

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ
УЧРЕЖДЕНИЕ «ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
ИНСТИТУТ ЦИТОЛОГИИ И ГЕНЕТИКИ
СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК»**

На правах рукописи

Лашин Сергей Александрович

**МНОГОУРОВНЕВЫЕ КОМПЬЮТЕРНЫЕ МОДЕЛИ
ЭВОЛЮЦИОННЫХ ПРОЦЕССОВ**

1.5.8. - математическая биология, биоинформатика

Диссертация на соискание учёной степени
доктора биологических наук

Научный консультант:

Академик РАН

Колчанов Николай Александрович

Новосибирск – 2024

Оглавление

Список сокращений	7
Введение.....	8
Актуальность работы.....	8
Цели и задачи работы:	11
Научная новизна.....	13
Положения, выносимые на защиту	15
Теоретическая и практическая значимость работы.....	16
Апробация результатов	17
Список публикаций автора по теме диссертации.....	22
1. Обзор литературы: о проблемах и перспективах моделирования эволюционно-биологических процессов	29
1.1. Уровни биологической организации и их взаимосвязь	29
1.1.1. Молекулярно-генетический уровень организации.....	31
1.1.2. Клеточный уровень организации	37
1.1.3. Тканевый/органный уровень организации.....	40
1.1.4. Организменный уровень организации.....	48
1.1.5. Популяционный уровень организации	49
1.1.6. Экологический/биоценотический уровень организации	55
1.2. Заключение по литературному обзору	58
2. Моделирование микробных сообществ.....	60
2.1. Методика моделирования микробных сообществ, основанная на концепции супериндивида	62
2.2. Расширенная версия ГЭЖ для моделирования пространственно- распределённых сообществ.....	67

2.2.1. Моделирование пространственно-распределённой среды обитания	67
2.2.2. Моделирование протока.....	69
2.2.3. Моделирование диффузии	69
2.2.4. Моделирование хемотаксиса.....	71
2.2.5. Факторы пространственного перераспределения и энергетические резервы клеток	73
2.3. Моделирование питания и синтез питательных веществ	74
2.3.1. Моделирование потребления питательных веществ	74
2.3.2. Моделирование синтеза питательных веществ	75
2.4. Моделирование генетического разнообразия популяции	76
2.5. Моделирование фаговой инфекции	81
2.6. Программный комплекс «Гаплоидный эволюционный конструктор» .	83
2.6.1. Общая архитектура ПК ГЭК.....	83
2.6.2. Организация потока вычислений ПК ГЭК.....	85
2.6.3. Модули высокопроизводительных вычислений с параллелизацией по генетическому разнообразию сообщества	86
2.6.4. Модули высокопроизводительных вычислений с параллелизацией по пространственной структуре среды обитания сообщества	92
2.6.5. Модули интеграции ПК ГЭК с существующими средствами моделирования генных сетей и базами данных.....	102
2.6.6. Графический интерфейс пользователя ПК ГЭК и веб-приложения	108
2.7. Содержательные биологические результаты, полученные с использованием ПК ГЭК.....	113

2.7.1. Компьютерное моделирование эволюции микробной популяции: преодоление локальных минимумов при достижении пика на ландшафте приспособленности.....	113
2.7.2. Моделирование формирования и конкуренции популяций подвижных микроорганизмов с различными адаптационными стратегиями в периодически изменяющейся среде	132
2.7.3. Моделирование эволюции в популяциях подвижных микроорганизмов с различными адаптационными стратегиями в периодически изменяющейся среде.....	153
2.7.4. Моделирование эволюции в системе «прокариотическое сообщество – бактериофаг»	163
2.8. Методика моделирования микробных сообществ, основанная на агентно-ориентированной концепции	178
2.8.1. Описание внутренних метаболитов клетки	181
2.8.2. Описание генной сети общего метаболизма клетки	182
2.8.3. Описание генной сети транспортных процессов клетки	183
2.8.4. Описание генной сети клеточного деления	183
2.8.5. Описание генной сети хемотаксиса	184
2.8.6. Моделирование жизненного цикла и деления клетки	184
2.8.7 Программный комплекс Microcosm.....	185
2.8.9. Верификация программного комплекса Microcosm.....	188
2.9. Заключение к главе 2	189
3. Моделирование популяций диплоидных организмов.....	193
3.1. Методика моделирования популяций сложноорганизованных диплоидных организмов.....	193

3.2. Программный комплекс «Диплоидный эволюционный конструктор»	201
3.2.1. Модули поддержки высокопроизводительных вычислений ПК ДЭК	207
3.2.2. Графический интерфейс пользователя ПК ДЭК	220
3.3. Содержательные биологические результаты, полученные с использованием ПК ДЭК	222
3.3.1. Верификация ПК ДЭК на классических задачах генетики популяций	223
3.3.2. Исследование инбридинга и аутбридинга в популяциях диплоидных организмов	225
3.3.3. Моделирование популяционно-генетических и миграционных процессов в популяциях древних людей на территории Евразии	227
3.3.4. Модель эволюции геномной сети с отрицательными обратными связями при разных типах отбора	231
3.4. Заключение к главе 3	235
4. Моделирование эволюции популяций с учётом социальных факторов	238
4.1. Методы моделирования биологических популяций с учётом социальных факторов	240
4.1.1. Модель «Агрессор — Защитник»	240
4.1.2. Расширенная модель «Агрессор - Защитник»	241
4.1.3. Основные режимы динамики модели «агрессор — защитник»	246
4.2. Модель распространения генетической глухоты в популяциях людей с разными социально-демографическими структурами	252
4.2.1. Итерационный цикл модели PGD-S	259
4.2.2. Стратегия выбора партнеров (СВП)	259

4.2.3. Программный комплекс PGD-S.....	263
4.2.4. Моделирование трендов распространения глухоты в программе PGD-S	265
Заключение к главе 4	267
Заключение	268
Выводы	272
Список литературы	275
Приложение 1. Язык описания моделей ГЭК	316
Приложение 2. Система плагинов ДЭК.....	323
Приложение 3. Язык описания моделей ДЭК.....	324
Приложение 4. Сценарий моделирования популяционно-генетических и миграционных процессов в популяциях древних людей на территории Евразии	326
Приложение 5. Общая архитектура программного комплекса PGD-S	329

Список сокращений

ГП – горизонтальный перенос генетического материала;

ГС – генная сеть;

ГЭК – Гаплоидный Эволюционный Конструктор;

ДЭК – Диплоидный Эволюционный Конструктор;

ОДУ – обыкновенные дифференциальные уравнения;

НС – неспецифический субстрат;

ПК – программный комплекс;

СС – специфический субстрат;

PGD-S – Population Genetics of Deafness - Simulator

Введение

Актуальность работы

Биологические системы – это *сложные иерархические системы*, функционирующие на разных уровнях биологической организации – от молекулярно-генетического до экосистемного и биоценотического. Корректное и всестороннее описание таких систем с помощью математического и компьютерного моделирования требует разработки новых методов, которые могли бы связать в единой модели данные, относящиеся если не ко всем, то хотя бы к нескольким уровням биологической организации.

Ключевым свойством биологических систем является их *способность к эволюции*, то есть изменению их свойств в ряду поколений живых организмов. Исследование механизмов, лежащих в основе эволюционных процессов, является одной из важнейших задач современной биологии, для решения которой зачастую привлекается аппарат других наук – физики, математики, информатики. Экспериментальное исследование эволюции затруднено для большинства таксонов: процессы макро- и микроэволюции протекают на интервалах от тысяч до миллионов лет и затрагивают множество организмов и систем организмов. Поэтому математическое и компьютерное моделирование становится важным инструментом исследования эволюции. Разработка реалистичных многопараметрических моделей микро- и макроэволюционных процессов и видообразования, учитывающих влияние условий окружающей среды, трофических взаимодействий, географических факторов, генетической структуры организмов и пр. является одной из основных задач для математической биологии и биоинформатики XXI века.

Первые два десятилетия XXI века ознаменовались значительным развитием высокопроизводительных методов полногеномного секвенирования ДНК, что открыло новые возможности для реконструкции методами молекулярной генетики сценариев эволюции различных видов (Bleidorn, 2016), в том числе бактерий (Herring et al., 2006; Shendure et al.,

2005), растений (Deschamps, Campbell, 2010; Sloan et al., 2012), животных (Amores et al., 2011; Christodoulou et al., 2010; Mehinto et al., 2012) и, наконец, человека (Lander, 2011; Olson, Varki, 2003). Эти экспериментальные и аналитические методы могут быть применены в том числе и к палеогенетическим данным, что даёт возможность оценивать эволюционные сценарии человечества (Green et al., 2010; Krause et al., 2010; Posth et al., 2016; Гунбин et al., 2012; Пономаренко et al., 2014). Разработка новых методов моделирования функционирования и эволюции сложных биологических систем, таких как популяции и сообщества организмов, а также высокоэффективных компьютерных программ для такого моделирования, являются важными этапами в решении как фундаментальных научных проблем теории эволюции, так и прикладных задач биотехнологии, эпидемиологии, медицины, экологии и т.д.

За последние десять-пятнадцать лет в моделировании сложноорганизованных биологических систем был достигнут значительный прогресс. Так, в 2012 году в журнале *Cell* была опубликована первая полногеномная компьютерная модель бактериальной клетки (Karr et al., 2012). Предложенная авторами модель патогенной бактерии *Mycoplasma genitalium* включала в себя описание всех основных молекулярных компонент и биохимических реакций с их участием, описание процессов транскрипции и трансляции наряду с описанием жизненного цикла клетки. Модель была реализована с использованием «гибридного» подхода – для описания различных процессов использовался аппарат обыкновенных дифференциальных уравнений (ОДУ), булевых сетей, подходов, основанных на ограничениях (англ. constraint-based modeling), и стохастическое моделирование. Примечательно, что авторам удалось предсказать поведение мутантных вариантов этой бактерии, часть из которых была получена впервые. В дальнейшем эффективность моделей подобного типа для предсказания новых фенотипических свойств бактерий была показана и

другими исследователями (Bordbar et al., 2014; O'Brien et al., 2015; Yilmaz, Walhout, 2017).

Гибридные подходы математического, а затем и компьютерного моделирования входят в арсенал математической биологии примерно с 1960-х – 1970-х годов. Немалую роль в их развитии и внедрении сыграли работы А.А. Ляпунова (Кулагина, Ляпунов, 1966; Ляпунов, 1962), И.А. Полетаева (Кудрина, Полетаев, 1974; Полетаев, 1975, 1965) и В.А. Ратнера (Ратнер, 2001, 1992; Ратнер В.А., 1966) – основоположников Новосибирской школы математической биологии (Акбердин et al., 2013; Ратушный et al., 2005). Одним из первых гибридных методов, применённых в задачах биологии, является пороговый метод моделирования, разработанный этой школой. Данный метод сочетает подходы теории автоматов и аппарат линейных дифференциальных уравнений (Чураев, Ратнер, 1972).

Расцвет гибридных подходов моделирования в биологии, обусловленный, с одной стороны, существенным ростом вычислительной мощности компьютеров, а, с другой, развитием методов объектно-ориентированного проектирования и программирования, вкуче с появлением достаточно большого количества высокоуровневых языков программирования, пришёл на начало XXI-го века. В разных науках, в том числе, в разных разделах биологии, такие подходы получили название *индивидуально-ориентированного* (англ. individual-based) моделирования (Doebeli, Dieckmann, 2004; Feng et al., 2006; Grimm, 2019; Mooij et al., 2003; Романюха, Каркач, 2003; Семовский С. В.; et al., 2004) или *агентно-ориентированного* (англ. agent-based) моделирования, которое для краткости часто называют *агентным* моделированием (Abar et al., 2017; Grimm et al., 2006, 2005; Machado et al., 2011).

В 2016 году в программной статье классиков экологического моделирования Уолкера Гримма и Уты Бергер были сформулированы основные принципы «моделирования нового поколения» (англ. next-

generations modelling), согласно которым главными признаками моделей нового поколения являются: (1) структурный реализм; (2) эмерджентность; (3) предсказательность (Grimm, Berger, 2016). Расширенный список свойств таких моделей, согласно авторам, следующий:

1. Построение модели «от первых принципов» (т.е. без привлечения дополнительных эмпирических предположений);
2. Гетерогенность и динамичность среды обитания;
3. Связь нижележащих уровней организации (авторы статьи указывают в качестве примера физиологический уровень) с вышестоящими уровнями;
4. Сочетание агентного и феноменологического подходов;
5. Микроэволюция;
6. Использование стандартизованных подмоделей.

Методы моделирования эволюции биологических систем, разрабатываемые автором в течение последних десяти лет и представленные в данной диссертации, не просто удовлетворяют вышеописанным свойствам, но и расширяют спектр областей их применения (в частности, учитывая молекулярно-генетический уровень).

Цели и задачи работы:

Целью данной работы является теоретическое исследование генетических, эволюционных и экосистемных процессов в популяциях с использованием компьютерного моделирования биологических систем с учётом их структурно-функциональной организации.

Задачи, поставленные для достижения поставленной цели:

- Разработать методы моделирования эволюции сложных, иерархически организованных биологических систем – популяций и сообществ одноклеточных гаплоидных и многоклеточных диплоидных организмов

с учётом таких уровней биологической организации как молекулярно-генетический, клеточный, популяционный, экосистемный.

- Реализовать разработанные методы в виде универсальных программных комплексов, в том числе – высокопроизводительных версий, работающих на суперкомпьютерах.
- Провести численные эксперименты по изучению эволюции микробных сообществ и популяций человека на основе разработанных программных комплексов.

Фактический материал и методы исследований

При построении компьютерных моделей использовались данные, относящиеся к различным уровням биологической организации, а именно:

- Морфологические, физические и популяционные характеристики бактерий *Escherichia coli* и других бактерий, а также их сред обитания;
- Генетические и метаболические данные, полученные на основе анализа геномов бактерий и археобактерий, включая данные по генетической регуляции и регуляции ферментативной активности;
- Демографические параметры популяций человека;
- Данные по нормальным и патологическим фенотипам человека, их связь с демографическими параметрами и приспособленностью;

На стадии проектирования и реализации программных комплексов, а также для анализа результатов моделирования, использовались следующие алгоритмы, языки программирования и технологии:

- Паттерны объектно-ориентированного проектирования (Буч et al., 2017; Гамма et al., 2015) – для разработки архитектуры программных комплексов;
- Язык C++ – для написания основного каркаса и вычислительного ядра приложений;

- Платформа Qt (<https://www.qt.io/>) – для реализации графического интерфейса пользователя, а также для реализации многопоточных версий ГЭК с использованием библиотеки Qt Concurrent;
- Программный интерфейс MPI – для реализации параллельных версий программных комплексов, работающих на высокопроизводительных кластерах;
- Библиотека OpenMP – для реализации параллельных версий программных комплексов, работающих на рабочих станциях с многоядерными процессорами;
- Язык Python – для работы с данными, генерации отчётов, рисунков и т.п.;
- Язык SBML (<http://sbml.org/>), а также библиотеки libSBML, sosLib – для импорта, экспорта, хранения и расчёта математических моделей молекулярно-генетических систем;
- Язык JavaScript и библиотека webix (<https://webix.com/>) – для реализации веб-версии ГЭК;
- Системы контроля версий svn и github.

Научная новизна

Впервые на момент опубликования были представлены методы моделирования функционирования и эволюции комплексных биологических систем, *учитывающие в одной модели* максимально широкий спектр уровней биологической организации (молекулярно-генетический, метаболический, клеточный, популяционный и экологический уровни). Методы были реализованы в программных комплексах «Гаплоидный Эволюционный конструктор 3D» (ГЭК 3D) (Klimenko et al., 2015, 2019; Lashin, Matushkin, 2012; Mustafin et al., 2012) и «Диплоидный Эволюционный Конструктор» (ДЭК) (Lashin, Matushkin, 2013). Ключевой особенностью предложенных методов является то, что конструируемые модели обладают многоуровневой (англ. multiscale или multilevel, реже multilayer) архитектурой, при этом каждый слой представляет собой подмодель, ответственную за описание

определённого уровня биологической организации: генетического, метаболического, популяционного и, наконец, экологического уровней. Такая архитектура позволяет разрабатывать множество разных подмоделей для одного слоя, фактически создавая библиотеки подмоделей. При этом, поскольку интерфейсы взаимодействия между подмоделями разных слоёв (в частности, входные и выходные данные) чётко специфицированы, при построении общей модели появляется возможность комбинирования различных сочетаний подмоделей разных слоёв, не накладывая жёстких ограничений на формализм, в котором записываются эти подмодели. Последнее позволяет описывать один слой, к примеру, с помощью ОДУ, а другой с помощью клеточных автоматов и т.п.

Разработанные программные комплексы позволяют исследовать различные аспекты эволюционного процесса в рамках одного программного средства. Предложенная архитектура построения моделей эволюционных и популяционно-генетических процессов являлась полностью оригинальной на момент её разработки и опубликования.

С помощью разработанных программных комплексов были получены оригинальные биологически содержательные результаты, касающиеся:

- эволюции пространственно-гетерогенных микробных сообществ и сообществ вида фаг-бактерия (Klimenko et al., 2016, 2015, 2019; Lashin et al., 2020, 2012a; Лашин et al., 2012a);
- влияния социальных факторов на распространение генетической глухоты в популяциях человека (Posukh et al., 2016; Romanov et al., 2018);
- популяционно-генетических и миграционных процессов в популяциях древних людей. В частности, на основе полученных оценок популяционно-генетических параметров были построены модели эволюции предков современных людей на территории Евразии [Gunbin, Lashin, 2013];

- эволюции генных сетей (ГС) с регуляторными контурами, работающим по принципу отрицательной обратной связи. В частности, показано, что в популяциях носителей ГС с обратной связью существенно дольше сохраняется биоразнообразие, что при меняющихся условиях среды может играть существенную роль в сохранении популяции (Trut et al., 2017).

