если для перехода от одного мутанта к другому необходимо более двух мутаций, считается, что вероятность такого перехода равна нулю. Далее для всех моделей использованы приведенные значения вероятностей переходов от одного типа мутанта к другому. С учётом этих условий получаем преобразованную модель в следующем виде:

$$\dot{v}_i = F_i(\mathbf{v}), i = \overline{1,8},$$

$$\begin{aligned}
F_1(\mathbf{v}) &= a_1 Q_{11} v_1 + a_2 Q_{21} v_2 + a_3 Q_{31} v_3 + a_4 Q_{41} v_4 - v_1 d(\mathbf{v}), \\
F_2(\mathbf{v}) &= a_1 Q_{12} v_1 + a_2 Q_{22} v_2 + a_6 Q_{62} v_6 + a_7 Q_{72} v_7 - v_2 d(\mathbf{v}), \\
F_3(\mathbf{v}) &= a_1 Q_{13} v_1 + a_3 Q_{33} v_3 + a_5 Q_{53} v_5 + a_6 Q_{63} v_6 - v_3 d(\mathbf{v}), \\
F_4(\mathbf{v}) &= a_1 Q_{41} v_1 + a_4 Q_{44} v_4 + a_5 Q_{54} v_5 + a_7 Q_{74} v_7 - v_4 d(\mathbf{v}), \\
F_5(\mathbf{v}) &= a_3 Q_{35} v_3 + a_4 Q_{45} v_4 + a_5 Q_{55} v_5 + a_8 Q_{85} v_8 - v_5 d(\mathbf{v}), \\
F_6(\mathbf{v}) &= a_2 Q_{26} v_2 + a_3 Q_{36} v_3 + a_6 Q_{66} v_6 + a_8 Q_{86} v_8 - v_6 d(\mathbf{v}), \\
F_7(\mathbf{v}) &= a_2 Q_{27} v_2 + a_4 Q_{47} v_4 + a_7 Q_{77} v_7 + a_8 Q_{87} v_8 - v_7 d(\mathbf{v}), \\
F_8(\mathbf{v}) &= a_5 Q_{58} v_5 + a_6 Q_{68} v_6 + a_7 Q_{78} v_7 + a_8 Q_{88} v_8 - v_8 d(\mathbf{v}).
\end{aligned} \tag{8}$$

Если считать концентрацию противовирусного препарата в организме пациента постоянной, (8) является системой ОДУ с постоянными коэффициентами. Для исследования динамики популяции следует задать начальные условия – размеры численности каждого квазивида до начала терапии.

Далее система (8) решается для случая изначально однородной популяции при условиях (6) и разнородной популяции при условиях (7).

На основе численных экспериментов, изменяя начальные условия и параметры модели (коэффициенты a_i зависят от условий среды, в которой находится популяция), можно изучить влияние биологических характеристик системы на ее динамику.

Для численного решения системы уравнений использовался алгоритм ode15s для решения систем ОДУ [17–19], реализованный в пакете MATLAB. Это многошаговый метод переменного порядка (от 1 до 5, по умолчанию 5), использующий формулы численного дифференцирования. Решением соответствующей начальной задачи являются функции, описывающие эволюцию численности популяций соответствующих мутантов.

7 Результаты работы дифференциальной модели

Упрощающее анализ предположение о постоянстве концентрации препарата AZT в организме человека не вполне реалистично. В действительности, после приёма препарата его концентрация резко возрастает, а потом, по мере выведения лекарства из организма, падает, что, в свою очередь, отражается на скорости репликации мутантов и динамике изменения их численности, однако на данном этапе моделирования принято допущение о постоянстве концентрации противовирусного препарата.

7.1 Результаты работы модели без введения препарата

Изменение численности восьми типов геномов, составляющих популяцию, без введения противовирусного препарата показано на рисунке 13. В данном случае вычисления проводились для начальных данных (6), что соответствует начальной популяции, состоящей из дикого типа.

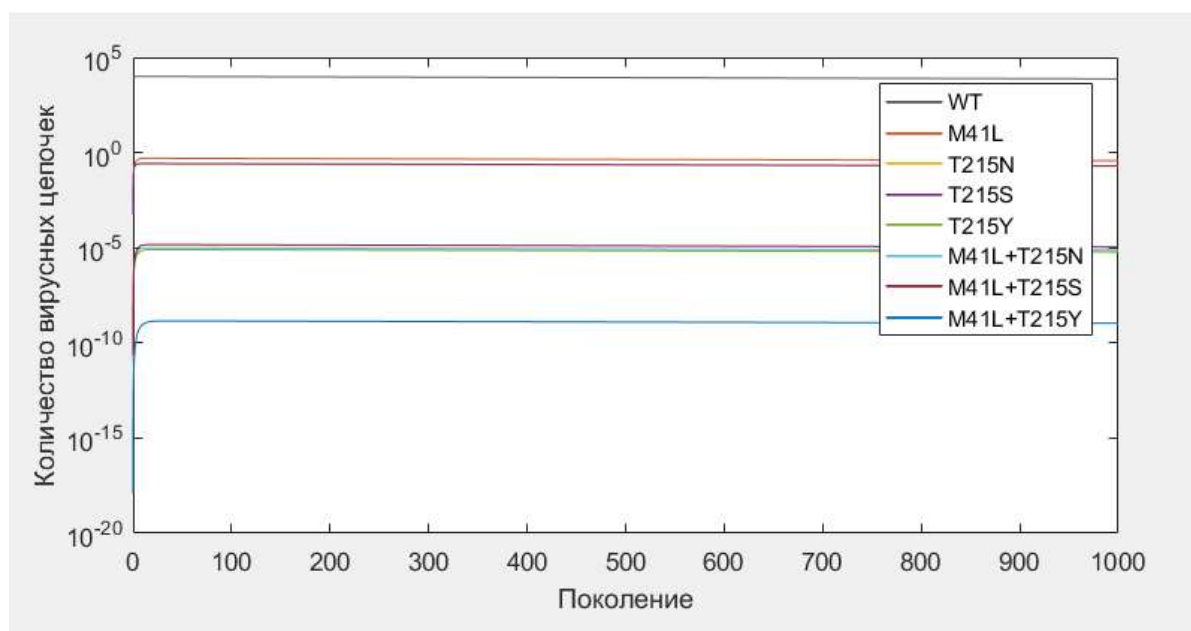


Рисунок 13 – Численность квазивидов при концентрации $AZT_{\mu M} = 0$

Из графика видно, что без применения препарата численность

квазивидов в популяции устанавливается менее чем за 100 поколений. При этом доминирующим является дикий тип, имеющий наибольшее значение функции приспособленности.

На рисунке 14 представлено изменение численности квазивидов в вирусной популяции для начального условия (7), то есть когда в начальный момент времени численности всех типов в популяции равны.

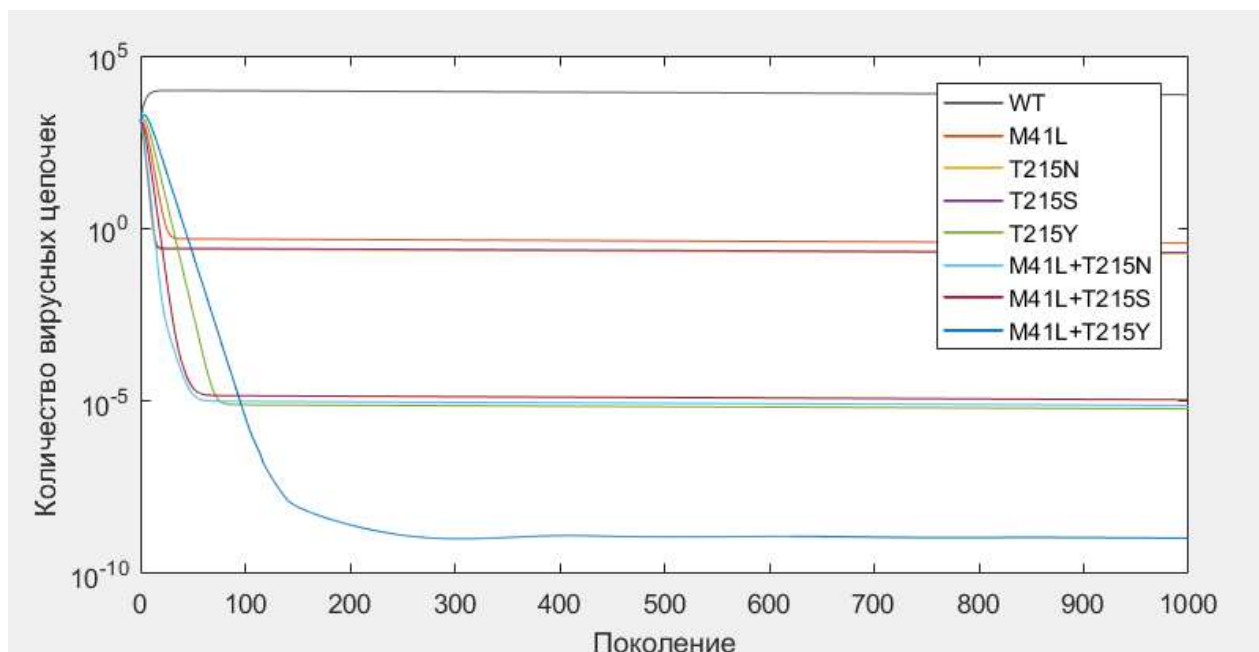


Рисунок 14 – Численность квазивидов при концентрации AZT μ M = 0

Из рисунка видно, что численность мутантов устанавливается примерно за 400 циклов, при этом численность дикого типа также становится преобладающей.

7.2 Результаты работы модели для случая применения противовирусной терапии

Предположим, что при введении противовирусного препарата его концентрация в организме остается постоянной, а выживаемость мутантов различных типов характеризуется величинами, приведенными в таблице 2.

На рисунке 15 показана динамика численности квазивидов при

концентрации препарата, равной 0,03 и начальных условиях (6), которые соответствуют однородной начальной популяции.

Численность резистентных мутантов начинает круто расти. Преобладающими с течением времени становятся квазивиды M41L+T215Y и Y, характеризующиеся наибольшими значениями функции приспособленности при заданной концентрации лекарства.

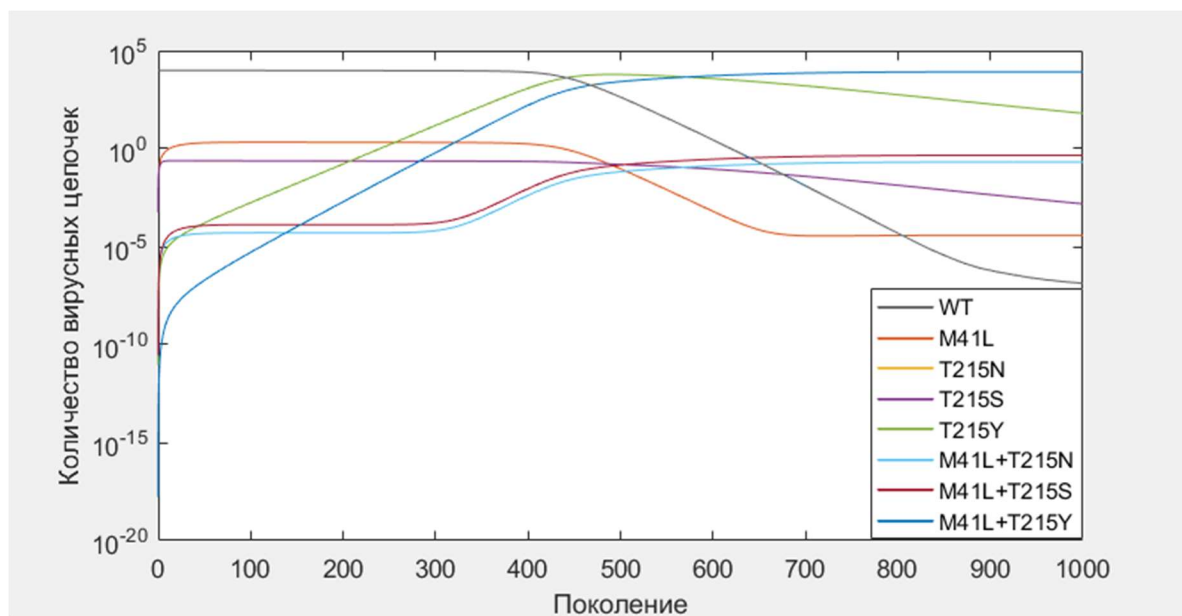


Рисунок 15 – Численность квазивидов при концентрации $AZT_{\mu M} = 0,03$

На рисунке 16 представлена динамика численности квазивидов при концентрации лекарства 0,3 и начальных условиях (6).

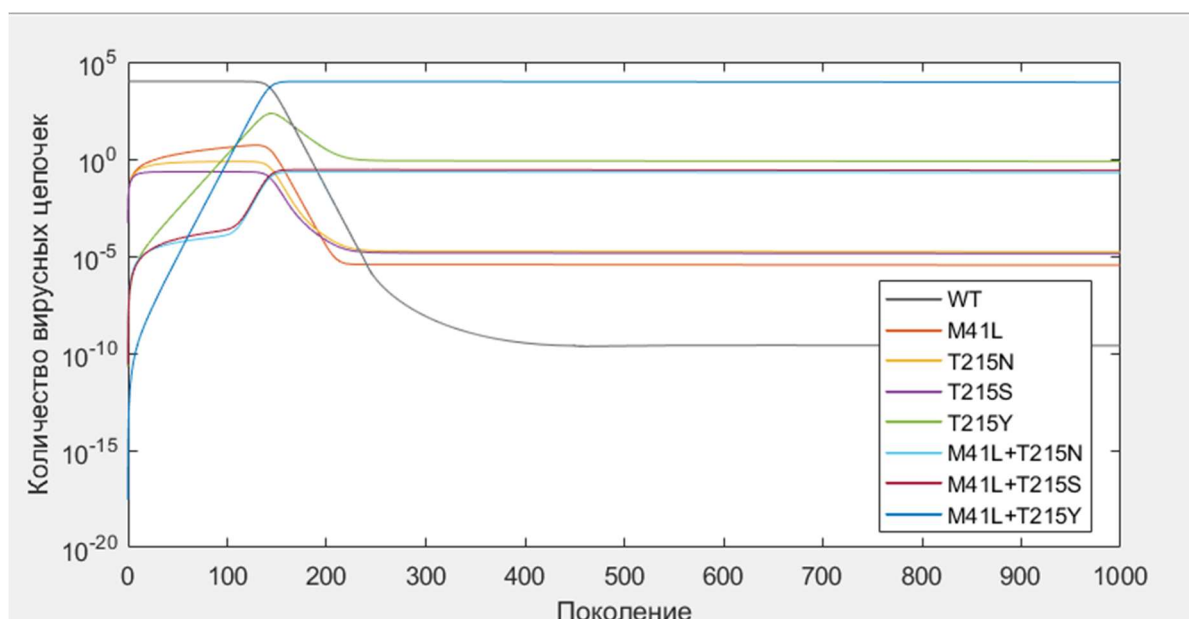


Рисунок 16 – Численность квазивидов при концентрации AZTμM = 0,3

При введении противовирусной терапии численность квазивидов в популяции устанавливается только около 400 цикла, доминирующим является квазивид M41L+T215Y, за ним следуют T215Y, M41L+T215S и N. Дикий тип имеет уже наименьшую численность.