Положения, выносимые на защиту

1. Разработанные компьютерные методы построения и анализа иерархических многоуровневых моделей биологических систем, а также реализованные на их основе программные комплексы «Гаплоидный Эволюционный Конструктор» (ГЭК) и «Диплоидный Эволюционный Конструктор» (ДЭК) являются эффективными инструментами для решения как фундаментальных, так и прикладных задач различной направленности.

2. С помощью программного комплекса ГЭК, разработанного для моделирования функционирования и эволюции пространственно-распределённых микробных сообществ с учетом диффузии метаболитов и двигательной активности бактерий в одно-, двух- и трёхмерном случаях, установлено, что

(а) экологические факторы способствуют поддержанию различных эволюционных стратегий развития пространственно-распределённых микробных сообществ, направленных как на усложнение, так и упрощение метаболизма микроорганизмов;

(б) эффективность влияния бактериофагов на структуру бактериальных сообществ в пространственно-распределённых средах обитания и скорость видообразования зависят от экологических факторов и состояния микробного сообщества в момент инвазии фага.

3. С помощью программного комплекса ДЭК, разработанного для моделирования эволюционно-генетических процессов в популяциях

высокоорганизованных диплоидных организмов, установлено, что наличие регуляторных отрицательных обратных связей в генных сетях, отвечающих за приспособленность особей к условиям окружающей среды, способствует сохранению биоразнообразия популяции и формированию пула скрытой изменчивости – одного из важных факторов эволюции в изменяющихся условиях внешней среды.

4. На базе ДЭК разработан программный комплекс «PGD-S», позволяющий анализировать влияние социальных факторов на динамику распространения наследственных заболеваний в изолированных популяциях человека. С его помощью установлено, что увеличение частоты встречаемости наследственной глухоты в популяциях человека, обусловленной наличием аутосомно-рецессивного гена *DFNB1A*, связано с преимущественным заключением браков между глухими людьми. Показано, что в перспективе этот фактор может быть нивелирован ростом социального равенства глухих людей, о чем свидетельствуют результаты моделирования нейтрального отбора.

Теоретическая и практическая значимость работы

Разработанные программные комплексы используются в ИЦиГ СО РАН для теоретических исследований микробной эволюции, моделирования эпидемических процессов, распространения генетических болезней. Программные комплексы также входят в методический арсенал кафедры информационной биологии факультета естественных наук Новосибирского государственного университета и используются в курсах «Математические основы системной биологии: моделирование молекулярно-генетических систем», «Современные проблемы биоинформатики и системной биологии», «Информационные технологии и языки программирования», «Моделирование молекулярно-генетических систем», «Анализ и моделирование сложных биологических систем», «Системная биология в биомедицине» и

используются в курсе «Математическое моделирование в биологии», читаемом в аспирантуре ИЦиГ СО РАН.

Разработанные модели цитируются в работах по эволюции бактериофагов (Silveira, Rohwer, 2016), их классификации (Schiettekatte et al., 2018; Vincent et al., 2017); функционированию микробиомов коралловых экосистем (Morrow et al., 2018); микробиоте кишечника человека (Ghattargi et al., 2018), рубца жвачных животных Бразилии (Mantovani et al., 2017), отложений мангровых лесов (Wallace Ribeiro Nunes Neto et al., 2019), а также в работах по созданию компьютерных моделей патогенных бактерий сельскохозяйственных культур (Bergholz, Wiedmann, 2016), что свидетельствует о том, что подобные модели помогают пониманию биологических механизмов функционирования и эволюции широкого класса биологических систем.

Разработанные подходы к построению многослойных моделей иерархических биологических систем легли в основу работ, проводимых в ФИЦ ИЦиГ СО РАН (работы по созданию информационно-компьютерной платформы для поддержки биоресурсных коллекций Российской Федерации - <http://www.biores.cytogen.ru/>) и Курчатовском Геномном Центре ИЦиГ СО РАН (работы по построению математических моделей метаболизма и генетической регуляции биотехнологически значимых штаммов микроорганизмов на основе данных геномного секвенирования; работы по созданию информационно-компьютерной платформы для планирования экспериментов *in silico* – <https://cgimu.sysbio.cytogen.ru/>).

Апробация результатов

Основные результаты диссертации докладывались и обсуждались на 36-ти всероссийских и международных научных конференциях: «Актуальные проблемы вычислительной и прикладной математики» (2015, 2019, Новосибирск – приглашённые устные доклады); «The International Society for Ecological Modelling Global Conference» (2019, Вена, Австрия – устный

доклад); 7th, 8th, 9th, 10th, 11th, 12th и 13th International Multiconference “Bioinformatics of Genome Regulation and Structure\Systems Biology” — BGRS\SB (2010, 2012, 2014, 2016, 2018, 2020, 2022 Новосибирск – устные доклады); 1st, 2nd, 3rd International Symposium Mathematical Modeling and High-Performance Computing in Bioinformatics, Biomedicine and Biotechnology (MM-HPC-BBB) (2014, 2016, 2018, Новосибирск – устные доклады); «The European Human Genetics Conference in conjunction with the European Meeting on Psychosocial Aspects of Genetics» (2018, Милан, постерный доклад); «Генетика микроорганизмов: от геномики к биоэкономике» (2018, Пущино – приглашённый пленарный доклад); V, VI, VII Международная конференция «Математическая биология и биоинформатика» (2014, 2016, 2018, Пущино – устные доклады); IV, V Международная научно-практическая конференция «Постгеномные методы анализа в биологии, лабораторной и клинической медицине» (2014, 2018, Казань – устные доклады); «Математика в современном мире. Международная конференция, посвящённая 60-летию Института математики им. С.Л. Соболева» (2017, Новосибирск – устный доклад); «Беляевские чтения, Международная конференция, посвященная 100-летию со дня рождения академика АН СССР Д.К. Беляева» (2017, Новосибирск – устный доклад); Международный конгресс «Биотехнология: состояние и перспективы развития» (2011, 2017, Москва – устные доклады); «Systems Biology and Bioinformatics (SBBI’2016)» (2016, Санкт-Петербург – постерный доклад); the 66th Annual Meeting of The American Society of Human Genetics (2016, Ванкувер, Канада); XVII Всероссийская конференция молодых ученых по математическому моделированию и информационным технологиям (YM2016) (2016, Новосибирск – устный доклад); VIII Сибирский форум индустрии информационных систем (SIIS 2016) (2016, Новосибирск – устный доклад); FAIRDOM Barcelona user meeting (2016, Барселона, Испания – приглашённый устный доклад); EMBO | EMBL Symposium “New Approaches and Concepts in Microbiology” (2015, Гейдельберг, Германия – постерный доклад); Седьмая международная молодежная научная школа-конференция

«Теория и численные методы решения обратных и некорректных задач», посвященная 90-летию со дня рождения академика Гурия Ивановича Марчука (2015, Новосибирск – устный доклад); VI и VII съезды Вавиловского общества генетиков и селекционеров (2014, 2019, Ростов-на-Дону и Санкт-Петербург, соответственно – устные доклады); Federation of European biochemical societies congress (2013, Санкт-Петербург – постерный доклад); The International Moscow Conference On Computational Molecular Biology (MCCMB'11) (2011, Москва – устный доклад); III International Conference “Biosphere Origin and Evolution” (2011, Ретимно, Греция – устный доклад); ERASysBio Summer School Data Management for Systems Biology III (2010, Тенерифе, Испания – устный доклад).

Отдельные результаты диссертации использовались при чтении лекций и проведении практических занятий в рамках следующих курсов Новосибирского Государственного Университета: «Введение в информационную биологию», «Современные проблемы биоинформатики и системной биологии», «Математические основы системной биологии: моделирование молекулярно-генетических систем» а также курсов «Математическая биология, биоинформатика» и «Математическое моделирование в биологии» для аспирантов ИЦиГ СО РАН.

Гранты и проекты

Работа над диссертацией проводилась при частичной финансовой поддержке следующих грантов и проектов:

- Грант РФФИ № 10-04-01310-А «Математическое и компьютерное моделирование коэволюции сообществ трофически связанных организмов», 2010-2012 гг., руководитель к.б.н. Матушкин Ю.Г.
- Грант РФФИ № 12-07-00671-А «Разработка программного комплекса для моделирования эволюционных и популяционно-генетических

- процессов в популяциях диплоидных организмов», 2012-2014 гг., руководитель Лашин С.А.
- Стипендия Президента Российской Федерации для молодых ученых и аспирантов, осуществляющих перспективные научные исследования и разработки по приоритетным направлениям модернизации российской экономики СП 1941.2012.5 «Разработка программных средств для высокопроизводительного моделирования эволюционно-генетических процессов», 2012-2014 гг., получатель Лашин С.А.
 - Междисциплинарный интеграционный проект СО РАН №21 «Исследование закономерностей и тенденций развития самоорганизующихся систем на примере веб-пространства и биологических сообществ», 2012-2014 гг., руководитель чл.-корр. РАН Федотов А.М.
 - Междисциплинарный интеграционный проект СО РАН №47 «Суперкомпьютерная реализация стохастической эволюции ансамблей взаимодействующих частиц различной природы для решения естественно-научных и нанотехнологичных задач», 2012-2014 гг., руководитель чл.-корр. РАН Михайлов Г.А.
 - Грант РФФИ №13-04-00620-А «Компьютерное моделирование коэволюции сообществ гаплоидных взаимодействующих организмов», 2013-2015 гг., руководитель к.б.н. Матушкин Ю.Г.
 - Грант РФФИ № 14-24-00123 «Системная компьютерная биология: анализ и моделирование структурно-функциональной организации и эволюции генных сетей», 2014-2016 гг., руководитель академик РАН Колчанов Н.А.
 - Грант РФФИ № 15-07-03879-А «Программный комплекс Гаплоидный Эволюционный конструктор 3D для моделирования пространственно распределённых сложноорганизованных микробных сообществ», 2015-2017 гг., руководитель Лашин С.А.

- Грант РФФИ № 15-04-04860-А «Оценка потенциальной роли социально-демографической структуры сообществ глухих людей в распространенности наследуемых форм потери слуха в регионах Сибири», 2015-2017 гг., руководитель к.б.н. Посух О.Л.
- Грант РФФИ № 17-29-06016-офи-м «Изучение разнообразия генетического контроля наследуемой потери слуха у коренного населения Сибири и создание специфичной панели генов для персонифицированного ДНК-тестирования», 2017-2019 гг., руководитель к.б.н. Посух О.Л.
- Грант РФФИ № 20-04-00885-А «Разработка и апробация методов анализа макро- и микроэволюционных характеристик генных сетей», 2020-2022 гг., руководитель Лашин С.А.
- Бюджетные проекты Отдела Системной Биологии ИЦиГ СО РАН с 2010 по 2023 гг.

Публикации. По материалам диссертационной работы опубликовано 106 научных работ, из них – 27 статей в журналах из Перечня ВАК (все индексируются в Web of Science и Scopus), 4 главы в монографии, 5 РИД.

Структура и объём работы. Диссертация состоит из введения, четырех глав, заключения, выводов, списка литературы из 349 наименований и 5 приложений. Содержание основного текста диссертации изложено на 330 страницах, содержит 93 иллюстрации и 41 таблицу.

Личный вклад автора

Автор внёс решающий вклад в постановку цели и задач исследования, разработку методов моделирования, проектирование и реализацию программных комплексов, формулировку постановок численных экспериментов, анализ и интерпретацию результатов моделирования. В большинстве публикаций по теме диссертации автор являлся первым автором и/или автором для переписки. Содержание диссертации и основные

положения, выносимые на защиту, отражают персональный вклад автора в опубликованные работы. У автора отсутствует конфликт интересов с соавторами по публикациям.

Благодарности

Автор благодарен коллегам и соавторам за плодотворное обсуждение различных аспектов работы на всём протяжении её выполнения: академику Н.А. Колчанову, к.б.н. Ю.Г. Матушкину, д.б.н. В.А. Лихошваю – за научное руководство, знакомство с математической биологией, обсуждение основных идей работы с самого её начала и постоянную поддержку; к.б.н. Д.А. Афонникову и Д.А. Рассказову – за острые дискуссии и помощь в работе с высокопроизводительными кластерами; к.б.н. О.Л. Посух, к.б.н. Н.А. Барашкову и к.б.н. Г.П. Романову – за совместные работы в области исследования распространения генетической глухоты; сотрудникам сектора компьютерного анализа и моделирования биологических систем ИЦиГ СО РАН : к.б.н. Ф.В. Казанцеву, к.б.н. А.И. Клименко, к.б.н. З.С. Мустафину, Е.А. Мамонтовой, А.И. Скибе, А.А. Смирновой, Т.Н. Лаховой, А.В. Кравчуку, С.П. Куценогому, А.Д. Чеканцеву, Р.К. Зудину, И.С. Дьяченко, Э.А. Магеррамову, Б.С. Зураеву, А.В. Кузьмичёву – за совместную реализацию наших планов; сотрудникам отдела системной биологии ИЦиГ СО РАН за творческую атмосферу и плодотворные дискуссии.

[Список публикаций автора по теме диссертации](#)

Главы в монографиях:

1. **Лашин С.А.,** Суслов В.В., Матушкин Ю.Г., Колчанов Н.А. Моделирование коэволюции одноклеточных гаплоидных организмов с помощью программного комплекса «Эволюционный конструктор» // В монографии: Роль микроорганизмов в функционировании живых систем: фундаментальные проблемы и биоинженерные приложения. Ред. Власов В.В., Дегерменджи А.Г., Колчанов Н.А., Пармон В.Н., Репин В.Е. Новосибирск: СО РАН. 2010, с. 347-375.

2. Лихошвай В.А., Хлебодарова Т.М., Ратушный А.В., **Лашин С.А.**, Турнаев И.И., Подколотная О.А., Ананько Е.А., Смирнова О.Г., Ибрагимова С.С., Колчанов Н.А. Компьютерный генетический конструктор: математическое моделирование генетических и метаболических подсистем *E. coli* // В монографии: Роль микроорганизмов в функционировании живых систем: фундаментальные проблемы и биоинженерные приложения. Ред. Власов В.В., Дегерменджи А.Г., Колчанов Н.А., Пармон В.Н., Репин В.Е. Новосибирск: СО РАН. 2010, с.376-391.
3. **Lashin S.A.**, Matushkin Yu.G., Klimenko A.I., Suslov V.V., Kolchanov N.A. (2014). Evolution, Biodiversity and Ecology in Microbial Communities: Mathematical Modeling and Simulation with the “Haploid Evolutionary Constructor” Software Tool, Biodiversity - The Dynamic Balance of the Planet, PhD. Oscar Grillo (Ed.), p.111-136. ISBN: 978-953-51-1315-7, InTech.
4. Trut L.N., Herbek Yu.E., Trapezov O.V., **Lashin S.A.**, Matushkin Yu.G., Markel A.L., Kolchanov N.A. The animal domestication experiment as a model of the evolutionary process: a new insight into evolution under selection targeting regulatory systems // In: Korogodina V.L., Mothersill C., Inge-Vechtomov S.G., Seymour C. (eds). Genetics, Evolution and Radiation Crossing Borders, The Interdisciplinary Legacy of Nikolay W. Timofeeff-Ressovsky. The Springer, Netherlands, pp. 455-478.