На рисунках 17 и 18 показана динамика квазивидов для вирусной популяции, состоящей в начальный момент времени из одинакового количества геномов всех видов, что соответствует начальным условиям (7), под воздействием противовирусного препарата с концентрацией 0,03 и 0,3 соответственно.

Рисунок 17 показывает установление численности квазивидов примерно за 700 циклов. Лучшую сопротивляемость лекарству и наибольшую численность геномов показывает мутант M41L+T215Y, наименьшую сопротивляемость – дикий тип.

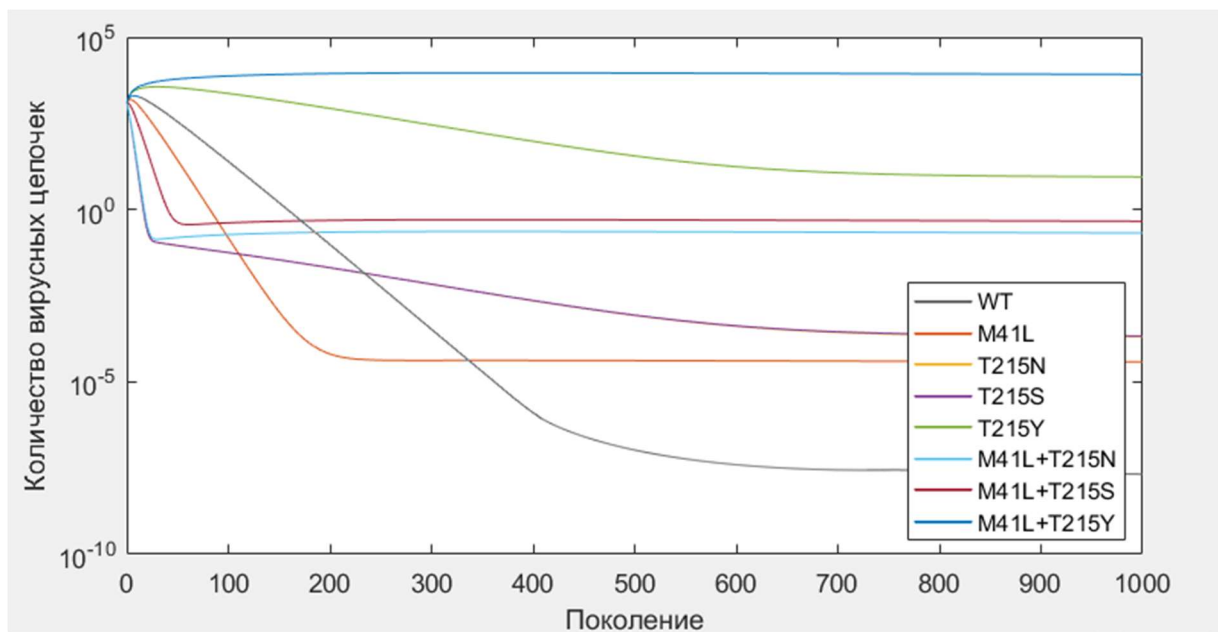


Рисунок 17 – Численность квазивидов при концентрации $AZT_{\mu M} = 0,03$

Рисунок 18 показывает динамику квазивидов при введении лекарства с концентрацией 0,3. При этом численности геномов устанавливаются за 200 циклов. Доминирующим видом так же является мутант M41L+T215Y.

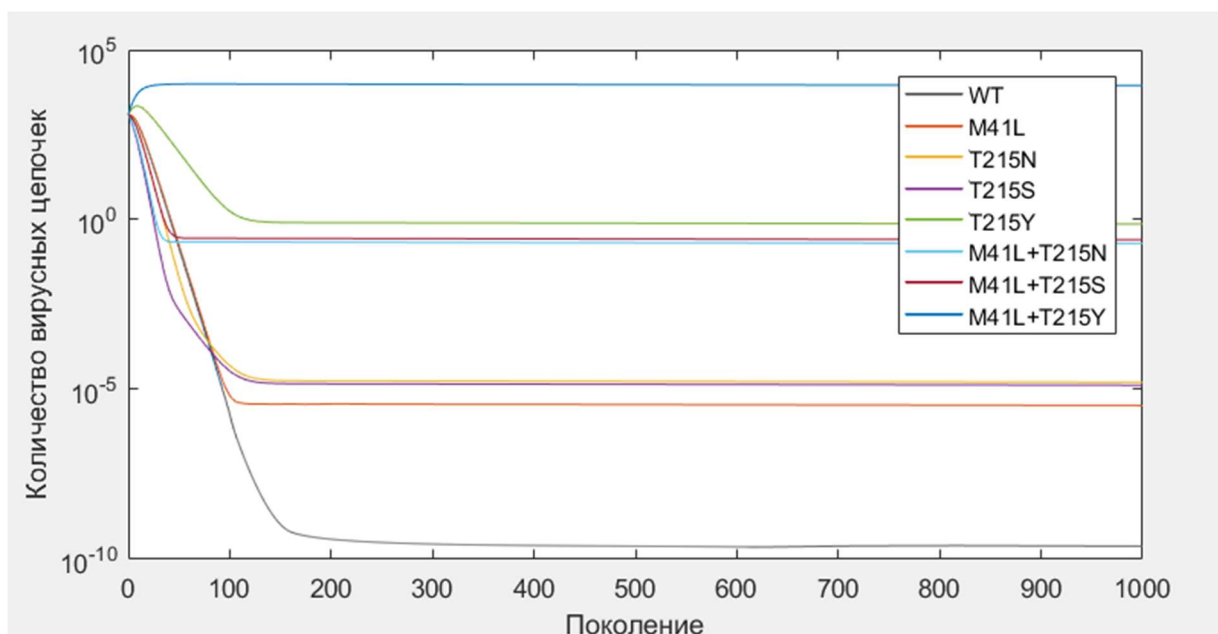


Рисунок 18 – Численность квазивидов при концентрации $AZT_{\mu M} = 0,3$

Численные результаты работы модели показывают, что в процессе воспроизводства популяция по-прежнему остается неоднородной, но

соотношение численности составляющих ее квазивидов существенно изменяется в зависимости от условий – концентрации противовирусного препарата.

Для удобства сравнения результатов с представленными в работе [9] графики (рисунки 13–18) построены для логарифмической шкалы [20]. Полученные результаты согласуются с приведенными в [9]. Другой вариант представления результатов приведен в приложении Б.

8 Описание приложения

8.1 Среда реализации и структура приложения

В состав MatLab входит среда GUIDE для создания приложений с графическим интерфейсом пользователя [21]. Работа в этой среде достаточно проста – элементы управления размещаются при помощи мыши, а затем программируются события, которые возникают при обращении пользователя к данным элементам управления.

Приложение может состоять как из одного основного окна, так и нескольких окон, и осуществлять вывод графической и текстовой информации в основное окно приложения и в отдельные окна. Используя редактор GUIDE, можно создать свой пользовательский интерфейс. Затем GUIDE автоматически сгенерирует код MatLab для создания пользовательского интерфейса, который затем можно изменить, чтобы запрограммировать поведение создаваемого приложения.

Для упрощения запуска моделей, представленных в работе, воспользуемся средой GUIDE и создадим графический интерфейс. Приложение будет состоять из главного окна, предоставляющего выбор параметров и типа запускаемой модели, и дополнительных окон, открывающихся по нажатию кнопки.