Публикации в изданиях, индексируемых Web of Science и Scopus:

1. **Lashin S.A.**, Suslov V.V., Matushkin Yu.G. Comparative modeling of coevolution in communities of unicellular organisms: adaptability and biodiversity // Journal of Bioinformatics and Computational Biology, 2010, V 8, N. 3, p. 627-643.
2. **Лашин С.А.**, Матушкин Ю.Г., Суслов В.В., Колчанов Н.А. Эволюционные тренды в системах “прокариотическое сообщество” и

- “прокариотическое сообщество–фар” // Генетика, 2011, т.47, №12, с. 1676–1685.
3. **Лашин С. А.**, Суслов В. В., Матушкин Ю. Г. Теории биологической эволюции с позиции современного развития системной биологии // Генетика, 2012, т.48, №5, с.573-589.
 4. **Lashin S.A.**, Matushkin Yu.G. Haploid evolutionary constructor: new features and further challenges // *In Silico Biology*, 2012. V.11. N. 3. P. 125-135.
 5. **Lashin S.A.**, Matushkin Yu.G., Suslov V.V., Kolchanov N.A. Computer modeling of genome complexity variation trends in prokaryotic communities under varying habitat conditions // *Ecological Modelling*, 2012, Vol. 224, No.1, pp. 124-129, doi: 10.1016/j.ecolmodel.2011.11.004.
 6. **Лашин С. А.**, Мамонтова Е.А., Матушкин Ю. Г. Разработка пространственно распределённой модели эволюции прокариотических сообществ // *Вавиловский журнал генетики и селекции*, 2012, т.16, №4/1, с. 830-837.
 7. Мустафин З.С., Матушкин Ю.Г., **Лашин С.А.** Высокопроизводительное моделирование эволюции прокариотических сообществ с использованием программного комплекса «гаплоидный эволюционный конструктор» // *Вавиловский журнал генетики и селекции*, 2012, т.16, №4/1, с. 825-829.
 8. **Lashin S.A.**, Matushkin Yu.G. Multi-layer computer models of gene network evolution in diploid populations // *FEBS Journal*, 2013, Vol. 280 (Suppl. 1), p. 563.
 9. Matushkin Yu.G., **Lashin S.A.** Spatially distributed models of evolution in symbiotic/antagonistic prokaryotic communities // *FEBS Journal*, 2013, Vol. 280 (Suppl. 1), pp. 563-564.
 10. **Lashin S.A.**, Klimenko A.I., Mustafin Z.S., Kolchanov N.A., Matushkin Yu.G. HEC 2.0: improved simulation of the evolution of prokaryotic

- communities // Математическая биология и биоинформатика, 2014, Т.9, №2, стр. 585-596.
11. Mustafin Z.S., Matushkin Yu.G., **Lashin S.A.** High performance simulations of population-genetic processes in bacterial communities using the haploid evolutionary constructor software // Вавиловский журнал генетики и селекции, 2014, Т.18, №4/2 стр. 1039-1045.
12. **Лашин С.А.**, Мамонтова Е.А., Матушкин Ю.Г., Колчанов Н.А. Компьютерное моделирование механизмов формирования и распространения социально-биологических связей и динамики поведения // Вавиловский журнал генетики и селекции, 2014, Т.18, №4/3, стр. 1249-1257.
13. Klimenko A.I., Matushkin Yu.G., Kolchanov N.A., **Lashin S.A.** Modeling evolution of spatially distributed bacterial communities: a simulation with the haploid evolutionary constructor» // BMC Evol Biol, 2015, 15 (Suppl 1):S3.
14. Клименко А.И., Мустафин З.С., Чеканцев А.Д., Зудин Р.К., Матушкин Ю.Г., **Лашин С.А.** Современные подходы к математическому и компьютерному моделированию в микробиологии // Вавиловский журнал генетики и селекции, 2015, Т.19(6): 74-82, DOI 10.18699/VJ15.095.
15. Klimenko A.I., Matushkin Yu.G., Kolchanov N.A., **Lashin S.A.** Bacteriophages affect evolution of bacterial communities in spatially distributed habitats: a simulation study // BMC Microbiol, 2016, 16 (Suppl 1):S10, DOI: 10.1186/s12866-015-0620-4.
16. Посух О.Л., Бады-Хоо М.С., Зыцарь М.В., Михальская В.Ю., **Лашин С.А.**, Барашков Н.А., Романов Г.П. Роль социально-демографической структуры сообществ глухих людей в распространенности наследуемых форм потери слуха // Вавиловский журнал генетики и селекции, 2016, Т. 20(1):7-15, DOI: 10.18699/VJ15.098.
17. **Лашин С.А.**, Афонников Д.А., Генаев М.А., Казанцев Ф.В., Комышев Е.Г., Ощепкова Е.А., Петров А.В., Рассказов Д.А., Смирнова А.А.,

- Колчанов Н.А. Информационная система по биоресурсным коллекциям институтов ФАНО России // Вавиловский журнал генетики и селекции, 2018, 22, 3, 386-393, DOI: 10.18699/VJ18.360.
18. Казанцев Ф.В., Смирнова А.А., Розанов А.С., Уварова Ю.Е., Афонников Д.А., Пельтек С.Е., **Лашин С.А.** Протокол работы с информационной системой по биоресурсным коллекциям институтов ФАНО России на примере коллекции микроорганизмов // Вавиловский журнал генетики и селекции, 2018, 22, 2, 279-284, DOI: 10.18699/VJ18.361.
19. Kazantsev F.V., Akberdin I.R., **Lashin S.A.**, Ree N.A., Timonov V.S., Ratushny A.V., Khlebodarova T.M., Likhoshvai V.A. MAMMOTH: a new database for curated MAtheMatical Models of bioMOlecular sysTems // Journal of bioinformatics and computational biology, 2018, V 16, 01, 1740010, DOI: 10.1142/S0219720017400108.
20. Romanov G.P., Barashkov N.A, Teryutin F.M., **Lashin S.A.**, Solovyev A.V., Pshennikova V.G., Bondar A.A., Morozov I.V., Sazonov N.N., Tomsky M.I., Dzhemileva L.U., Khusnutdinova E.K., Posukh O.L., Fedorova S.A. Marital structure, genetic fitness, and the GJB2 gene mutations among deaf people in Yakutia (Eastern Siberia, Russia) // Russian journal of genetics, 2018, 54 (5): 554-561, DOI: 10.1134/S1022795418050071.
21. Колчанов Н.А., Пельтек С.Е., Розанов А.С., **Лашин С.А.** Системная компьютерная биология и биоинформатика: моделирование и компьютерный дизайн экспериментов по созданию штаммов с целевыми свойствами // Биотехнология. 2018. Т. 34. № С1. С. 11, DOI: 10.21519/0234-2758-2018-S-11-11.
22. **Лашин С.А.**, Матушкин Ю.Г., Клименко А.И., Афонников Д.А., Казанцев Ф.В., Лихошвай В.А., Мустафин З.С., Колчанов Н.А. Программные средства для построения и анализа сложных моделей микробного метаболизма и микробных сообществ // Биотехнология. 2018. Т. 34. № С1. С. 56, DOI: 10.21519/0234-2758-2018-S-11-56.

23. Posukh O.L., Bady-Khoo M.S., Romanov G.P., Barashkov N.A., Smirnova A.A., Zytsar M.V., Maslova E.A., Danilchenko V.Y., Posukh O.V., **Lashin S.A.** Insight into genetic and social aspects of modern communities of deaf people in Siberia for forecasting the prevalence of hereditary deafness // European journal of human genetics, 2019, 27(S1):1038, DOI: 10.1038/s41431-019-0408-3.
24. Klimenko A.I., Matushkin Yu.G., Kolchanov N.A., **Lashin S.A.** Spatial heterogeneity promotes antagonistic evolutionary scenarios in microbial community explained by ecological stratification: a simulation study // Ecological modelling, 2019, V. 399, p. 66-76. DOI: 10.1016/j.ecolmodel.2019.02.007.
25. **Лашин С. А.**, Мустафин З. С., Клименко А. И., Афонников Д. А., Матушкин Ю.Г. Компьютерное моделирование эволюции микробной популяции: преодоление локальных минимумов при достижении пика на ландшафте приспособленности // Генетика, 2020, т. 56, № 2, с. 234–246, DOI: 10.31857/S0016675820020071.
26. Klimenko A.I., Matushkin Yu.G., Kolchanov N.A., **Lashin S.A.** Leave or Stay: Simulating Motility and Fitness of Microorganisms in Dynamic Aquatic Ecosystems // Biology. 2021. V10, N 10, P. 1019. DOI: 10.3390/biology10101019.
27. Romanov G.P., Smirnova A.A., Zamyatin V.I., Mukhin A.M., Kazantsev F.V., Pshennikova V.G., Teryutin F.M., Solovyev A.V., Fedorova S.A., Posukh O.L., **Lashin S.A.**, Barashkov N.A. Agent-Based Modeling of Autosomal Recessive Deafness 1A (DFNB1A) Prevalence with Regard to Intensity of Selection Pressure in Isolated Human Population // Biology. 2022. 11(2), 257. DOI: 10.3390/biology11020257.

Результаты интеллектуальной деятельности по теме диссертации:

1. Мухин А.М., Лахова Т.Н., Клименко А.И., Казанцев Ф.В., Деменков П.С., **Лашин С.А.** Биотехнологически значимые штаммы бактерий

- (МикроБиотех) / Biotechnologically significant strains of bacteria (MicroBiotech) // Свидетельство №2021621155, от 31.05.2021.
2. **Лашин С.А.**, Матушкин Ю.Г., Афонников Д.А., Мустафин З.С., Чеканцев А.Д., Дьяченко И.С., Магеррамов Э.А. Программный комплекс для моделирования эволюционно-генетических процессов в популяциях диплоидных организмов «Диплоидный эволюционный конструктор» (ДЭК) / Software package for simulation evolutionary-genetic processes in population of diploid organisms «Diploid Evolutionary Constructor» (DEC) // Свидетельство № 2015662838 от 04.12.2015.
 3. **Лашин С.А.**, Матушкин Ю.Г., Афонников Д.А., Мустафин З.С., Чеканцев А.Д., Клименко А.И., Зудин Р.К. Программный комплекс «Гаплоидный эволюционный конструктор» для моделирования популяционно-экологических процессов в микробных сообществах (ГЭК) / Software package «Haploid Evolutionary Constructor» for simulation populational-ecological processes in microbial communities (HEC)» // Свидетельство № 2015662840 от 04.12.2015.
 4. Акбердин И.Р., Казанцев Ф.В., **Лашин С.А.**, Лихошвай В.А., Ощепкова Е.А., Ри Н.А., Хлебодарова Т.М. База данных "Элементарные подсистемы: метаболизм E.coli" (ЭлСи: E.coli)/ Database "Elementary subsystems: E.coli metabolism" (ElSy: E.coli) // свидетельство №2012620064 от 16.01.2012.
 5. Лихошвай В.А., Казанцев Ф.В., Акбердин И.Р., Безматерных К.Д., **Лашин С.А.**, Подколотная Н.Н., Ратушный А.В. Компьютерная система для конструирования, расчета и анализа моделей молекулярно-генетических систем (МГСмоделлер)/ A computer system for reconstruction/ Calculation and analysis mathematical models of molecular genetic system (MGSmodeller) // свидетельство №2008612820 от 09.06.2008.

1. Обзор литературы: о проблемах и перспективах моделирования эволюционно-биологических процессов

1.1. Уровни биологической организации и их взаимосвязь

Биологические системы – это сложноорганизованные иерархические системы, различия в характерных размерах объектов которых составляют до десяти порядков, а в характерных временах процессов – до 18 порядков (Ризниченко, 2003; Шумный et al., 2006) (рис. 1.1.1).

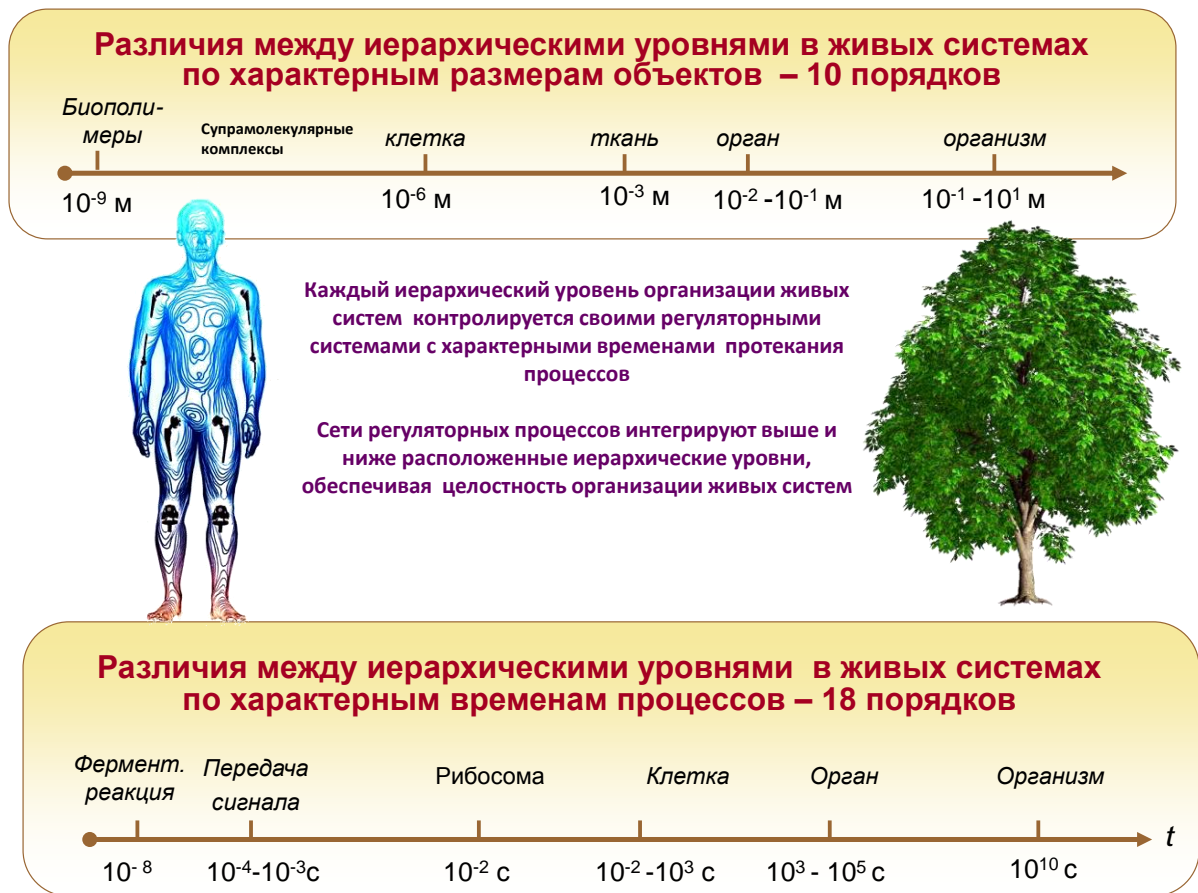


Рис. 1.1.1. Характерные размеры биологических объектов и характерные времена биологических процессов (адаптировано по: (Шумный et al., 2006)).

Выделяя наиболее проработанные с точки зрения математического описания уровни биологической организации, отметим следующие:

- молекулярно-генетический;

- клеточный;
- тканевый/органный;
- организменный;
- популяционный;
- экологический/биоценотический.

Каждый из этих уровней характеризуется определёнными объектами и процессами и, соответственно, требует для своего описания собственных математических и компьютерных моделей (см. подразделы 1.1.1-1.1.6). Отметим также, что такие понятия, как *генотип* и *фенотип*, являясь одними из наиболее часто используемых терминов в биологии, могут быть описаны, с точки зрения математического моделирования, совершенно по-разному, и, в зависимости от этого описания, формально относиться к различным уровням биологической организации.

Генотип – это совокупность генов конкретного живого организма; фенотип – совокупность его описательных признаков. В более узком смысле под генотипом подразумевается набор аллельных вариантов генов организма (Камшилов, 1972). В зависимости от контекста рассматриваемой ситуации, эти аллельные варианты могут описываться как символьные последовательности, числа и т.д. Разнообразие способов описания фенотипических признаков организма ещё шире – это могут быть дискретные признаки (например, цвет глаз или волос, цветка) или непрерывные признаки (например, рост или вес). Признаки могут быть результатом измерений, проведённых как с помощью простых приборов типа линейки или весов, так и с помощью сложных физических и/или биохимических приборов (например, морфофизиологические признаки).

Перейдём к более детальному описанию перечисленных выше иерархических уровней и обзору математических и компьютерных моделей, используемых для их исследований.

1.1.1. Молекулярно-генетический уровень организации

Молекулярно-генетическая система управления (МГСУ) клетки – совокупность её нерегулярных полимеров (ДНК, РНК и белков), а также молекулярных подсистем, выполняющих различные биохимические процессы над нерегулярными полимерами ((Ратнер et al., 1985)):

- синтез
- преобразование
- распад
- перенос и т.д.

Среди методологических подходы для исследования МГСУ можно выделить три основных:

1. Структурно-функциональный подход – концентрирует внимание на материальных свойствах макромолекул и МГСУ в связи с их функцией. Изучаются структурные, физико-химические и энергетические закономерности строения макромолекул, термодинамику и кинетику процессов и т.п.
2. Информационно-кибернетический подход – сосредотачивается на выявлении принципов организации и управления МГСУ, отвлекаясь от их структурных особенностей. Исследуются самовоспроизведение, информационные процессы, кодирование, память, надёжность, системы регулирования и т.д.
3. Эволюционный подход – выявляет пути возникновения и эволюции МГСУ в целом, различных подсистем и фракций макромолекул, а также факторы эволюции, типы эволюционной динамики и т.д. (Ратнер et al., 1985).

Частным случаем МГСУ являются т.н. генные сети – группы координированно функционирующих генов, взаимодействующих друг с другом как через свои первичные продукты (РНК и белки), так и через

разнообразные метаболиты и другие вторичные продукты функционирования генных сетей (Колчанов et al., 2013).

Модели биологических систем на молекулярно-генетическом уровне включают в себя простейшие ОДУ и системы ОДУ из нескольких уравнений, дискретные модели, построенные с использованием разных формализмов (булевы сети, сети Петри, клеточные автоматы и др.), дискретно-непрерывные модели, а также всевозможные компьютерные имитационные и агентно-ориентированные модели.

Простейшие модели МГС строились на основе уравнений химической кинетики вида (1.1.1.1) представляли собой различные способы представления *кинетического закона действующих масс* и обобщённого химико-кинетического метода моделирования (ОХКММ), предложенного В.А. Лихошваем (Лихошвай et al., 2000):

$$\frac{dX}{dt} = V(Y, K), (1.1.1.1)$$

где

X - вектор (список) управляемых переменных,

Y - вектор (список) управляющих переменных,

K – список параметров.

Допускается вхождение одних и тех же переменных в оба списка, но, в общем случае, списки X и Y не совпадают и могут вообще не пересекаться. Переменные обычно имеют смысл концентраций веществ или вероятностей реализации выделенных состояний веществ. Переменные из списка Y , не входящие в список X являются параметрами для текущей элементарной модели. Функционал V описывает закон скоростей изменения концентраций веществ из списка X .

Кинетический закон действующих масс выводится на основе теории соударений. Пусть имеется биохимическая реакция



Тогда скорость $V_{A+B \rightarrow C}$ формирования комплекса C в текущий момент времени t равна

$$V_{A+B \rightarrow C}(t) = k_1[A][B],$$

а скорость $V_{A+B \rightarrow C}$ распада комплекса C на составляющие A и B в текущий момент времени t равна

$$V_{C \rightarrow A+B}(t) = k_2[C].$$

Если имеется конкретная биохимическая схема реакций вида (1.1.1.2), то мгновенная скорость изменения концентрации любого вещества равняется сумме локальных скоростей изменения концентрации данного вещества в каждой реакции, в которой данное вещество участвует (Акбердин et al., 2013; Казанцев et al., 2009). Это простое правило позволяет легко выписать итоговую систему дифференциальных уравнений, описывающих целевую биохимическую схему, используя при этом в качестве правых частей уравнений системы (1.1.1.1) только полиномы первого и второго порядка. Теоретической основой для этого служит важнейшая для моделирования химической кинетики теорема Корзухина, которая утверждает: «Для любой совокупности неотрицательных кривых, заданных на конечном интервале времени, и любой заданной точности, существует такая (может быть не одна) биохимическая схема, составленная только из бимолекулярных и мономолекулярных реакций, что математическая модель, построенная по данной биохимической схеме, приближает заданную совокупность кривых с заданной точностью» (Жаботинский, 1974).

Важнейшим элементом построения более сложных моделей МГС с помощью ОХКММ (Лихошвай et al., 2000) является правило суммирования локальных скоростей протекания биохимических реакций: общая скорость изменения концентраций компонентов системы является суммой скоростей

изменения концентрации данного компонента во всех элементарных процессах, подробнее см. таблицу 1.1.1.

Таблица 1.1.1. Некоторые формальные блоки ОХКММ (по: (Лихошвай, 2008)).

<p>1. Обратимая бимолекулярная реакция: $x_1 + x_2 \xrightleftharpoons[k_2]{k_1} x_3$</p>	<p>$\bar{X} = (x_1, x_2, x_3), \bar{P} = (k_1, k_2),$ $\bar{F}: \frac{dx_1}{dt} = k_2 \cdot x_3 - k_1 \cdot x_1 \cdot x_2, \frac{dx_1}{dt} = \frac{dx_2}{dt} = -\frac{dx_3}{dt}.$</p>
<p>2. Необратимая мономолекулярная реакция: $x \xrightarrow{k} y_1 + y_2 + \dots + y_n$</p>	<p>$\bar{X} = (x, y_1, y_2, \dots, y_n), \bar{P} = (k),$ $\bar{F}: \frac{dx}{dt} = -\frac{dy_1}{dt} = -\frac{dy_2}{dt} = \dots = -\frac{dy_n}{dt} = -k \cdot x, n \geq 0.$</p>
<p>3. Конститутивный синтез: $\xrightarrow{k} x_1 + \dots + x_n$</p>	<p>$\bar{X} = (x_1, \dots, x_n), \bar{P} = (k), F: \frac{dx_i}{dt} = k, i = 1, \dots, n, n \geq 1.$</p>
<p>4. Обобщенная схема 1 Михаэлиса – Ментен</p>	<p>$\bar{X} = (x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n), \bar{P} = (m, k_i, k_d, a_1, \dots, a_m, b_1, \dots, b_n),$ $\bar{F}: \frac{dx_j}{dt} = -a_j \cdot Z, j = 1, \dots, m, \frac{dy_l}{dt} = b_l \cdot Z, l = 1, \dots, n,$ $z \partial_e Z = \frac{k_d k_i x_1 \dots x_m}{(k_d + x_1) \dots (k_d + x_m) - x_1 \dots x_m}, m \geq 2, n \geq 0.$</p>
<p>5. Обобщенная схема 2 Михаэлиса – Ментен</p>	<p>$\bar{X} = (e, x_1, \dots, x_m, y_1, \dots, y_n),$ $\bar{P} = (m, k, s_1, \dots, s_m, a_0, a_1, \dots, a_m, b_1, \dots, b_n), m \geq 1, n \geq 0,$ $\bar{F}: \frac{de}{dt} = -a_0 \cdot Z, \frac{dx_j}{dt} = -a_j \cdot Z, j = 1, \dots, m, \frac{dy_l}{dt} = b_l \cdot k \cdot Z, l = 1, \dots, n$ $z \partial_e Z = \frac{k s e x_1 \dots x_m}{(s + e)(s_1 + x_1) \dots (s_m + x_m) - e x_1 \dots x_m}, s = \frac{\frac{s_1}{x_1} + \dots + \frac{s_m}{x_m}}{\frac{1}{x_1} + \dots + \frac{1}{x_m}}$</p>
<p>6. Обобщенная схема 3 Михаэлиса – Ментен</p>	<p>$\bar{X} = (e_1, \dots, e_l, x_1, \dots, x_m, y_1, \dots, y_n),$ $\bar{P} = (l, m, k, s_1, \dots, s_m, a_0, a_1, \dots, a_m, b_1, \dots, b_n), m \geq 1, n \geq 0,$ $\bar{F}: \frac{de}{dt} = -a_0 \cdot Z, \frac{dx_j}{dt} = -a_j \cdot Z, j = 1, \dots, m, \frac{dy_l}{dt} = b_l \cdot k \cdot Z, l = 1, \dots, n$ $z \partial_e Z = \frac{k e_1 \dots e_l x_1 \dots x_m}{(s_1 + x_1) \dots (s_m + x_m)},$</p>

<p>7. "реакция" $x_1 + \dots + x_m \Leftrightarrow y_1 + \dots + y_n$</p>	$\bar{X} = (x_1, \dots, x_m, y_1, \dots, y_n),$ $\bar{P} = (m, k_1, k_2, a_1, s_{x_1}, s_{y_1}, \dots, a_m, s_{x_m}, s_{y_m}, b_1, r_{x_1}, r_{y_1}, \dots, b_n, r_{x_n}, r_{y_n}),$ $\bar{F}: \frac{dx_i}{dt} = -s_{x_i} k_1 x_1^{a_1} \dots x_m^{a_m} + s_{y_i} k_2 y_1^{b_1} \dots y_n^{b_n}, \quad i = 1, \dots, m,$ $\frac{dy_j}{dt} = r_{x_j} k_1 x_1^{a_1} \dots x_m^{a_m} - r_{y_j} k_2 y_1^{b_1} \dots y_n^{b_n}, \quad j = 1, \dots, n.$
<p>8. Суммирование: $x = x_1 + \dots + x_n.$</p>	$\bar{X} = (x, x_1, \dots, x_n), \quad \bar{F}: x = x_1 + \dots + x_n.$
<p>9. Пороговый автомат</p>	$\bar{X} = (x, x_1, x_2), \quad \bar{F}: x = x_1, \text{ если } x_1 \geq x_2.$
<p>10. Стохастический автомат</p>	$\bar{X} = (x, x_1, y_1, y_2) \quad \bar{F}: x = x_1, \text{ если } y_1 \geq y_2,$ <p>где u_1 – случайная величина, равномерно распределенная на отрезке $[0, 1]$, y_2 – число из интервала $[0, 1]$.</p>

Логические подходы для моделирования МГСУ, введённые ещё в 60-х годах 20-го века, основаны на применении для описания молекулярно-генетических механизмов в терминах логики и дискретной математики (Kauffman, 1969). Приведем пример описания работы генной сети в терминах булевой логики.

Генная сеть представляется в виде булевой сети – направленного графа, где вершины представляют состояния генов, а дуги – регуляторные события, т.е. действие генов на другие гены (Varbuti et al., 2020; Tran, 2016; Ратушный et al., 2005). Текущее состояние генной сети описывается списком экспрессирующихся (работающих) и не экспрессирующихся (не работающих) генов в дискретные моменты времени. Считается, что гены могут находиться только в двух состояниях: «экспрессируется» (true,1), либо «не экспрессируется» (false,0). Состояние генной сети, таким образом, представляет собой булевский вектор

$$X = (A_i), i \in \overline{1, N}$$

где A_i – состояние i -го гена. Смена состояний сети описывается с помощью булевой вектор-функции f :

$$X(t + 1) = f(X(t)) = (f_{A_1}(X(t)), f_{A_2}(X(t)), \dots, f_{A_N}(X(t))),$$

где $f_{A_i}: \{0,1\}^N \rightarrow \{0,1\}$ – описывает воздействие экспрессии всех генов на экспрессию i -го гена. Пример простой булевой генной сети показан на рисунке 1.1.2.

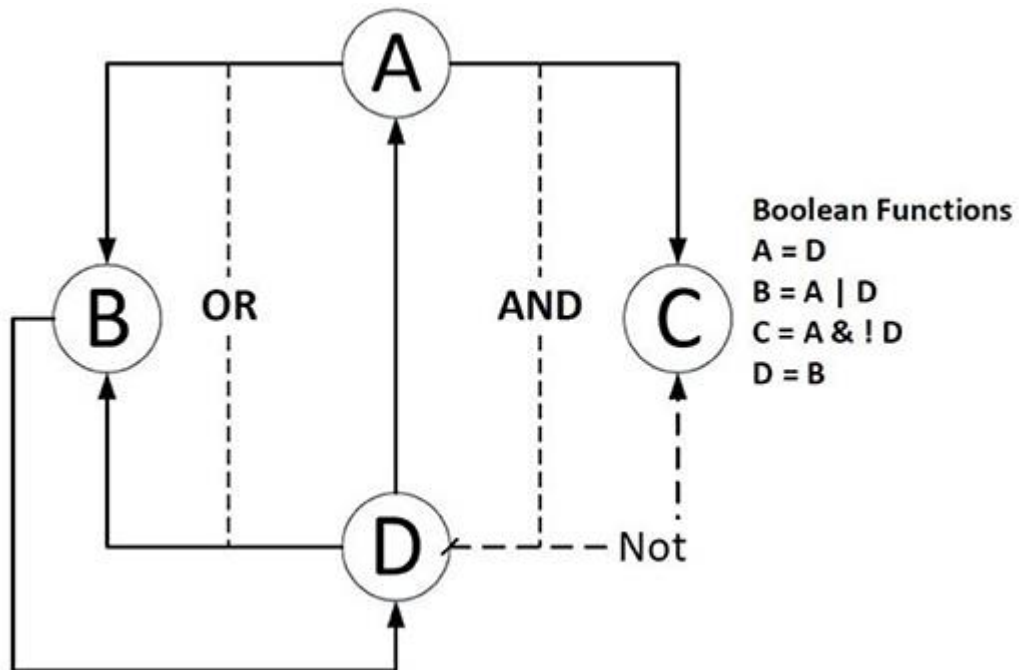


Рис. 1.1.2. Пример простой булевой генной сети. В правой части перечислены булевы функции смены состояний; пунктирная линия означает, что исходный ген должен быть ингибирован; сплошная линия означает, что исходный ген должен быть активирован (по: (Chen et al., 2018)).

В настоящее время наиболее часто используемыми подходами к описанию МГС являются различные *гибридные* методы, сочетающие применение непрерывных, дискретных методов и стохастических методов моделирования. Одним из первых подобных методов является разработанный в Новосибирске Р.Н. Чураевым ещё в 70-х годах 20-го века обобщённый пороговый метод моделирования (Чураев, 2005; Чураев, Ратнер, 1972). В данном методе сочетаются подходы теории автоматов и линейных ОДУ. В основе идеи лежит разбиение фазового пространства системы на области, внутри которых поведение системы описывается линейными **ОДУ**, условия

перехода между областями описываются булевыми функциями. Решения линейных систем «сшиваются» на границах зон. Другим примером гибридных подходов является *моделирование на основе правил* (англ. rule-based modeling), в котором модель определяется косвенно, посредством определённого набора правил (Blinov et al., 2004; Harris et al., 2016).

Дальнейшее развитие в ИЦиГ СО РАН ОХКММ и методов численного моделирования МГС на его основе также велось в парадигме гибридных подходов (Kazantsev et al., 2009, 2018).

1.1.2. Клеточный уровень организации

Математическое моделирование на клеточном уровне организации рассматривает такие биологические процессы, как транспорт веществ и энергии из окружающей среды в клетку и обратно и их метаболизм, передвижение клеток в пространстве и их деление. В части моделирования метаболических процессов клеточный уровень организации напрямую связан с молекулярно-генетическим, рассмотренном в предыдущем разделе. Появившийся в конце 20-го-начале 21-го века термин «Электронная клетка» (Ishii et al., 2004; Karr et al., 2012; Price et al., 2004; Tomita, 2001; Tomita et al., 1999; Акбердин et al., 2013) подразумевает моделирование клетки прежде всего на молекулярно-генетическом уровне биологической организации.

Тем не менее, в отличие от моделей, рассмотренных в предыдущем разделе (подглава 1.1.1), в которых интерес представляют сами МГС, их состояния, режимы функционирования и т.п., модели клетки могут фокусироваться и на других событиях и процессах, рассматривая модели МГС, входящие в них, в качестве «подручных инструментов». Например, японский системный биолог и биоинформатик Масару Томита, лидер проекта E-CELL – одного из первых успешных проектов по моделированию электронной бактериальной клетки, выделяет среди процессов клеточного уровня организации, которые можно смоделировать с помощью процессов МГС, транспорт веществ через мембрану, клеточный цикл и деление клеток, а

также развитие патологических состояний (например, для клеток эритроцитов человек) (Hucka et al., 2003; Tomita, 2001; Tomita et al., 1999). Программы, разработанные в рамках проекта E-CELL, используют гибридный подход к моделированию жизнедеятельности клетки, в частности, в версии E-CELL 3 используются системы ОДУ, стохастическое моделирование с помощью алгоритма Гиллеспи, а также специальные алгоритмы «сшивки» решений.

Другим важным клеточным процессом, для моделирования которого также могут использоваться модели МГС, является *хемотаксис* – клеточное движение. Бактериальный хемотаксис представляет собой один из самых простых и хорошо изученных примеров поведения микробов. Он позволяет плавающим бактериальным клеткам следовать за химическими веществами в окружающей среде по градиенту концентрации. Молекулярные механизмы хемотаксиса в модельной бактерии *E. coli* изучались с большой детализацией в течение последних 50 лет, с использованием большого спектра экспериментальных, в основном биофизических, методов (Berg, Purcell, 1977; Kaizu et al., 2014; Vladimirov, Sourjik, 2009). Накопление экспериментальных данных в этой области привело к созданию нескольких моделей хемотаксиса. Так, например, модель, опубликованная группой (Bray et al., 1993), представляет собой гибридную блочно-модульную модель, в которой молекулярно-генетическая компонента описывается с помощью уравнений химической кинетики, а физическая – вращение специальных белков-моторов – с помощью конечного автомата. На рисунке 1.1.3 приведена принципиальная схема этой модели. Модель воспроизводит поведение более 30 мутантов, что говорит о её высоком качестве. Модель получила развитие и была реализована в виде программного комплекса ВСТ (BacterialChemoTaxis) (Bray, Bourret, 1995), что в дальнейшем позволило описать уже более 60 мутантов (Bray et al., 2007). Кроме того, для исследования хемотаксиса использовались агентно-ориентированные модели, сочетающие этот подход со стохастическим моделированием МГС (AgentCell (Emonet et al., 2005));

гибридным моделированием, комбинирующем блоки ОДУ и стохастического моделирования, а также аппроксимацию среднего поля для модели Моно-Уаймена-Шанжэ¹ (RapidCell (Vladimirov et al., 2008), см. также рис. 1.1.4).

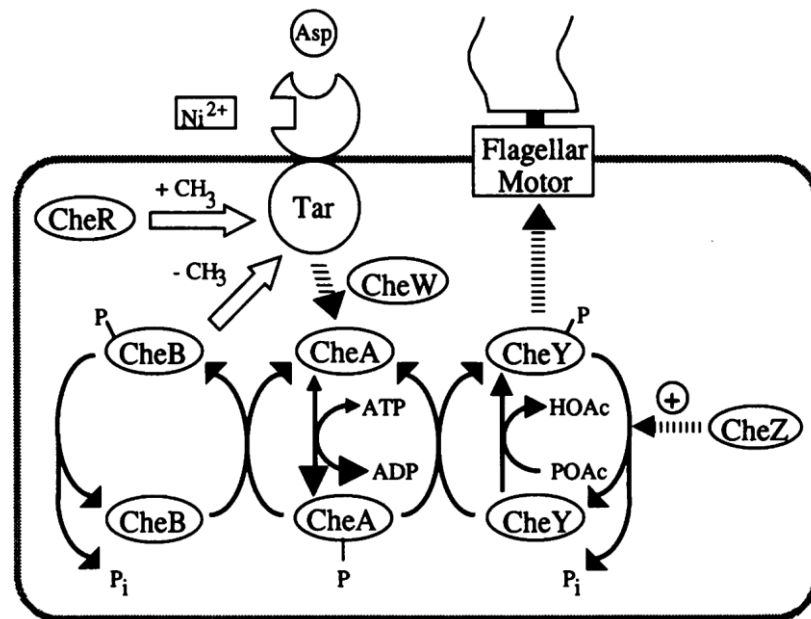
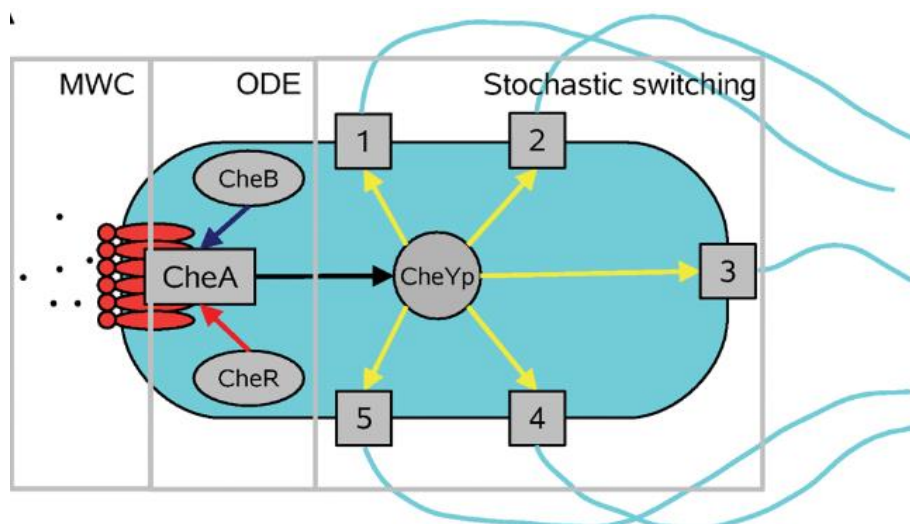


Рисунок 1.1.3. Схема пути передачи сигнала бактериального хемотаксиса. CheA, CheB, CheR, CheW, CheY, CheZ – белки хеморецепции Che (chemoreceptor proteins), передающие хемотаксический сигнал от белка-рецептора Tar до жгутикового мотора (по: (Bray et al., 1993)).