8.2 Графический интерфейс

Главное окно запущенного приложения показано на рисунке 19. В окне представлены средства для выбора модели вирусной динамики и концентрацию противовирусного препарата, а также начальные данные для модели, описанной ОДУ.

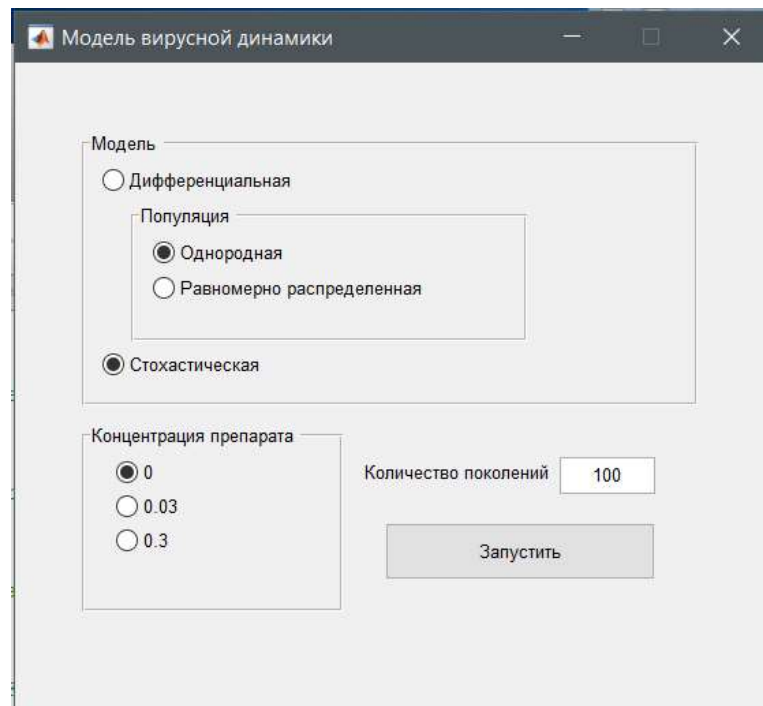


Рисунок 19 – Главное окно

Виды экрана приложения, демонстрирующие результаты запуска для обеих моделей показаны на рисунках 20 и 21.

Рисунок 20 иллюстрирует работу дифференциальной модели для случая, когда начальная популяция состоит из одинакового числа геномов каждого из восьми типов.

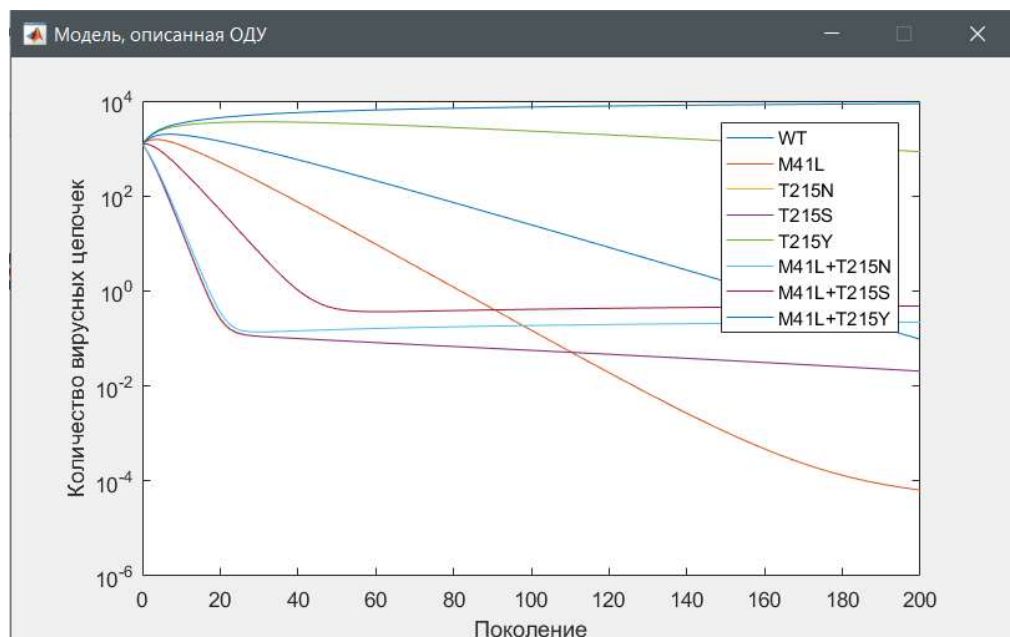


Рисунок 20 – Пример запуска модели, описанной ОДУ

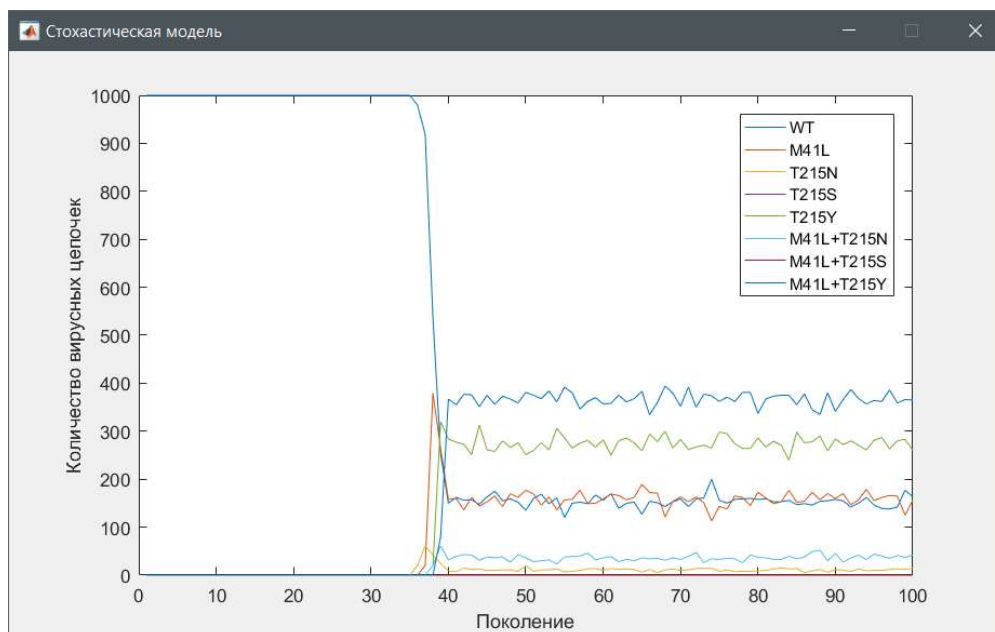


Рисунок 21 – Пример запуска стохастической модели

Рисунок 21 иллюстрирует работу стохастической модели для случая, когда начальная популяция является однородной и целиком состоит из дикого типа.

Результаты работы также могут быть сохранены в файл для последующей интерпретации в другом пакете, например в Excel.

ЗАКЛЮЧЕНИЕ

Для изучения процесса вирусной репликации при инфицировании вирусом иммунодефицита человека, отличающегося рядом особенностей, учет которых усложняет построение моделей на основе обыкновенных дифференциальных уравнений, можно использовать подходы на основе генетических алгоритмов. Если же численность популяции велика, эволюция может рассматриваться как детерминированный процесс. В этом случае эволюционная динамика популяции может быть описана системой обыкновенных дифференциальных уравнений.

Выпускная квалификационная работа посвящена исследованию закономерностей развития популяции ВИЧ в организме и построению моделей с использованием различных подходов к моделированию процесса: стохастической модели эволюции популяции квазивидов ВИЧ на основе генетического алгоритма и детерминированной модели, описанной с помощью ОДУ.