¹ Модель Моно-Уаймена-Шанжэ описывает регуляцию ферментативной активности белка, состоящего из одиноковых субъединиц, из-за аллостерического изменения его структуры (Monod et al., 1965).

Рисунок 1.1.4. Схема гибридной модели RapidCell. Активность рецепторного кластера зависит от локальной концентрации лиганда и уровня метилирования в соответствии с моделью MWC (Моно-Уаймена-Шанжэ). Связывание лигандов, переключение кластеров рецепторов, фосфорилирование белка CheY и переключение клеточных моторов рассматриваются как быстрые процессы и описываются алгебраическими уравнениями, в то время как кинетика метилирования и деметилирования относительно медленная и моделируется с помощью ОДУ. Переключение моторов моделируется стохастически (по: (Vladimirov et al., 2008)).

Молекулярные механизмы движения более сложных клеток эукариот принципиально отличаются от таковых у бактерий. Тем не менее, методический репертуар для математического моделирования этих процессов в целом такой же. Например, математическая модель регуляции волосковых клеток — рецепторов слуховой системы и вестибулярного аппарата животных и человека, описывающая ионный транспорт, мембранный потенциал и движение клеток (O’Beirne, Patuzzi, 2007), была реализована на основе уравнения Больцмана.

Несмотря на имеющийся большой задел по моделированию биологических систем на клеточном уровне организации, в большинстве работ, проводящихся в настоящее время, модели отдельных клеток используются в качестве вспомогательных инструментов. Это относится к уровням биологической организации, рассматривающим модели клеточных ансамблей — тканевом, организменном и популяционном (см. следующие разделы 1.1.3 – 1.1.5).

1.1.3. Тканевый/органный уровень организации

Биологическая *ткань* — это система клеток, сходных по происхождению, строению и выполняемым функциям в организме, а также межклеточных веществ и структур — продуктов их жизнедеятельности (Гиляров et al., 1986). *Животные ткани* разделяются на четыре основных типа: соединительные,

мышечные, нервные и эпителиальные. Ткани растений разделяются на *простые*, состоящие из клеток одного типа (например, колленхима), и *сложные*, состоящие из разных (например, эпидерма). В зависимости от классификации, выделяют два, либо три типа растительных тканей. В первом случае ткани подразделяются на *меристематические* (активно делящиеся клетки, например, в корнях или кончиках стеблей) и *постоянные* (утратившие способность к делению). Во втором случае ткани делятся на *тканевые системы: эпидермис, механическая ткань и проводящая ткань* (Гиляров et al., 1986). Совокупность различных и взаимодействующих тканей образуют органы.

Учитывая такое разнообразие свойств и функций биологических тканей и органов, репертуар математических и компьютерных моделей, представляющих этот уровень биологической организации, чрезвычайно велик. Методический арсенал для построения подобных моделей широк и включает в себя как классические методы моделирования с помощью систем ОДУ и дифференциальных уравнений в частных производных (чаще используемых для моделирования биологических тканей), так и современные методы гибридного моделирования, включающие в себя агентно-ориентированные подходы и т.п.

Математическое и компьютерное моделирование биологических тканей и органов вызывает наибольший интерес медиков и биологов в следующих контекстах: (1) моделирование развития, т.н. *морфогенеза, эмбриогенеза* и т.п.; (2) моделирование патологий (болезней) и стратегий их коррекции. В ряде случаев одна и та же модель может рассматриваться в обоих контекстах одновременно, например, модель развития раковой опухоли, позволяющая численно исследовать различные стратегии лечения.

Классической моделью теории морфогенеза считается модель «реакция-диффузия», предложенная в середине 20-го века Аланом Тьюрингом (Turing, 1952):

$$\begin{cases} \frac{\partial x}{\partial t} = P(x, y) + D_x \frac{\partial^2 x}{\partial r^2}, \\ \frac{\partial y}{\partial t} = Q(x, y) + D_y \frac{\partial^2 y}{\partial r^2}, \end{cases} \quad (1.1.3.1)$$

где r – пространственная координата, $D_x \partial^2 x / \partial r^2$ и $D_y \partial^2 y / \partial r^2$ описывают диффузию веществ x и y вдоль этой координаты. Диффузия нелинейно связанных компонент x и y в данной системе приводит не к выравниванию, а периодическому во времени и неравномерному в пространстве распределению (Ризниченко, 2003).

Модификации модели «реакция-диффузия», активно исследовались и применялись для широкого круга биологических задач, включая моделирование тканей и органов. В частности, математическая модель, реализованная в виде системы (1.1.3.2), применялась для исследования паттернов окраски шкуры животных. В зависимости от параметров модели, которым соответствуют разные сочетания морфогенов (веществ, влияющих на индивидуальное развитие организма), паттерны окраски варьируют в широком пределе (рис. 1.1.5).

$$\begin{aligned} \frac{\partial u}{\partial t} &= \gamma f(u, v) + \nabla^2 u, & \frac{\partial v}{\partial t} &= \gamma g(u, v) + d \nabla^2 v \\ f(u, v) &= a - u - h(u, v), & g(u, v) &= \alpha(b - v) - h(u, v) \\ h(u, v) &= \frac{\rho uv}{1 + u + Ku^2}. \end{aligned} \quad (1.1.3.2)$$

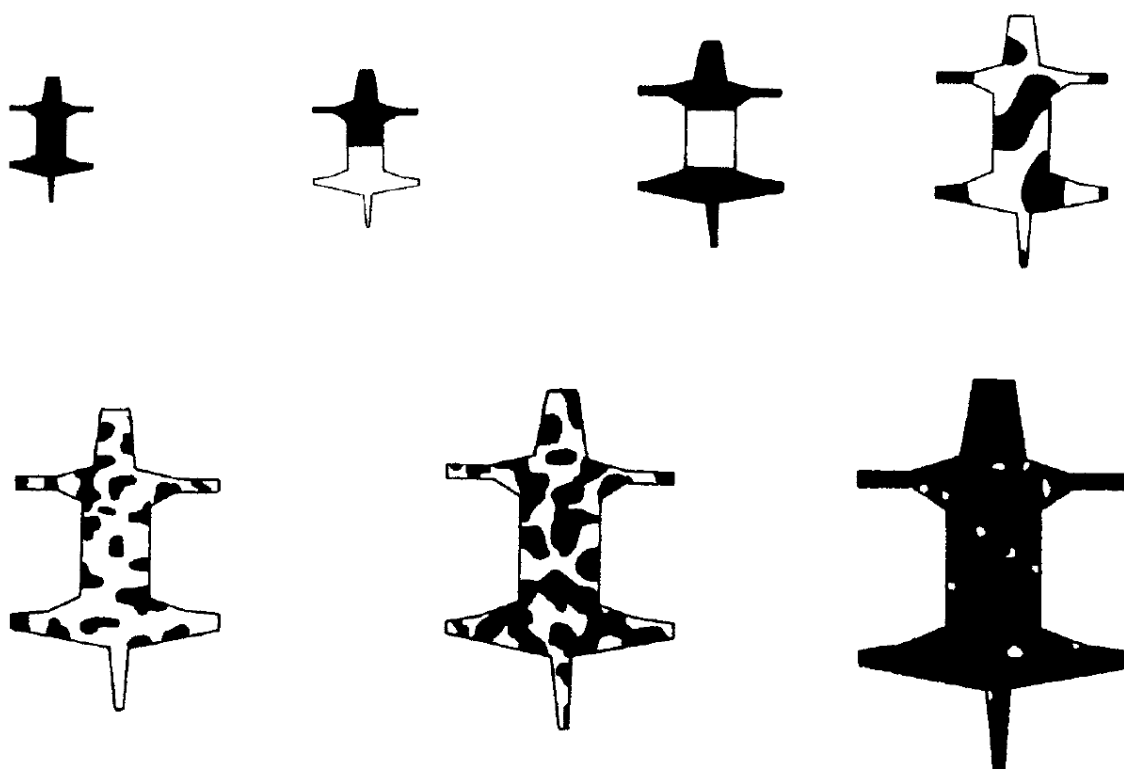


Рисунок 1.1.5. Влияние масштаба поверхности тела на пространственные паттерны, формируемые механизмом диффузии реакции (2) со значениями параметров $\alpha = 1.5$, $K = 0.125$, $\rho = 13$, $a = 103$, $b = 77$ (устойчивое состояние $u_s = 23$, $v_s = 24$), $d = 7$. Размерность домена напрямую связана с γ . Сверху вниз, слева направо, значения γ следующие: $\gamma < 0.1$; $\gamma = 0,5$; $\gamma = 25$; $\gamma = 250$; $\gamma = 1250$; $\gamma = 3000$; $\gamma = 5000$. Одна и та же форма размеров использовалась для всех моделирования (по: (Murray, 2003)).

Другой подход к моделированию биологических систем на тканевом и органном уровне организации был предложен В.А. Лихошваем и позднее получил развитие в работах под руководством В.В. Мироновой по моделированию тканей растений (Mironova et al., 2010, 2012; Novoselova et al., 2013; Pasternak et al., 2019; Лихошвай et al., 2007; Фадеев et al., 2008). Его суть состоит в описании ткани растения в виде системы ОДУ достаточно большой размерности – от сотен до нескольких тысяч уравнений. При этом отдельной клетке ткани соответствует несколько уравнений, в большинстве работ – всего четыре уравнения, описывающих динамику интересующих исследователя веществ: белков и гормонов. Клетки рассматриваются как системы с полным

перемешиванием. Перераспределение веществ между клетками посредством диффузии и процессов активного транспорта описываются уравнениями переноса вещества. Пример таких уравнений для гормона ауксин приведён в следующем фрагменте системы (ур. 1.1.3.3) из статьи (Лихошвай et al., 2007) (схема модели приведена ниже на рисунке 1.1.6 а).

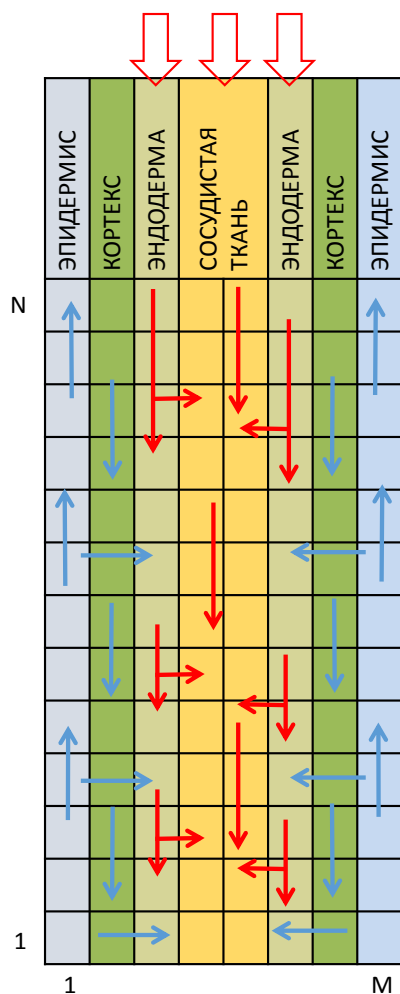
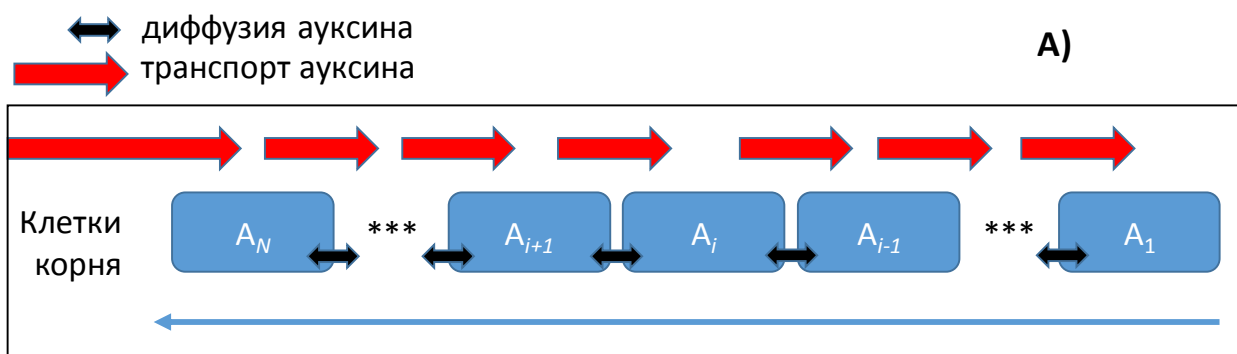
$$\begin{aligned} \frac{da_n}{dt} &= \alpha + P_t a_{n-1} - P_t a_n - K_d a_n + K_0 a_n f(a_n) \\ \frac{da_i}{dt} &= P_t (a_{i+1} + a_{i-1}) + K_0 a_{i+1} f(a_{i+1}) - 2P_t a_i - K_d a_i - K_0 a_i f(a_i), i = \overline{n-1, 2} \\ \frac{da_1}{dt} &= -P_t a_1 - K_d a_1 + P_t a_2 + K_0 a_2 f(a_2) \end{aligned} \quad (1.1.3.3)$$

где a_i – концентрация гормона ауксин в i -ой клетке корня растения (сам корень рассматривается в этой модели как одномерный массив клеток, см. также рис. 1.1.6 а); параметр α – описывает потоянный приток ауксина в систему; P_t – диффузию ауксина между клетками; K_d – параметр деградации ауксина в клетке; функция $f(a_i)$ описывает активный направленный транспорт ауксина через мембрану посредством действия специальных белков-транспортёров (ур.1.1.3.4):

$$f(a_i) = \left(\frac{\left(\frac{a_i}{q_{11}} \right)^{p_1}}{1 + \left(\frac{a_i}{q_{12}} \right)^{p_1}} \right) \times \left(\frac{1}{1 + \left(\frac{a_i}{q_2} \right)^{p_2}} \right) \quad (1.1.3.4)$$

В этой функции, свою очередь, q_{11} – константа порога активации ауксин-зависимого транспорта; q_{12} – константа порога насыщения ауксин-зависимого транспорта; q_2 – константа порога ингибирования ауксин-зависимого транспорта; p_1 и p_2 – коэффициенты нелинейности механизмов активации и ингибирования, соответственно.

В вышеупомянутой серии работ по моделированию органов и тканей растений рассматривались как одномерные (1D), так и двумерные (2D) модели (рис 1.1.6). Хотя эти модели представляют из себя системы ОДУ (зачастую большой размерности), идеологически они близки к клеточно-автоматным моделям, которые также часто используются для моделирования биологических систем на тканевом и органном уровнях организации.



Процессы, рассмотренные в модели

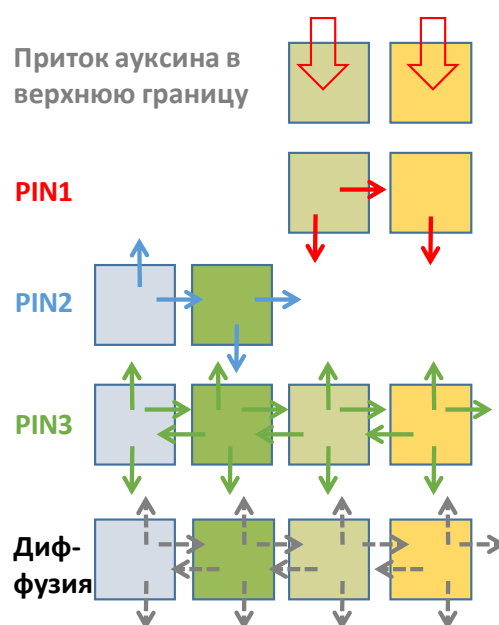


Рисунок 1.1.6. а) схема линейной 1D-модели распределения ауксина в корне растения (по: (Лихошвай et al., 2007)); б) схема 2D-модели меристемы корня *A.thaliana*, рассматривающая клетки четырёх типов (эпидермис, кортекс, эндодерма и сосудистая ткань), с детализацией процессов транспорта веществ: пассивный – путём диффузии, активный – с помощью белков PIN1, PIN2, PIN3 (по: (Mironova et al., 2012)).

Так, моделирование с помощью клеточных автоматов использовалось для исследования процессов онтогенеза (Akberdin et al., 2007; Markus et al., 1999; Pałubicki et al., 2019; Витвицкий, 2014) и вегетации (Colasanti et al., 2007; Komarov et al., 2003) растений. Активно используется этот подход и для моделирования тканей и органов животных и человека, в частности, болезненных состояний, таких как онкологические заболевания (Brüningk et al., 2019; Gevertz, Torquato, 2006; Salguero et al., 2019; Szabó, Merks, 2013), иммунные (Bezzi et al., 1997), инфекционные (Slimi et al., 2009) и другие (Talaminos-Barroso et al., 2020).

В большинстве из перечисленных выше работ, а также во множестве других, не вошедших в данный обзор, методология клеточно-автоматного моделирования используется в сочетании с другими подходами, в частности, агентно-ориентированным моделированием и моделированием, основанном на правилах (рис. 1.1.7), ранее упоминавшемся в разделе 1.1.1. Такие гибридные модели, используемые прежде всего для моделирования тканей и органов, получили название *клеточные модели Поттса* (Cellular Potts Models) (Glazier, Graner, 1993; Marée et al., 2007; Voss-Böhme, 2012).