В работе генетические алгоритмы и ОДУ использованы для моделирования эволюции популяции вирусов с учетом действия таких процессов, как рекомбинации, точечные мутации, репликация вирусных геномов и отбор потомков по величине функции приспособленности. Используется четырехбуквенный алфавит, соответствующий реальной кодировке нуклеотидов, реализованы однократные инфекции клеток–мишеней. Дифференциальная модель не позволяет учитывать рекомбинации, что несколько затрудняет сопоставление результатов, полученных с помощью различных подходов. Для численной реализации модели написана программа при помощи средств MATLAB.

В результате построения моделей было показано, что соотношение численности составляющих популяцию квазивидов существенно изменяется в зависимости от условий – концентрации противовирусного препарата, которые отражены в значениях функции приспособленности. Результатом

эволюции является разнородная вирусная популяция, меняется лишь соотношение между числом цепочек различных типов в пользу наиболее приспособленных (резистентных) мутантов.

Некоторые результаты исследования были представлены на XIV Ежегодной молодежной научной конференции студентов, аспирантов и молодых ученых «Достижения и перспективы молодых ученых в интересах развития Юга России» (17–27 апреля 2018 г., Ростов-на-Дону) и опубликованы в сборнике докладов [22].

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Цинкернагель, Р. Основы иммунологии / Р. Цинкернагель. – М.: Мир, 2008. – 135 с.
- 2 Timing the Emergence of Resistance to Anti-HIV Drugs with Large Genetic Barriers // PLoS Computational Biology. – 2009. – Vol.5, Issue 3. – pp. 1–10.
- 3 Putting more genetics into genetic algorithms / D.S. Burke [et al.] // Evolutionary computation. – 1998. – Vol. 3. – pp. 387–410.
- 4 Genome length as an evolutionary self-adaptation / K.L. Ramsey [et al.] // Lecture notes in computer science. PPSN V: Parallel problem solving from nature. 1998. – Vol. 1498. – pp. 345–353.
- 5 Recombination increases human immunodeficiency virus fitness, but not necessarily diversity / N.V. Vijay [et al.] // Journal of General Virology. – 2008. – 89. – pp. 1467–1477.
- 6 Recombination in HIV and the evolution of drug resistance: for better or for worse? / M.T. Bretscher [et al.] // BioEssays 26:180–188, Wiley Periodicals, Inc. 2004. – pp. 180–188.
- 7 Rouzine, I.M. Evolution of Human Immunodeficiency Virus Under Selection and Weak Recombination / I.M. Rouzine, J.M. Goffin // Copyright © 2005 by the Genetics Society of America DOI: 10.1534/genetics.104.029926.
- 8 Гладков, Л.А. Генетические алгоритмы / Л.А. Гладков, В.В. Курейчик, В.М. Курейчик. – М.: Физматлит, 2006. – 320 с.
- 9 Игнатович, А.Н. Математические технологии моделирования динамики вирусов и иммунных реакций / А.Н. Игнатович // Структура и динамика молекулярных систем: Электронный журнал. – 2008. – № 4. – с. 350–386.
- 10 Математические методы интегративного моделирования ВИЧ инфекции / Г.А Бочаров [и др.] // Российский иммунологический журнал. – 2014. – Т. 8. – С. 782–785.

11 Mathematical modelling of the within-host hiv quasispecies dynamics in response to antiviral treatment / G.A. Bocharov [et al.] // Russian Journal of Numerical Analysis and Mathematical Modelling. – 2015 – V. 30/ – pp. 157–170.

12 Pharmacokinetics of Lamivudine, Zidovudine and Nevirapine Administered as a Fixed-Dose Combination Formulation Versus Coadministration of the Individual Products / J.F. Marier [et al.] // Journal of Pharmacology. – 2007. – Vol. 47. – pp. 1381–1389.

13 Эйген, М. Гиперцикл. Принципы самоорганизации макромолекул / М. Эйген, П. Шустер. – М.: Мир, 1982. – 270 с.

14 Nowak, M.A. Virus dynamics: mathematical principles of immunology and virology / M.A. Nowak, R. May. – London: Oxford University Press, 2000. – 237 p.

15 Estimating the relative contribution of dNTP pool imbalance and APOBEC3G/3F editing to HIV evolution in Vivo / K. Deforche [et al.] // Journal of Computational Biology. – 2007. – Vol. 14(8). – pp. 1105 –1114.

16 Pharmacokinetics of Lamivudine, Zidovudine and Nevirapine Administered as a Fixed-Dose Combination Formulation Versus Coadministration of the Individual Products / J.F. Marier [et al.] // Journal of Pharmacology. – 2007. – Vol. 47. – pp. 1381–1389.

17 Решение ОДУ в MATLAB [Электронный ресурс]. – URL: <http://apeshnik.narod.ru/matlab/part3.htm>. – (дата обращения 04.06.2018).

18 Решатели ОДУ в MATLAB [Электронный ресурс]. – URL: <https://studfiles.net/preview/2652877/page:12/>. – (дата обращения 04.06.2018).

19 Solve stiff differential equations and DAEs [Электронный ресурс]. – URL: <https://www.mathworks.com/help/matlab/ref/ode15s.html>. – (дата обращения 04.06.2018).

20 Графики в полулогарифмическом масштабе [Электронный ресурс]. – URL: <http://lib.qrz.ru/node/8058>. – (дата обращения 04.06.2018).

21 Create Apps with Graphical User Interfaces in MATLAB [Электронный ресурс]. – URL: <https://www.mathworks.com/discovery/matlab->

gui.html. – (дата обращения 04.06.2018).

22 Вербицкая, А.А. К моделированию антигенной изменчивости вирусной популяции / А.А. Вербицкая // XIV Ежегодная молодежная научная конференция студентов, аспирантов и молодых ученых «Достижения и перспективы молодых ученых в интересах развития Юга». – Ростов н/Д: изд-во ЮНЦ РАН. – 2018. – С. 84.

ПРИЛОЖЕНИЕ А

Словарь новых понятий

Виртуальный геном – последовательность символов, кодирующая цепочку вирусных РНК;

Квазивид – генетически близкородственные варианты вирусных цепочек в популяции;

Популяция – набор вирусных цепочек РНК;

Вирусная частица – пара вирусных цепочек;

Мутации – случайные изменения одной или нескольких позиций в случайно выбранных цепочках;

Рекомбинация – обмен цепочек своими частями;

Репликация – размножение цепочек;

Отбор – выборка из размножившейся популяции исходного числа хромосом;

Резистентность – устойчивость.

ПРИЛОЖЕНИЕ Б

Результаты вычислительных экспериментов на базе дифференциальной модели

Численные результаты работы модели, основанной на системе ОДУ, представленные в линейной шкале.

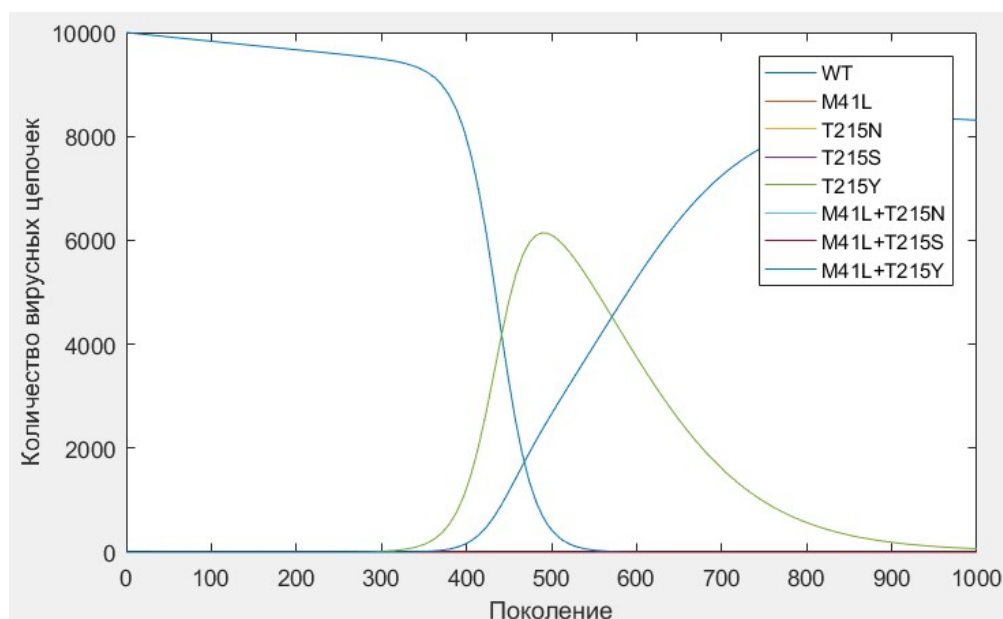


Рисунок Б1 – Численность квазивидов при концентрации $AZT_{\mu M} = 0,03$, однородная начальная популяция