Дальнейшее развитие методов моделирования биологических тканей и органов привело к появлению программных комплексов, библиотек и платформ, адаптирующих модели Поттса и агентно-ориентированное моделирование для решения конкретных содержательных задач в биологии. Отметим такие программные комплексы как CellSys (Hoehme, Drasdo, 2010), EPISIM (Sutterlin et al., 2013; Sütterlin et al., 2017), CompuCell3D (Swat et al., 2012) и др. Отдельно отметим проекты по моделированию целых органов, например, платформу для моделирования печени человека VirtualLiver (Drasdo et al., 2014; Holzhütter et al., 2012).

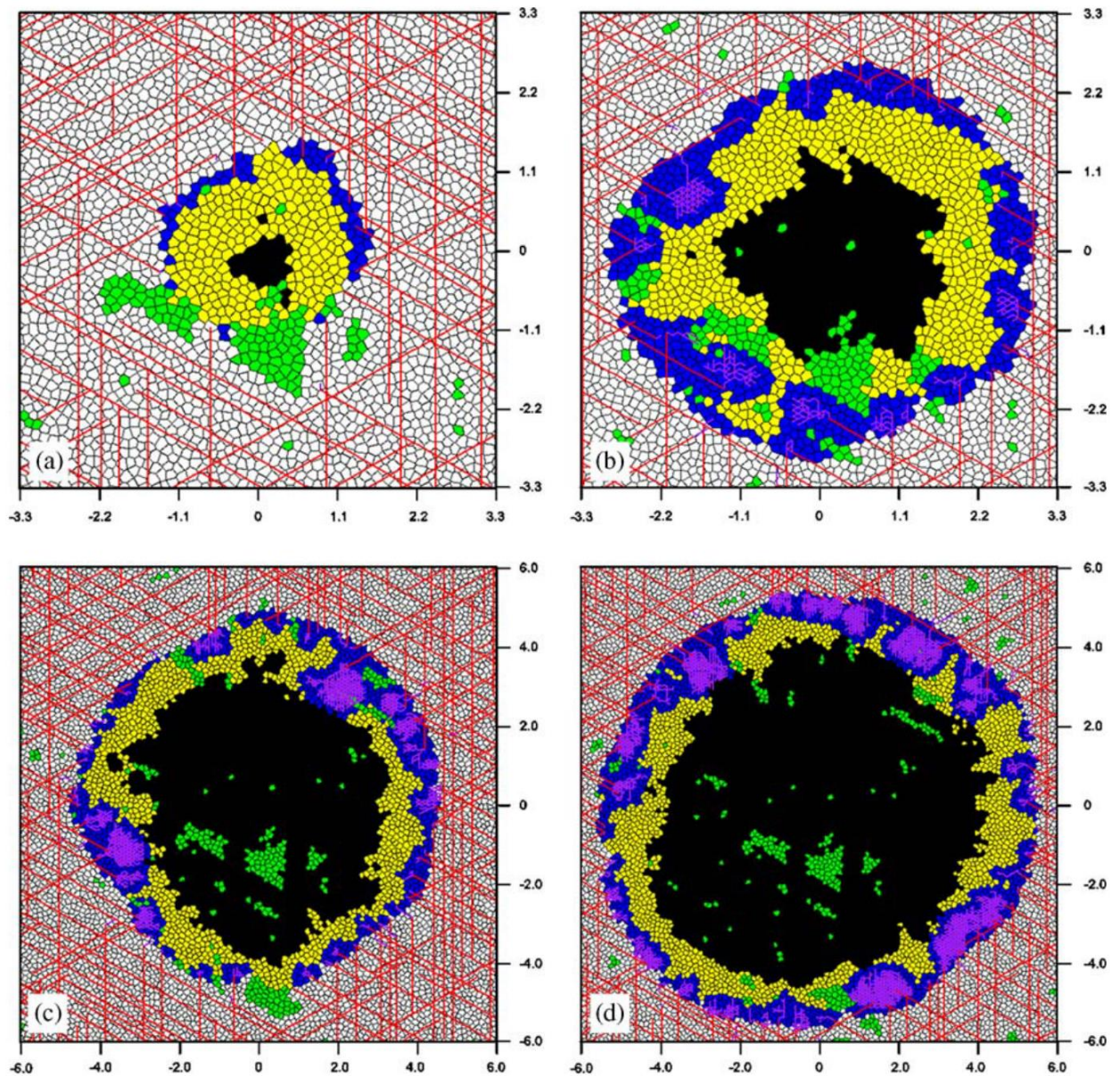


Рисунок 1.1.7. Клеточно-автоматная модель временного развития поперечного центрального среза опухоли при наличии правильно функционирующих ангиогенных механизмов. (a) Радиус опухоли на 40-й день - 1,58 мм. (b) Радиус опухоли на 70-й день - 3,09 мм. (c) Радиус опухоли на 100-й день - 4,53 мм. (d) Радиус опухоли на 130-й день - 5,84 мм. (по: (Biess et al., 2006)).

Развитие методов моделирования биологических тканей и органов подготовило почву для появления компьютерных моделей целых многоклеточных организмов, о чём будет сказано в следующем разделе.

1.1.4. Организменный уровень организации

История математического и компьютерного моделирования функционирования отдельных организмов берёт своё начало ещё в 60-х годах XX-го века. Первая и простейшая модель дерева была представлена И.А. Полетаевым и его учениками в 1965 году (Полетаев, 1966, 1965). Несмотря на простоту, модель, базируясь на физических принципах, давала ответ на вопрос «почему дерево не растёт в высоту бесконечно?». В дальнейшем на базе простейшей модели были построены и исследованы более сложные варианты модели (Карев, Скоморовский, 1999; Колобов, Фрисман, 2008).

Настоящий расцвет детальных моделей отдельных многоклеточных организмов пришёлся на последние 10-15 лет. Это во многом было обусловлено колоссальным прогрессом, которого достигла производительность компьютеров, а именно – быстродействие, объём оперативной памяти и объёмы запоминающих устройств. Стало возможным создавать реалистичные модели живых организмов. Появился термин «цифровые двойники живых организмов» (англ. Digital twin of biological organism) (Barnabas, Raj, 2020; Mieke et al., 2021; Tellechea-Luzardo et al., 2020) и т.п.

Методический арсенал, используемый для моделирования на организменном уровне биологической организации, естественным образом включает в себя все те подходы, о которых шла речь в предыдущих разделах. В зависимости от целей и задач, которые авторы «цифровых организмов» ставят перед собой, одни уровни биологической организации, а также соответствующие биологические процессы могут прописываться в модели более детально, а другие – менее. Так, например, в известном проекте OpenWorm («Открытый червь»), посвящённом созданию компьютерной модели червя *Caenorhabditis elegans* (Sarma et al., 2018; Szigeti et al., 2014; Пальянов, 2019), основное внимание уделено моделированию биомеханики движения и биофизике передачи нейроимпульсов. А, например, проект Digital

Salmon («Цифровой лосось») (Omholt et al., 2013) в большей степени сосредоточен на процессах метаболизма и онтогенеза лосося.

Отметим, что работы по созданию компьютерных моделей целых организмов – «цифровых двойников» – в настоящее время ведутся большими научными коллективами, а зачастую и консорциумами, состоящими из многих коллективов. Как правило, в рамках этих работ разрабатываются целые программные комплексы и даже программные платформы, которые способствуют развитию методической базы математической и системной биологии.

1.1.5. Популяционный уровень организации

Биологическая *популяция* (лат. *populus* – население) – совокупность особей одного вида, обладающих общим генофондом и занимающих определённую территорию (Гиляров et al., 1986). История математической биологии связана прежде всего с моделированием биологических популяций. Начиная с Л. Фибоначчи, который в своей книге по арифметике «Книга абака» предложил модель изменения размера популяции кроликов с течением времени (решение модели – знаменитые числа Фибоначчи), продолжая работами Т.Р. Мальтуса «О законе народонаселения» (Malthus, 1798) и П.Ф. Ферхюльста «Обратите внимание на закон, которому следует население в своем росте» (Verhulst, 1838), именно моделирование динамики популяций становится движущей силой развития математической биологии. Описанные в упомянутых работах модели, конечно, очень просты (см. ур. 1.1.5.1 и 1.1.5.2), но они послужили фундаментом для более сложных моделей не только в биологии, но и в других областях науки – физике, химии и др. Ниже представлено уравнение Мальтуса:

$$\frac{dN}{dt} = aN, \quad a > 0 \tag{1.1.5.1}$$

где N – численность популяции, a – коэффициент прироста численности популяции.

Уравнение Ферхюльста выглядит следующим образом:

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right), \quad r > 0, K > 0 \quad (1.1.5.2)$$

где r – коэффициент прироста численности популяции, K – максимальная численность популяции.

Модели взаимодействующих популяций были впервые предложены А. Лоткой (Lotka, 1925, 1920, 1909) и В. Вольтерра (Vol'terra, 1928; Вольтерра, 1976) и получили в дальнейшем названия «модели Лотки-Вольтерра» или «модели хищник-жертва» (ур. 1.1.5.3):

$$\begin{aligned} \frac{dx}{dt} &= x(\varepsilon_x - \gamma_{xy}y) \\ \frac{dy}{dt} &= y(\gamma_{yx}x - \varepsilon_y) \end{aligned} \quad (1.1.5.3)$$

где x – количество жертв, y – количество хищников, коэффициенты: ε_x – естественного прироста жертв, ε_y – естественной убыли хищников (в отсутствие жертв), γ_{xy} – влияния хищников на численность жертв, γ_{yx} – влияния жертв на численность хищников.

Обобщённая модель Вольтерра (ур. 1.7) позволяет рассматривать помимо отношений «хищник-жертва» другие типы взаимодействий между двумя популяциями.

$$\begin{aligned} \frac{dN_1}{dt} &= a_1N_1 + b_{12}N_1N_2 - c_1N_1^2 \\ \frac{dN_2}{dt} &= a_2N_2 + b_{21}N_1N_2 - c_2N_2^2 \end{aligned} \quad (1.1.5.4)$$

где N_1, N_2 – численность (плотность) соответствующей популяции, a_1, a_2 – скорость прироста соответствующей популяции, c_1, c_2 – коэффициент смертности соответствующей популяции, b_{12}, b_{21} – коэффициенты влияния

популяций друг на друга. В зависимости от значений параметров b_{12} и b_{21} определяется тип взаимодействия между популяциями (табл. 1.1.2).

Таблица 1.1.2. Типа взаимодействий между популяциями, рассматриваемыми в обобщённой модели Вольтерра (ур.1.1.5.4) (по: (Одум, 1975)).

Тип отношений	Влияние вида 1	Влияние вида 2	Параметры
Симбиоз	+	+	$b_{12}, b_{21} > 0$
Комменсализм	+	0	$b_{12} > 0, b_{21} = 0$
Хищник-жертва	+	-	$b_{12} > 0, b_{21} < 0$
Аменсализм	0	-	$b_{12} = 0, b_{21} < 0$
Конкуренция	-	-	$b_{12}, b_{21} < 0$
Нейтрализм	0	0	$b_{12}, b_{21} = 0$

Режимы функционирования, реализуемые в моделях (1.1.5.3-1.1.5.4), сводятся к двум типам – стационарные состояния системы и осциллирующие режимы (Ризниченко, 2017, 2002). При взаимодействии нескольких (более двух) популяций между собой богатство динамических режимов моделей резко возрастает. Так, например, в системе «хищник – две жертвы» была показана возможность существования хаотических режимов (ур. 1.1.5.5)) (Апони́на et al., 1982) (см. также рис. 1.1.8):

$$\begin{aligned}
 \frac{du_1}{dt} &= u_1 (\alpha_1 - u_1 - 6u_2 - 4v) \\
 \frac{du_2}{dt} &= u_2 (\alpha_2 - u_2 - u_1 - 10v) \\
 \frac{dv}{dt} &= -v(1 - 0,25u_1 - 4u_2 + v)
 \end{aligned}
 \tag{1.1.5.5}$$

где u_1, u_2 – безразмерные численности жертв, v – безразмерная численность хищников, α_1, α_2 – параметры, соответствующие скоростям роста численностей жертв.

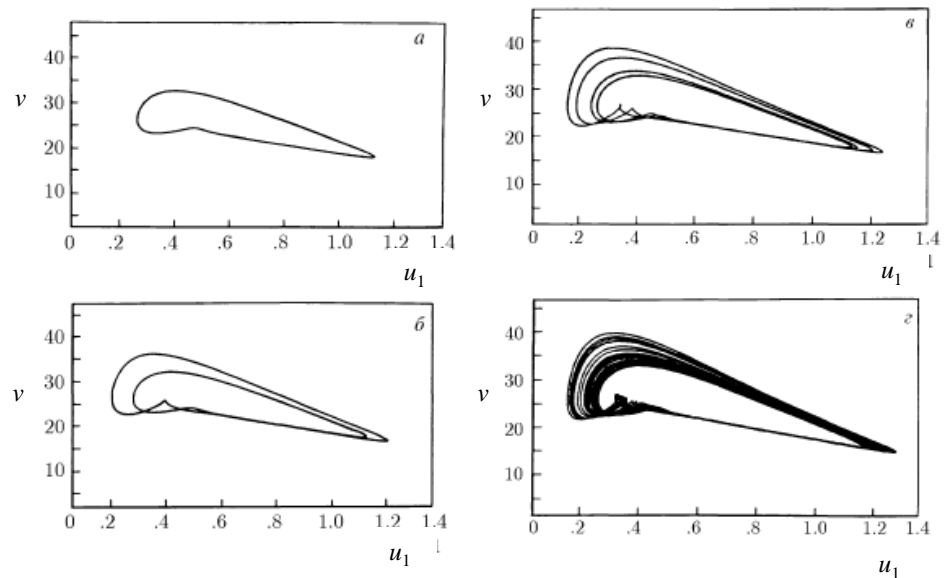


Рисунок 1.1.8. Проекция трёхмерной притягивающей траектории на координатную плоскость $\{u_1, v\}$ фазового пространства для системы (1.4) при $\alpha_1 = 2,4$ и α_2 : а) 1,7537, б) 1,7533, в) 1,7532, г) 1,7531 (результаты численного эксперимента) (по: (Апони́на et al., 1982)).

В книге А.Д. Базыкина «Нелинейная динамика взаимодействующих популяций» проведён исчерпывающий анализ моделей трёх взаимодействующих популяций типа обобщённых моделей Вольтерра (Базыкин, 2003).

Дискретные аналоги рассмотренных выше непрерывных моделей популяционной динамики – модели Морана и Рикера – рассматривают численность популяции в виде дискретной величины, изменяющейся в некоторые дискретные моменты времени, что соответствует экспериментальным данным по переписи реальных популяций. Если предположить, что численность популяции в момент времени t ($N_t, t = 0, 1, 2, \dots$) зависит от численностей в некоторые предшествующие моменты времени, то для описания динамики численности популяций можно применить аппарат рекуррентных или разностных уравнений (отображений):

$$N_t = F(N_{t-1}, N_{t-2}, \dots, N_{t-k}) \quad (1.1.5.6)$$

Решением этого уравнения является последовательность значений N_t , удовлетворяющие уравнению (1.1.5.6) при каждом t . Модель Морана и Рикера (ур. 1.1.5.7) были предложены для описания динамики численности популяций насекомых (Moran, 1950) и рыб (Ricker, 1958):

$$N_t = N_{t-1} \exp \left\{ r \left(1 - \frac{N_{t-1}}{K} \right) \right\} \quad (1.1.5.7)$$

Интересно, что даже в такой простой модели обнаруживаются очень разные режимы функционирования (рис. 1.1.9).

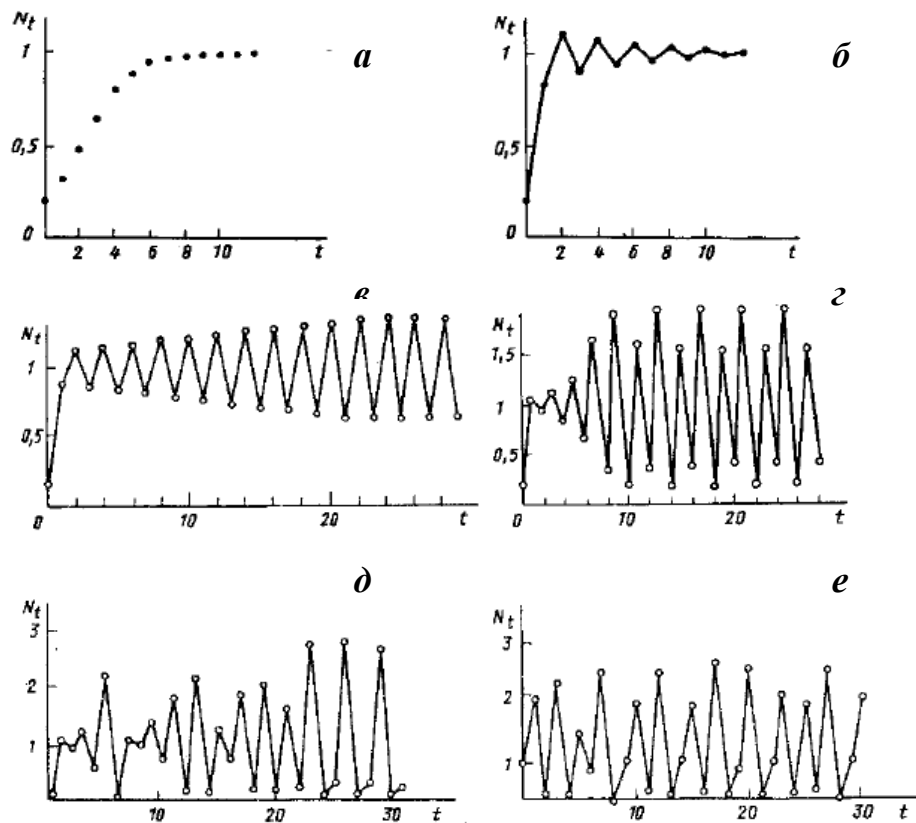


Рисунок 1.1.9. Типы динамики численности в модели популяции с неперекрывающимися поколениями при разных значениях собственной скорости роста: а – монотонный рост; б – затухающие колебания; в – двухточечный цикл; г – четырёхточечный цикл; д, е – квазистохастическое поведение (по: (Ризниченко, 2003)).

Другое большое направление в моделировании популяций – моделирование генетики популяций – было заложено классиками математической биологии: Рональдом Фишером, Дж. Холдейном, С. Райтом и другими (Fisher, 1930; Haldane, 1990, 1926, 1924; Wright, 1949, 1931). В отличие от моделей популяционной динамики, описывающих изменения численности популяций, модели популяционной генетики фокусируются на описании изменений частот аллелей (вариантов генов) в популяциях. Математический аппарат, используемый в классических моделях популяционной генетики, во многом напоминает таковой, используемый в моделях динамики популяций – это либо рекуррентные уравнения, либо ОДУ. В качестве примера, ниже приведена модель менделевской бесполой диплоидной панмиктической популяции с одним диаллельным локусом (геном, имеющим только два состояния – A_1 или A_2):

$$\frac{dp}{dt} = \frac{sp(1-p)}{1-s(1-p)} \approx sp(1-p) \quad (1.1.5.8)$$

где p – частота аллеля A_1 в популяции (тогда частота аллеля A_2 , $q = 1-p$), приспособленность гамет определяется как w_1 , w_2 , а их разница: $s = w_1 - w_2$. Уравнение 1.1.5.8 описывает изменение частот аллелей с течением времени.

До недавнего времени, в связи с недостатком широкомасштабных геномных данных, такие модели строились преимущественно исходя из неких биологически осмысленных предположений, что позволяло проводить теоретические исследования в этой области лишь на качественном уровне. Развитие методов секвенирования и последовавшее за этим появление большого объёма экспериментальных данных, привело к появлению компьютерных моделей популяционно-генетических процессов, учитывающих эти данные на количественном уровне. Имитационное моделирование эволюции генетических последовательностей получило название «моделирование коалесценций» (coalescence simulation).

Методически моделирование коалесценций представляет собой разновидность стохастического моделирования с применением различных подходов (методов Монте-Карло, марковских цепей и др.) (Salem et al., 2005). Большинство работ в этой области основывается на развитии и модификации классических моделей популяционной генетики, таких как модель Райта-Фишера (Hudson, 2002), островная модель (Wakeley, 2001) и другие. Схема последовательного усложнения популяционно-генетической модели путём добавления в неё дополнительных биологических параметров была предложена (Schaffner et al., 2005) совместно с алгоритмом верификации значений этих параметров. В настоящее время создан ряд программных пакетов для такого моделирования: GENOME (Liang et al., 2007), Migrate-n (Beerli, Palczewski, 2010), CoaSim (Mailund et al., 2005) и другие. С помощью этих пакетов были получены биологически значимые результаты. В работе (Bataillon et al., 2006) была проведена оценка эффективной численности популяции Исландии, скорости рекомбинации, а также ряда других популяционных параметров. Тестирование новых методов анализа генетических ассоциаций с болезнями человека с помощью компьютерного моделирования приводится в (Guan et al., 2009). Статистическая оценка альтернативных сценариев эволюции человека с помощью моделирования была проведена в работе (Fagundes et al., 2007).

1.1.6. Экологический/биоценологический уровень организации

Экология (от греч. oikos – дом и logos – учение), согласно Э. Геккелю – наука о взаимоотношениях организмов и их популяций друг с другом и со средой обитания (Haeckel, 1866). Экология изучает *биоценозы* и *экосистемы* как результат взаимообусловленной эволюции организмов (биоты) и биоценологической среды, с учётом осуществляемой на разных трофических уровнях деятельности популяций, определяющей мощность потоков энергии и веществ в экосистемах и общий круговорот веществ, а также об авторегуляции экосистем и их роли в биосфере планеты (Быков, 1983). В настоящее время функционирование экосистем планеты зависит от

социальных факторов и антропогенных влияний. В любом ограниченном пространстве обычно обитает множество видов, между которыми установились постоянные и сложные взаимоотношения. Иными словами, различные виды организмов, существующие в определенном пространстве с комплексом физико-химических условий, образуют сложную систему, более или менее длительно сохраняющуюся в природе. В экологии их называют *экосистемами* (Tansley, 1935) или *биогеоценозами* (Сукачѳв, 1972).

Термин «*экологическое моделирование*» включает в себя рассмотрение как моделей взаимодействия отдельных организмов со средой (*аутоэкология*), взаимодействия популяции со средой (*демэкология*), так и целых сообществ или биоценозов (*синэкология*). В основе экологических моделей лежит, прежде всего, описание переноса вещества, энергии и информации между различными частями экосистемы. Основное внимание уделяется тому, как эти части взаимодействуют, как они связаны друг с другом и влияют друг на друга, включая физическую среду.

Размерными единицами, используемыми в экологическом моделировании, обычно являются количество энергии или материи, перемещающейся через систему. Это является одним из основных отличий экологических моделей от популяционных, в которых единицами измерения обычно является размер популяции (Jørgensen, 2009).

Методический арсенал экологического моделирования во многом повторяет методы, использующиеся для описания молекулярно-генетического, клеточного, организменного и популяционного уровней, рассмотренных в предыдущих подглавах данной работы. В частности, в экологическом моделировании используется метод *анализа баланса потоков* (англ. flux balance analysis – FBA) (Allen, Gillooly, 2009; Orth et al., 2010). Основная идея метода состоит в описании потоков превращения веществ в виде задачи линейного программирования с ограничениями, что даёт возможности оценить скорости синтеза и деградации этих веществ. Пример

схематичного изображения модели, основанной на принципах баланса потоков, приведён ниже (рис. 1.1.10).

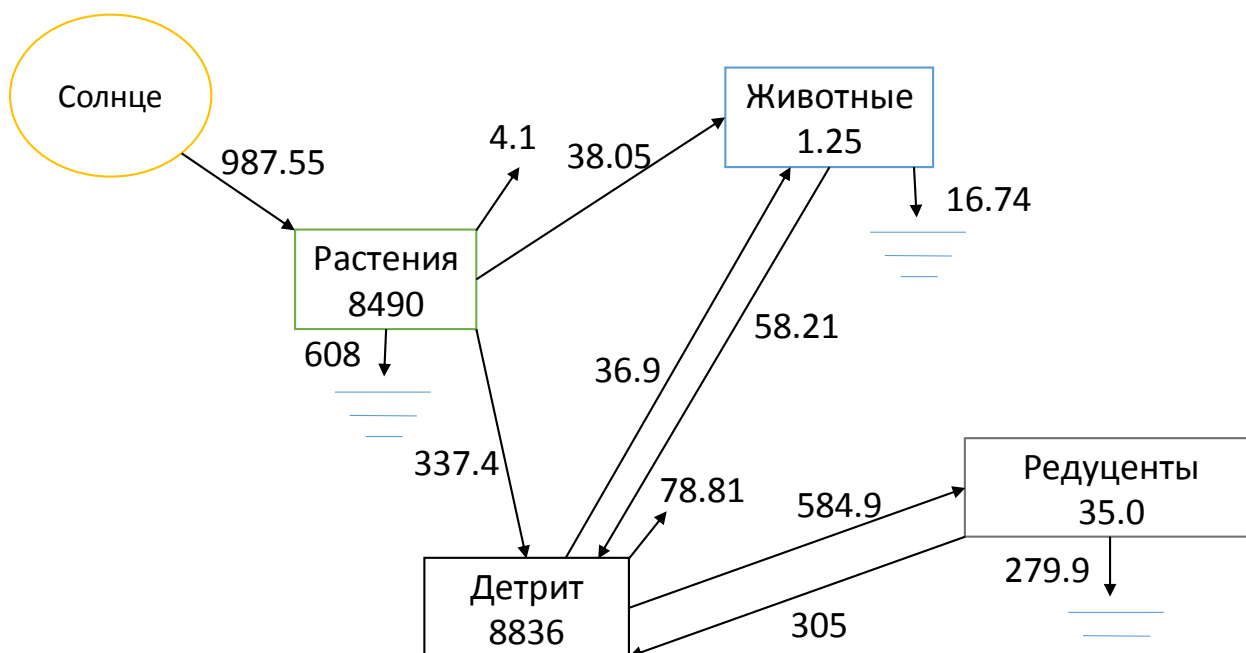


Рисунок 1.1.10. Энергетические потоки (г. фиксированного углерода на м² в год) в болоте (Россия). Детрит – это недифференцированный неживой материал, и поэтому не имеет метаболических потерь. Цифры в отдельных компартментах (прямоугольники) – запасы (г. фиксированного углерода на м²) (по: (Logofet, Alexandrov, 1984), адаптировано из: (Jørgensen, 2009)).

Активно используются в экологическом моделировании и методы ОДУ (Lavaud et al., 2020; Owolabi, Patidar, 2016; Цхай et al., 2001) и систем ДУ в частных производных (Holmes et al., 1994; Цхай et al., 2001), клеточные автоматы (Esteban, Rodríguez-Patón, 2011), стохастическое моделирование (Khatri et al., 2012; Kutalik et al., 2005; Phillips et al., 2006), графовый анализ (Fath et al., 2007) и другие методы.

Экологическое моделирование также является одним из направлений, в котором многомасштабное (англ. multiscale) и многоуровневое (англ. multilevel/multilayer) получило наиболее сильный импульс к развитию (Grimm et al., 2010, 2005; Grimm, Berger, 2016).

Частным случаем многомасштабных моделей являются, агентно-ориентированные модели, которые в экологическом моделировании традиционно называются индивидуально-ориентированными. Являясь по своей сути имитационными моделями, они покрывают значительный спектр вопросов функционирования экосистем как на уровне отдельных организмов, так и их популяций и сообществ (Doebeli, Dieckmann, 2004, 2003; Hellweger et al., 2016; Kreft et al., 1998; Widder et al., 2016). В индивидуально-ориентированной парадигме разработаны современные программные комплексы, в частности, для моделирования бактериальных сообществ. Например программа-симулятор BacSim (Kreft et al., 1998, 2001), а также программный комплекс, разработанный в работе (Xavier et al., 2005), описывают такие процессы жизнедеятельности бактерий, как поглощение (транспорт) субстратов, метаболизм, клеточное деление и гибель клеток, выделяя в качестве объекта отдельную клетку, рассматривая сообщества в виде ансамблей таких объектов. Они ориентированы, прежде всего, на исследование бактериальных сообществ в виде биоплёнок. Программа Micro-Gen Bacteria Simulator моделирует жизненный цикл растущей культуры бактерий и её взаимодействие с различными молекулами, например, антибиотиками (Murphy, Walshe, 2011). В этих программах детально описываются экологические, метаболические и популяционные компоненты, но отсутствует описание генетических процессов и наследования.

1.2. Заключение по литературному обзору

Современные методы моделирования биологических систем на разных иерархических уровнях организации базируются как на традиционных подходах (дифференциальные, алгебраические и стохастические уравнения, теория графов, клеточные автоматы и т.д.), так и на гибридных техниках, сочетающих объектно-ориентированный и агентно-ориентированный (индивидуально-ориентированный) подходы с упомянутыми выше традиционными. Хотя в совокупности они охватывают практически все

аспекты экологических (Jørgensen et al., 2009) и эволюционных (de Jong, 2002; Ferrer et al., 2008) процессов, сочетание экологической и эволюционной составляющих в рамках одной модели по-прежнему встречается достаточно редко.

Сравнивая традиционные подходы построения математических моделей с имитационным моделированием, мы видим, что и в том, и в другом случае присутствуют определённые ограничения, существенные для такого объекта моделирования, как сложно организованная биологическая система, такая, например, как микробное сообщество. В первом случае в качестве ограничения выступает статичная структура модели: число уравнений, переменных и параметров модели в процессе расчёта не меняется. В случае имитационного моделирования проблема статичности структуры модели решается, так как имитационные модели могут содержать переменное количество объектов (например, особей). Однако имитационные модели эволюции и динамики популяций весьма требовательны к размеру оперативной памяти, а также имеют большую вычислительную сложность.

Исходя из вышесказанного, разработка методов моделирования сложных иерархически организованных биологических систем, с учётом разномасштабных процессов, протекающих в этих системах, а также с учётом их эволюции – важная задача современной математической биологии. Не менее важной является и задача разработки программных комплексов, позволяющих эффективно решать содержательные задачи биологии с помощью математического и компьютерного моделирования.

2. Моделирование микробных сообществ

Бактерии и археи – безъядерные микроорганизмы (прокариоты), являющиеся древнейшими живыми организмами и неотъемлемой частью биосферы Земли. Несмотря на свой микроскопический размер (в среднем порядка 0.5–5 мкм (Нетрусов, Котова, 2018)), прокариоты обладают существенно более широким биохимическим потенциалом в сравнении с другими группами живых организмов. Прокариотические сообщества обеспечивают протекание основных биогеохимических циклов. Функционирование и эволюция бактерий и архей в основном происходит в сообществах, бактериальных матах или биоплёнках, представляющих собой сложные пространственно-распределённые слоистые структуры (Dobretsov et al., 2008; Zavarzin, 2010; Заварзин, 2003). Многие виды прокариот некультивируемы вне их естественных сообществ, поэтому математическое и компьютерное моделирование играют важную роль в исследовании особенностей их функционирования и эволюции.

С экологической и эволюционной точки зрения пространственное распределение видов играет большую роль в локальной микробной кооперации и конкуренции. Миграционные процессы в популяциях являются существенным фактором эволюции. Взаимодействуя с другими эволюционными факторами, они способны существенно влиять на динамику частот аллелей в популяциях сообщества. Поэтому зависимость эволюции и функционирования прокариотических сообществ от пространственного распределения организмов и веществ представляет собой существенный теоретический интерес. Не менее важным представляется исследование того, как различные сообщества преобразуют окружающую среду, в которой они ведут свою жизнедеятельность.

В последние годы в научной литературе опубликовано много работ по моделированию различных аспектов жизнедеятельности бактерий. В одних статьях рассматривались биологические аспекты моделирования, такие, как

связь между индивидуальным и популяционным ростом бактериальных клеток (Kutalik et al., 2005), способность поддержания биологического разнообразия системы при разных ландшафтах приспособленности и частотах мутаций (Beardmore et al., 2011). В других рассматривались различные методики компьютерного моделирования, анализировалась целесообразность, а также плюсы и минусы использования индивидуально-ориентированного моделирования вместо классических методов (DeAngelis, Mooij, 2005; Grimm et al., 2006) или же использование клеточных автоматов (Esteban, Rodríguez-Patón, 2011).

В настоящее время существует ряд компьютерных средств, предназначенных для моделирования пространственно-распределённых бактериальных сообществ. Большинство из них, такие как клеточные автоматы UMCCA (Laspidou, Rittmann, 2004a, 2004b), мультиагентная программная система AgentCell (Emonet et al., 2005), AQUASIM (Mburu et al., 2014; Wanner, Morgenroth, 2004), INDISIM (Ginovart et al., 2002) и другие, делают серьёзный упор на детальное описание собственно пространственного распределения клеток. Однако не менее интересным представляется исследование того, как процессы генетической изменчивости могут влиять на пространственную структуру сообщества. Те же программные средства, что фокусируются на процессах генетической изменчивости внутри популяций (такие как AEvol (Beslon et al., 2010; Knibbe et al., 2008)), описывают пространственную организацию сообществ недостаточно детально. К тому же, многие используемые исследователями средства моделирования на практике оказываются недоступны или же плохо документированы.

Особенностью методов и программ для моделирования эволюции микробных сообществ (далее – МС), представленных в данной работе, в сравнении с аналогами, является комплексный подход к описанию объекта исследования, позволяющий учесть, как факторы пространственного

распределения клеток и веществ, так и генетическую изменчивость на уровне отдельных популяций.

2.1. Методика моделирования микробных сообществ, основанная на концепции супериндивида

Методика моделирования микробных сообществ на основе концепции супериндивида была предложена нами ранее в диссертации на соискание учёной степени кандидата наук (Лашин, 2010) и получила развитие в последующих работах (Klimenko et al., 2016, 2015, 2019; Lashin et al., 2014, 2012a, 2010; Lashin, Matushkin, 2012; Matushkin, Lashin, 2013; С. А. Лашин et al., 2011; Лашин et al., 2012a; С А Лашин et al., 2011). Методика была реализована в виде программного комплекса «Гаплоидный эволюционный конструктор» (ГЭК), который описан в следующей подглаве. Далее мы будем обозначать и программный комплекс, и методику моделирования как ГЭК, уточняя в случае необходимости о чём именно идёт речь.

Ключевой единицей моделирования ГЭК является популяция метаболически однородных клеток, взаимодействующая с окружающей средой и с другими популяциями. В рамках данной методики рассматривается следующий набор объектов и процессов (см. также рис. 2.1.1).