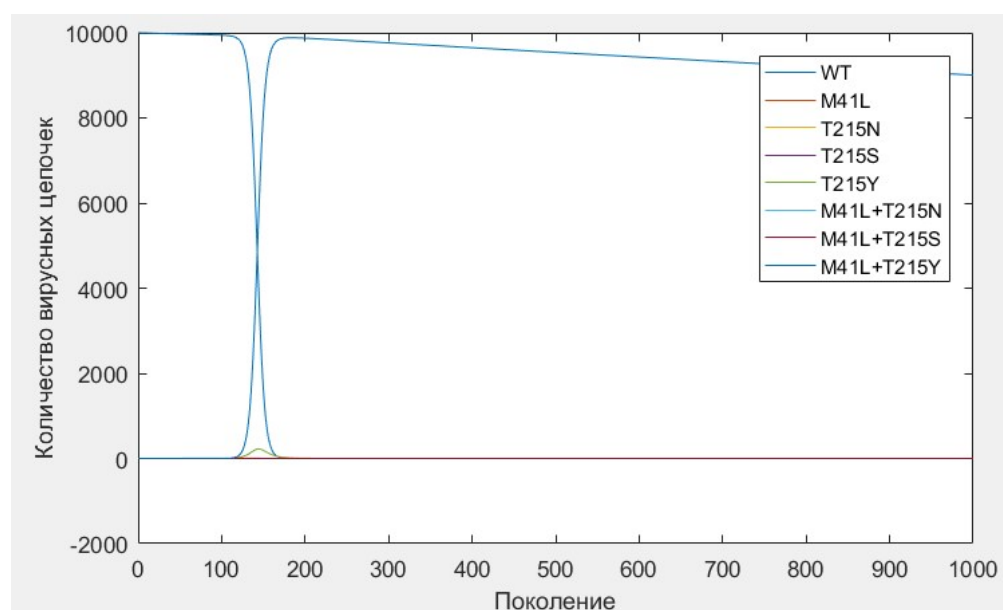


Рисунок Б2 – Численность квазивидов при концентрации $AZT_{\mu M} = 0,3$, однородная начальная популяция

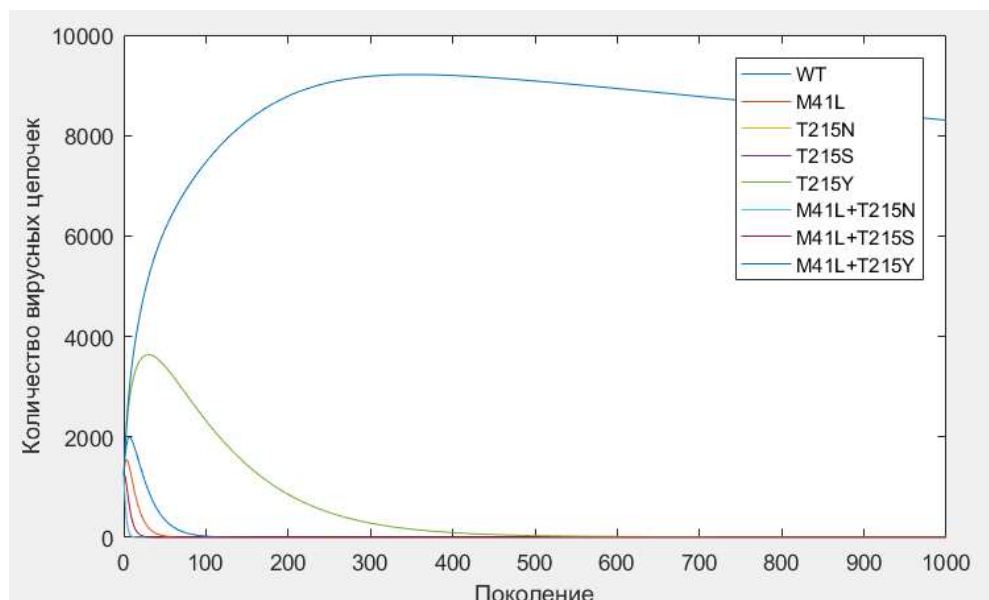


Рисунок Б3 – Численность квазивидов при концентрации $AZT_{\mu M} = 0,03$,
равномерно распределенная начальная популяция

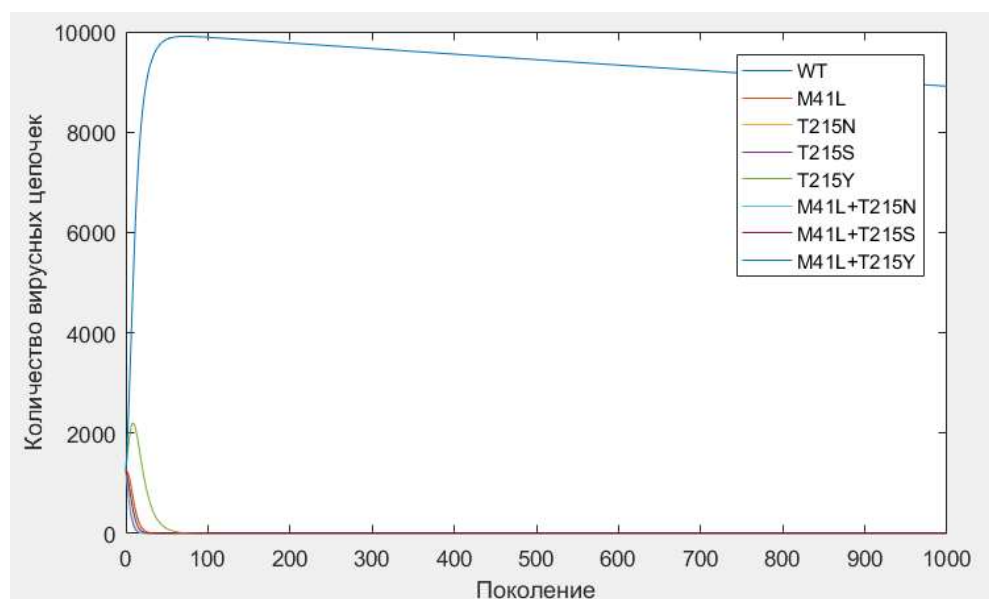


Рисунок Б6 – Численность квазивидов при концентрации $AZT_{\mu M} = 0,3$,
равномерно распределенная начальная популяция

ПРИЛОЖЕНИЕ В

Программа, реализующая стохастическую модель

Стохастическая компьютерная модель на основе генетического алгоритма учитывает происходящие в популяции точечные мутации, рекомбинации вирусных геномов и отбор потомков по величине функции приспособленности.

```
alphabet = ['A', 'G', 'T', 'C'];
alphabet1 = [1 2 3 4];
Npop = 1000; %численность популяции
Ngene = 1800; %длина генома
Pmut = 0.2; %вероятность мутации
Prec = 0.5; %вероятность рекомбинации
global Kiter glConc; %количество поколений и концентрация
препарата
Npairs = Npop*(Npop-1)/2; %количество пар геномов
HamGlobal = zeros(1, Kiter); % для хранения расстояния Хэмминга
MutSpeedA=[2.27743125852128, 0.318614435046603,
0.542883200161911];%A-G,A-T,A-C
MutSpeedG=[7.36088882597026, 0.655035702037978,
0.261999891600743];%G-A,G-T,G-C
MutSpeedT=[0.809503738020646, 0.58021101931253,
2.8093006895185];%T-A,T-G,T-C
MutSpeedC=[1.85795724212526, 0.228442777406494,
4.60051637685454];%C-A,C-G,C-T
%скорость мутации

RouletteA=[MutSpeedA(1)/sum(MutSpeedA),
(MutSpeedA(1)+MutSpeedA(2))/sum(MutSpeedA), 1];
%A->(G,T,C)
RouletteG=[MutSpeedG(1)/sum(MutSpeedG),
(MutSpeedG(1)+MutSpeedG(2))/sum(MutSpeedG), 1];
%G->(A,T,C)
RouletteT=[MutSpeedT(1)/sum(MutSpeedT),
(MutSpeedT(1)+MutSpeedT(2))/sum(MutSpeedT), 1];
%T->(A,G,C)
RouletteC=[MutSpeedC(1)/sum(MutSpeedC),
(MutSpeedC(1)+MutSpeedC(2))/sum(MutSpeedC), 1];
%C->(A,G,T)
%расчет секторов рулетки для мутации

Type=['WT','M41L','T215N','T215S','T215Y','M41L+T215N','M41L+T215S',
'M41L+T215Y'];
switch glConc
    case 1
        Func = [1 0.60398, 0.20344 0.25286 0.70386 0.29144 0.49304
0.78211]; %0
    case 2
```

```

        Func = [0.56022 0.50883 0.095828 0.10023 0.60567 0.12426
0.39282 0.61736]; %0.03
        case 3
            Func = [0.16226 0.16627 0.11927 0.01554 0.29251 0.038069
0.11565 0.38495]; %0.3
        end
        %функции приспособленности

Roulette = zeros(1,8);
Roulette(1) = Func(1)/sum(Func);
for i = 2:8
    Roulette(i) = Roulette(i-1)+ Func(i)/sum(Func);
end
%расчет рулетки для функции приспособленности

GlobType = zeros(Kiter, 8);

%% создание начальной популяции
pop1 = randsrc(Npop, Ngene, alphabet1);
pop = alphabet(pop1); % числа [1 2 3 4] -> буквы ['A', 'G', 'T',
'C']
clear pop1
pop(1, 121:123) = ['A' 'T' 'G'];
pop(1, 643:645) = ['A' 'C' 'C'];
for i = 2:Npop
    pop(i, :) = pop(1, :);
end
newpop = pop; %матрица, в которую будет отбираться новое
поколение

for k = 1:Kiter

    %% рекомбинация
    i = 1;
    while i<=(Npop-1)
        RandRec = rand; %(0, 1)
        if RandRec<=Prec
            RandGene = randi(Ngene); %[1, 1800]
            Temp = pop(i, RandGene:Ngene);
            pop(i, RandGene:Ngene) = pop(i+1, RandGene:Ngene);
            %pop(i+1, :) = pop(i, :);
            pop(i+1, RandGene:Ngene) = Temp;
            %pop(i+3, :) = pop(i+2, :);
            clear Temp;
            i = i + 2; %если произошла рекомбинация со следующим,
пропускаем его
        else
            i = i + 1; %если нет, то продолжаем со следующего
        end
    end

end

%% мутация

```

```

for i = 1:1:Npop
    RandMut = rand;
    if RandMut<=Pmut
        RandGene = randi(Ngene); %номер мутирующего гена
        RandMutGene = rand;
        switch pop(i, RandGene)
            case 'A' %мутирует A
                if RandMutGene<=RouletteA(1)
                    pop(i, RandGene) = 'G' ;
                elseif RandMutGene<=RouletteA(2)
                    pop(i, RandGene) = 'T';
                else
                    pop(i, RandGene) = 'C';
                end
            case 'G' %мутирует G
                if RandMutGene<=RouletteG(1)
                    pop(i, RandGene) = 'A' ;
                elseif RandMutGene<=RouletteG(2)
                    pop(i, RandGene) = 'T';
                else
                    pop(i, RandGene) = 'C';
                end
            case 'T' %мутирует T
                if RandMutGene<=RouletteT(1)
                    pop(i, RandGene) = 'A' ;
                elseif RandMutGene<=RouletteT(2)
                    pop(i, RandGene) = 'G';
                else
                    pop(i, RandGene) = 'C';
                end
            case 'C' %мутирует C
                if RandMutGene<=RouletteC(1)
                    pop(i, RandGene) = 'A' ;
                elseif RandMutGene<=RouletteC(2)
                    pop(i, RandGene) = 'G';
                else
                    pop(i, RandGene) = 'T';
                end
            end
        end
    end
end

%% пересчет
NType=zeros(1,8); %количество генов каждого типа в данной
популяции
for i = 1:1:Npop
    if strcmp(pop(i, 121:123),'TTG')
        switch pop(i, 643:645)
            case 'AAC' % (TTG AAC)- тип M41L+T215N
                NType(6)=NType(6) + 1;
            case 'TCC' % (TTG TCC)- тип M41L+T215S
                NType(7)=NType(7) + 1;
            case 'TAC' % (TTG TAC)- тип M41L+T215Y

```

```

        NType(8)=NType(8) + 1;
    otherwise % (TTG ---)- тип M41L
        NType(2)=NType(2) + 1;
    end
else
    switch pop(i, 643:645)
        case 'AAC' % (--- AAC)- тип T215N
            NType(3)=NType(3) + 1;
        case 'TCC' % (--- TCC)- тип T215S
            NType(4)=NType(4) + 1;
        case 'TAC' % (--- TAC)- тип T215Y
            NType(5)=NType(5) + 1;
        otherwise % тип WT
            NType(1)=NType(1) + 1;
    end
end
end

%% отбор ч.1
MType = zeros(1,8); %количество генов каждого типа, отбираемых в
новую популяцию из данной
Mpop = 0;
while Mpop<Npop
    RandSelect = rand;
    if RandSelect<=Roulette(1)
        if MType(1)<NType(1)*20
            MType(1)=MType(1)+1;
            Mpop = Mpop + 1;
        end
    elseif RandSelect<=Roulette(2)
        if MType(2)<NType(2)*20
            MType(2)=MType(2)+1;
            Mpop = Mpop + 1;
        end
    elseif RandSelect<=Roulette(3)
        if MType(3)<NType(3)*20
            MType(3)=MType(3)+1;
            Mpop = Mpop + 1;
        end
    elseif RandSelect<=Roulette(4)
        if MType(4)<NType(4)*20
            MType(4)=MType(4)+1;
            Mpop = Mpop + 1;
        end
    elseif RandSelect<=Roulette(5)
        if MType(5)<NType(5)*20
            MType(5)=MType(5)+1;
            Mpop = Mpop + 1;
        end
    elseif RandSelect<=Roulette(6)
        if MType(6)<NType(6)*20
            MType(6)=MType(6)+1;
            Mpop = Mpop + 1;

```

```

        end
    elseif RandSelect<=Roulette(7)
        if MType(7)<NType(7)*20
            MType(7)=MType(7)+1;
            Mpop = Mpop + 1;
        end
    else
        if MType(8)<NType(8)*20
            MType(8)=MType(8)+1;
            Mpop = Mpop + 1;
        end
    end
end
end

%% отбор ч.2
Nnewpop=0; %количество уже отобранных в новую популяцию цепочек
генов
NnewType=zeros(1, 8); %количество генов каждого типа, отобранных
в новую популяцию из данной
i = 1; %индекс для старой популяции
j = 1; %индекс для новой популяции
while Nnewpop<Npop
    if strcmp(pop(i, 121:123),'TTG') %проходим
        switch pop(i, 643:645)
            case 'AAC'
                if NnewType(6)<MType(6)
                    newpop(j, :) = pop (i, :);
                    NnewType(6)= NnewType(6) +1;
                    Nnewpop = Nnewpop + 1;
                    j = j + 1;
                end
            case 'TCC'
                if NnewType(7)<MType(7)
                    newpop(j, :) = pop (i, :);
                    NnewType(7)= NnewType(7) +1;
                    Nnewpop = Nnewpop + 1;
                    j = j + 1;
                end
            case 'TAC'
                if NnewType(8)<MType(8)
                    newpop(j, :) = pop (i, :);
                    NnewType(8)= NnewType(8) +1;
                    Nnewpop = Nnewpop + 1;
                    j = j + 1;
                end
            otherwise
                if NnewType(2)<MType(2)
                    newpop(j, :) = pop (i, :);
                    NnewType(2)= NnewType(2) +1;
                    Nnewpop = Nnewpop + 1;
                    j = j + 1;
                end
            end
        end
    end
end
end

```

```

else
    switch pop(i, 643:645)
        case 'AAC'
            if NnewType(3)<MType(3)
                newpop(j, :) = pop (i, :);
                NnewType(3)= NnewType(3) +1;
                Nnewpop = Nnewpop + 1;
                j = j + 1;
            end
        case 'TCC'
            if NnewType(4)<MType(4)
                newpop(j, :) = pop (i, :);
                NnewType(4)= NnewType(4) +1;
                Nnewpop = Nnewpop + 1;
                j = j + 1;
            end
        case 'TAC'
            if NnewType(5)<MType(5)
                newpop(j, :) = pop (i, :);
                NnewType(5)= NnewType(5) +1;
                Nnewpop = Nnewpop + 1;
                j = j + 1;
            end
        otherwise
            if NnewType(1)<MType(1)
                newpop(j, :) = pop (i, :);
                NnewType(1)= NnewType(1) +1;
                Nnewpop = Nnewpop + 1;
                j = j + 1;
            end
        end
    end
end
if i>=Npop
    i = 1;
else
    i = i + 1;
end
end

%% подготовка к следующей итерации
pop = newpop;
GlobType(k, :) = MType;

%% подсчет расстояния Хэмминга

HamDifPairs = zeros(1, Ngene);
HamTotal = 0;
for j = 1:1:Ngene
    HamNums = zeros(1, 4); % количества разных нуклеотидов
    HamPairs = zeros(1, 4); % количество одинаковых пар для
    каждого нуклеотида
    for i = 1:1:Npop %ищем количества разных нуклеотидов в
    столбце j

```

```

switch pop(i,j)
case 'A'
    HamNums(1) = HamNums(1) + 1;
case 'G'
    HamNums(2) = HamNums(2) + 1;
case 'T'
    HamNums(3) = HamNums(3) + 1;
case 'C'
    HamNums(4) = HamNums(4) + 1;
end
end
for l = 1:1:4 % считаем количество одинаковых пар для
каждого нуклеотида
    if (HamNums(l) >= 2)
        HamPairs(l) = HamNums(l) * (HamNums(l) - 1) / 2;
    else
        HamPairs(l) = 0;
    end
end
HamDifPairs(1, j) = Npairs - sum(HamPairs);
% получим количество разных пар нуклеотидов в столбце j
end
HamTotal = sum(HamDifPairs) / Npairs;
% среднее расстояние Хэмминга для популяции
HamGlobal(k) = HamTotal;

end

%% построение графика квазитипов
plot(GlobType(:, :))
%title('Концентрация AZT = 0,03')
xlabel('Поколение')
ylabel('Количество вирусных цепочек')
legend('WT', 'M41L', 'T215N', 'T215S', 'T215Y', 'M41L+T215N', 'M41L+T215S', 'M41L+T215Y')

```

ПРИЛОЖЕНИЕ Г

Программа, реализующая дифференциальную модель

Модель на основе системы ОДУ, описывающая эволюционную динамику популяции вируса.

```
v01 = [10000 0 0 0 0 0 0 0];
v02= [1250 1250 1250 1250 1250 1250 1250 1250];
global Kiter glConc Vn;
v0 = [v01; v02];
t0 = [0 Kiter];
[t,v] = ode15s(@s8, t0, v0(Vn,:));
semilogy(t,v)
xlabel('Поколение')
ylabel('Количество вирусных цепочек')
legend('WT', 'M41L', 'T215N', 'T215S', 'T215Y', 'M41L+T215N', 'M41L+T215S', 'M41L+T215Y')

function vect=s8(t,v)
global glConc;
p = 0.00002;
a0 = [1 0.60398 0.20344 0.25286 0.70386 0.29144 0.49304
0.78211]; % 0
a003 = [0.56022 0.50883 0.095828 0.10023 0.60567 0.12426 0.39282
0.61736]; %0,03
a03 = [0.16226 0.16627 0.11927 0.01554 0.29251 0.038069 0.11565
0.38495]; %0,3
a = [a0; a003; a03];
Qii= (1-3*p)^6;
Qij= p*((1-3*p)^5);
vect = zeros(8,1);
vect(1)=
a(glConc,1)*Qii*v(1)+a(glConc,2)*Qij*v(2)+a(glConc,3)*Qij*v(3)+a
(glConc,4)*Qij*v(4)-v(1)*((a(glConc,:)*v)/sum(v));
vect(2)=
a(glConc,1)*Qij*v(1)+a(glConc,2)*Qii*v(2)+a(glConc,6)*Qij*v(6)+a
(glConc,7)*Qij*v(7)-v(2)*((a(glConc,:)*v)/sum(v));
vect(3)=
a(glConc,1)*Qij*v(1)+a(glConc,3)*Qii*v(3)+a(glConc,5)*Qij*v(5)+a
(glConc,6)*Qij*v(3)-v(3)*((a(glConc,:)*v)/sum(v));
vect(4)=
a(glConc,1)*Qij*v(1)+a(glConc,4)*Qii*v(4)+a(glConc,5)*Qij*v(5)+a
(glConc,7)*Qij*v(7)-v(4)*((a(glConc,:)*v)/sum(v));
vect(5)=
a(glConc,3)*Qij*v(3)+a(glConc,4)*Qij*v(4)+a(glConc,5)*Qii*v(5)+a
(glConc,8)*Qij*v(8)-v(5)*((a(glConc,:)*v)/sum(v));
vect(6)=
a(glConc,2)*Qij*v(2)+a(glConc,3)*Qij*v(3)+a(glConc,6)*Qii*v(6)+a
(glConc,8)*Qij*v(8)-v(6)*((a(glConc,:)*v)/sum(v));
```



```

vect(7)=
a(glConc,2)*Qij*v(2)+a(glConc,4)*Qij*v(4)+a(glConc,7)*Qii*v(7)+a
(glConc,8)*Qij*v(8)-v(7)*(a(glConc,:)*v)/sum(v);
vect(8)=
a(glConc,5)*Qij*v(5)+a(glConc,6)*Qij*v(6)+a(glConc,7)*Qij*v(7)+a
(glConc,8)*Qii*v(8)-v(8)*(a(glConc,:)*v)/sum(v);
end

```