Объекты:

- **Среда.** Рассматривается жидкая проточная среда с равномерно распределёнными клетками и веществами;
- **Субстраты.** Субстратом (*нутриентом*) называется вещество (*метаболит*), необходимое клеткам для их жизнедеятельности. Субстраты могут секретироваться клетками популяций (т.н. *специфические субстраты*) или поглощаться ими для нужд собственного роста. Субстраты, поступающие в среду вместе с протоком извне, называются *неспецифическими* и рассматриваются в методике как базовый источник энергии. Субстраты могут

оказывать как ускоряющее, так и замедляющее (вплоть до отрицательного) воздействие на рост популяций;

- **Популяции.** Популяция характеризуется набором метаболических систем (модельных генов), отвечающих за то, какие субстраты и с какой скоростью клетки данной популяции способны потреблять и производить.

Процессы:

- **Поглощение субстратов** из среды и секреция продуктов клетками популяций;
- **Обмен метаболитами** между популяциями внутри среды;
- **Самовоспроизведение популяций** и естественный отбор;
- **Мутации.** Мутация представляет собой изменение значения аллеля определённого гена клеткой(-ами) популяции;
- **Потеря генов.** Клетки популяции могут потерять часть генов, таким образом, образовав новую популяцию с геномом, отличным от исходного;
- **Горизонтальный перенос генетического материала.** Клетки различных популяций могут обмениваться между собой генами, образуя новые популяции.

В рамках данного подхода под популяцией понимается множество клеток, обладающих одинаковыми свойствами в отношении потребления и производства субстратов, т.е. одними и теми же метаболическими системами. Клетки считаются принадлежащими одной популяции (одному виду), если они:

1. потребляют (утилизируют) одинаковое количество неспецифических и специфических субстратов;
2. продуцируют одинаковое количество специфических продуктов;

3. имеют одну и ту же модель метаболизма, описывающую превращение множества входных субстратов в множество выходных с заданной стехиометрией;
4. обладают одной трофической стратегией;
5. обладают одной стратегией синтеза.

Таким образом, именно популяция в методике ГЭК выступает в роли супериндивида.

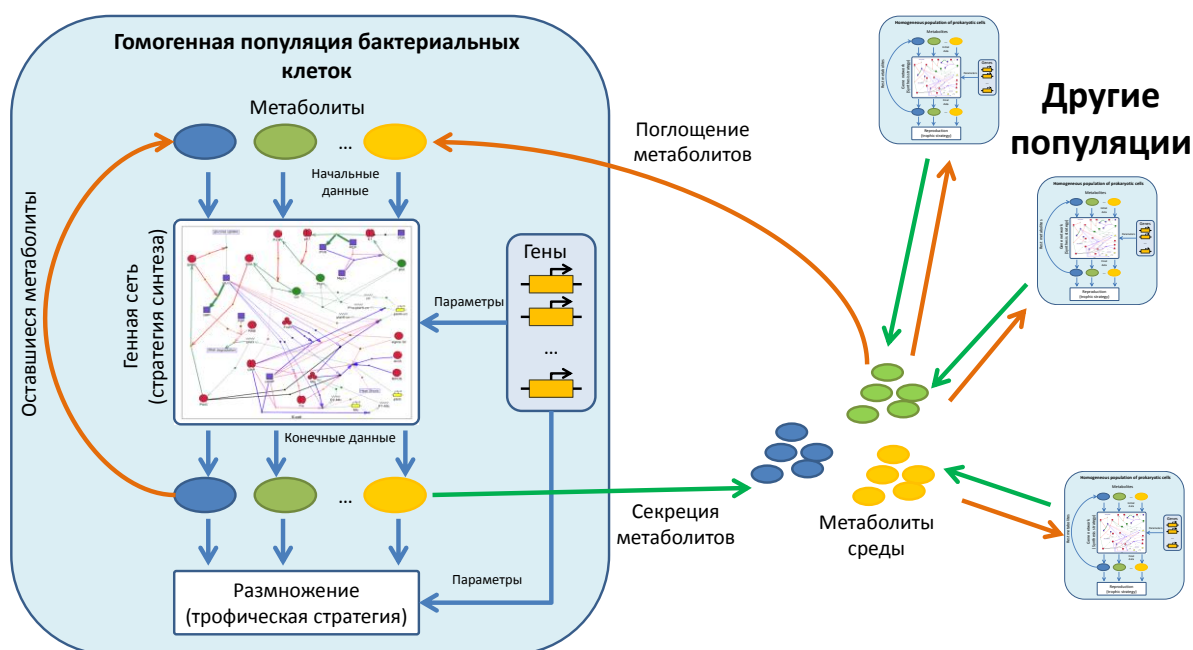


Рис. 2.1.1. Основные объекты и процессы ГЭК.

Под трофической стратегией понимается модель популяции, описывающая закон изменения её численности за одно поколение в зависимости от потреблённых клетками данной популяции субстратов. В рамках данной работы были использованы компенсаторная (ур. 2.1.1), некомпенсаторная (ур. 2.1.2) и ингибиторная (ур. 2.1.3) трофические стратегии.

$$F(n_0, \vec{S}, r_0, \vec{C}, P, G) = \sqrt{r_0 n_0(P) \sum_{i \in I_{consumed}} c_i s_i(P) - k_{death} P^2} \quad (2.1.1)$$

$$F(n_0, \vec{S}, r_0, \vec{C}, P, G) = P \frac{\left(\frac{n_0/P}{K_{01}(r_0)}\right)^{\gamma_0}}{1 + \left(\frac{n_0/P}{K_{02}(r_0)}\right)^{\gamma_0}} \prod_{i \in I_{consumed}} \frac{1 + \left(\frac{s_i/P}{K_{i1}(c_i)}\right)^{\gamma_i}}{1 + \left(\frac{s_i/P}{K_{i2}(c_i)}\right)^{\gamma_i}} - k_{death} P^2 \quad (2.1.2)$$

$$F(n_0, \vec{S}, r_0, \vec{C}, P, G) = a_{basal}(n_0)P - \sqrt{\sum_{i \in I_{consumed}} c_i s_i(P)} - k_{death} P^2 \quad (2.1.3)$$

где $I_{consumed}$ – набор индексов специфических субстратов, потребляемых клетками популяции; \vec{S} – вектор количественных значений потребляемых специфических субстратов; n_0 – количество неспецифического субстрата, потребляемого клетками популяции; r_0 – генетически предопределённая скорость утилизации неспецифического субстрата; P – численность популяции; k_{death} – коэффициент смертности; K_{ij} – коэффициенты, описывающие эффективность влияния субстратов на рост популяции (описывается в виде линейной комбинации одного или нескольких признаков), γ_i – коэффициенты, описывающие нелинейность влияния субстратов на рост популяции, a_{basal} – «естественный прирост» популяции.

Основные термины и понятия, которые используются в рамках методики ГЭК:

- **признак** – константа скорости синтеза или утилизации какого-либо определённого субстрата. Считается, что каждый признак однозначно определяется одним **геном**. **Ген** в данном случае рассматривается как единица наследования;
- **аллель** – это вариант гена, т.е. конкретное значение соответствующей константы;
- **генотип особи** представляет собой набор аллелей, разделённый на три группы. Первая группа (c_i) характеризует эффективность утилизации специфических субстратов (s_i), вторая группа (d_i) – скорость выработки субстратов и третья группа (r_i) – эффективность утилизации неспецифических субстратов;

- **мутация** представляет собой изменение величины соответствующего признака, что может интерпретироваться, как перевод гена в другое состояние (аллель).

С использованием данной терминологии формулируется понятие **мономорфной (изогенной) популяции** – популяции «генетически идентичных» клеток, где у всех клеток соответствующие гены представлены одними и теми же аллелями. Общий для всех клеток такой популяции генотип называется **генотипом мономорфной популяции**.

Аналогично понятию мономорфной популяции, мы определяем понятие **полиморфной популяции**, которое можно рассматривать как совокупность мономорфных субпопуляций. Клетки в полиморфной популяции имеют одно и то же разнообразие генов, в то время разные клетки могут иметь разные аллели одного или нескольких генов. Для полиморфной популяции характерен «обобщённый геном» – набор генетических спектров популяции. Генетический спектр – это многомерное распределение частот встречаемости аллелей для всех генов в популяции. Мутации в терминах генетического спектра означает изменение этого распределения (таким образом, возможно формирование новых аллелей). Изменение размера полиморфной популяции рассчитывается по схеме, предложенной ещё в (Лашин, 2010), согласно которой полиморфная популяция разбивается на множество мономорфных субпопуляций. Затем для каждой субпопуляции рассчитывается прирост по формулам, аналогичным приведенным выше (например, 2.1.1-2.1.3). После этого мономорфные популяции вновь объединяются в полиморфную популяцию. Полиморфная популяция разбивается на «воксели» пропорционально доле каждого аллеля для всех генов в геноме. Субстраты, потребляемые полиморфной популяцией, делятся пропорционально размерам отдельных мономорфных популяций. Примечательно, что рост каждой отдельной мономорфной популяции может заметно отличаться от роста других субпопуляций. Это зависит от генотипа мономорфной популяции,

размера популяции и количества определенных субстратов, потребляемых популяцией. Следовательно, доля аллелей в популяции может меняться (что может быть интерпретировано как адаптация популяции к определенным условиям).

Данная схема описания генетического разнообразия микробных популяций успешно использовалась во многих исследованиях ввиду вычислительной эффективности и низким требованиям к размеру оперативной памяти. Однако, в ряде случаев, эта схема могла служить источником систематических ошибок моделирования, поэтому в рамках данной работы была предложена более точная схема (см. раздел 2.4).

Помимо популяций бактериальных клеток в ГЭК рассматриваются популяции умеренных бактериофагов. Жизненный цикл умеренного фага в методике ГЭК подчиняется следующим правилам переключения между лизогенным и литическим циклами: в пессимальных условиях вирус встраивается в геном клеток в виде профага, а в оптимальных условиях, способствующих быстрому росту популяции бактерий, бактериофаг вызывает лизис клеток с образованием новых фаговых частиц (подробнее см. раздел 2.5). Также возможны переключения между режимами в случае изменения условий, в которых находится заражённая популяция (подраздел 2.7.4).

2.2. Расширенная версия ГЭК для моделирования пространственно-распределённых сообществ

2.2.1. Моделирование пространственно-распределённой среды обитания

Расширенная версия программного комплекса ГЭК реализует развивает методологию, разработанную нами ранее для 0D-сред (т.е. сред с полным перемешиванием) (Lashin, Matushkin, 2012). В случае 0D среда обитания имеет только один пространственный параметр – объём, и никаких других. Предполагается, что клетки и вещества равномерно распределены в среде. Будем называть отдельную 0D-среду – «точечной средой» или «ячейкой». Пространственно распределённая (1D, 2D или 3D) среда может рассматриваться как набор таких ячеек. Расчёт новых состояний для каждой

ячейки включает моделирование следующих процессов: потребление питательных веществ (субстратов или нутриентов), утилизация питательных веществ, размножение, синтез субстратов или нутриентов и их секреция. Этот этап унаследован от прежней версии ГЭК 0D. Отметим, что описанная выше процедура является расчёта новых состояний является независимой и может выполняться параллельно (о чём более подробно написано в разделах 2.5.3, 2.5.4).

Дополнительным этапом моделирования в расширенной методике является перераспределение субстратов и клеток во всей среде, т.е. моделирование процессов переноса веществ и клеток между ячейками среды. Для переноса веществ используются два процесса: диффузия и поток. Для наглядности рассмотрим примеры для 2-мерного случая. Процессы в одномерном и трехмерном случае описываются аналогично, путем изменения размерности векторов и связности узлов. Пусть среда обитания описывается набором матриц

$$E_t^s = \begin{pmatrix} e_{11} & \cdots & e_{1n} \\ \vdots & \ddots & \vdots \\ e_{m1} & \cdots & e_{mn} \end{pmatrix}, \quad (2.2.1)$$

где e_{ij} количество субстрата s в (i, j) ячейке в момент времени t . Для того чтобы определить пространственное распределение вещества на следующем шаге, то есть вычислить E_{t+1}^s , необходимо учесть влияния протока и диффузии. В матричной записи это представлено следующим образом:

$$E_{t+1}^s = E_t^s + F(E_t^s) + D(E_t^s), \quad (2.2.2)$$

где $F(E_t^s)$ и $D(E_t^s)$ – это матрицы переходов состояний для веществ (их элементы могут быть положительными и отрицательными) в ячейке (i, j) для протока и диффузии, соответственно. Помимо протока и пассивного транспорта, действующего аналогично диффузии, транспорт клеток определяется хемотаксисом (см. подробно в подразделе 2.2.4).

На каждом шаге итерации перемещение клеток и питательных веществ

происходит между четырьмя соседними узлами для 2D-случая для каждого узла (2 соседа для 1D, 6 соседей для 3D). Поэтому перемещение через n узлов при таком подходе требует n итераций. Одномерный и трёхмерный случаи формализуются аналогично. В последнем случае матрицы $E_t^S = (e_{ijk})$ являются трёхмерными, а узлы – шестисторонне связанными.

2.2.2. Моделирование потока

Проток описывается вектором соответствующего размера, состоящим из коэффициентов интенсивности потока. Компоненты вектора определяют порции клеток и веществ, перемещаемых потоком в каждом направлении. Таким образом, пользователь может задать направление потока и его интенсивность.

Проток в среде задаётся вектором $\vec{fl} = \begin{pmatrix} fl_x \\ fl_y \end{pmatrix}$ в декартовой системе координат (в случаях одно- и трёхмерных сред векторы содержат одну и три компоненты, соответственно). Этот вектор описывает долю вещества, переносимую из ячейки за итерацию, нормированную с учётом выполнения равенства $fl_x + fl_y + 4k_d = 1$, где k_d – параметр скорости диффузии. Очевидно, что $|fl_x| + |fl_y| \leq 1$, поскольку поскольку количество удаляемого из ячейки вещества не может превышать его общее содержание.

Введём обозначение $\vec{fl}_{abs} = \begin{pmatrix} |fl_x| \\ |fl_y| \end{pmatrix}$ и обозначим базисные векторы как $\vec{i} = (1,0)$ и $\vec{j} = (0,1)$. Определим элемент матрицы $F(E_t^S) = (f_{ij})$ как скалярное произведение двух векторов: $f_{ij} = \vec{fl}_{abs} \cdot (c_x - e_{ij}, c_y - e_{ij})$, где

$$c_x = \begin{cases} e_{i,j-1}, & \text{if } \vec{fl} \cdot \vec{i} \geq 0, \\ e_{i,j+1}, & \text{иначе} \end{cases}, \quad c_y = \begin{cases} e_{i-1,j}, & \text{if } \vec{fl} \cdot \vec{j} \geq 0, \\ e_{i+1,j}, & \text{иначе} \end{cases}$$

Таким образом мы получаем значение слагаемого $F(E_t^S)$ из формулы (2.2.2).

2.2.3. Моделирование диффузии

Диффузия – это процесс ненаправленного распространения вещества, обусловленный случайным движением молекул в окружающей среде. В данном исследовании параметр скорости диффузии описывает часть вещества,

диффундирующего из текущей ячейки. Для описания эффекта диффузии, рассмотрим матрицу $D(E_t^s) = (d_{ij})$, где $d_{ij} = k_d \cdot (e_{i,j-1} + e_{i,j+1} + e_{i-1,j} + e_{i+1,j} - 4e_{ij})$, а k_d – параметр скорости диффузии, т.е. доля вещества, диффундирующая из текущей ячейки в одну из четырёх соседних ячеек (в 2D случае). Отметим, что в случае 3D-среды учёт 6 примыкающих граней к кубической ячейки устроен аналогично (в одномерном случае граничащих грани 2).

Согласно первому закону Фика, диффузионный поток вещества, переносимого молекулярной диффузией через поверхность S за время t , пропорционален градиенту концентрации:

$$J = \frac{m}{St} = -D \frac{dC}{dx}$$

где D – коэффициент диффузии. Введённый ранее коэффициент k_d подчиняется соотношению

$$k_d = M_a \cdot D$$

где M_a – молярная масса диффундирующего вещества. Например, если диффундирующее вещество – это глюкоза ($M_a = 180.156$ г/моль), размер отдельной точечной среды $V = 0.006$ л (т.е. $S = 3.3 \cdot 10^{-4}$ м² и $l = 1.817 \cdot 10^{-2}$ м²), а характерное время одной итерации $t = 30$ минут, то:

$$k_d = \frac{M_a \cdot dx}{S \cdot t \cdot dc} = \frac{M_a \cdot g \cdot 1.817 \cdot 10^{-2} \cdot m}{\frac{mol}{0.001 \cdot m^3} \cdot 3.3 \cdot 10^{-4} \cdot m^2 \cdot 1800 \cdot s} = \frac{M_a \cdot g \cdot m^2 \cdot 1.817 \cdot 10^{-5}}{mol \cdot 3.3 \cdot 0.18 \cdot s} = M_a \cdot 0.306 \cdot \frac{g \cdot cm^2}{mol \cdot s}$$

Таким образом, если принять параметр скорости диффузии $k_d = 0.001$, то соответствующий коэффициент диффузии может быть получен следующим образом:

$$D = \frac{k_d}{M_a} = \frac{0.001 \cdot 0.306}{180.156} \cdot \frac{cm^2}{s} = 1.6979 \cdot 10^{-6} \frac{cm^2}{s}$$

что является физически и биологически осмысленным значением

(Лобанов, 2017).

2.2.4. Моделирование хемотаксиса

Поток и диффузия одинаково затрагивают как микроорганизмы, так и вещества. Однако для адекватного моделирования движения клеток необходимо учитывать следующие факторы. Во-первых, поскольку микробные клетки могут прилипать к бактериальной плёнке или бактериальному мату, двигаться может лишь ограниченная часть «свободно» плавающих, планктонных, клеток. Во-вторых, это активное движение посредством хемотаксиса. В ГЭК часть свободных клеток описывается собственным коэффициентом k_{free} , а хемотаксис – с помощью отдельной вычислительной процедуры. Аналогично коэффициентам интенсивности диффузии и потока, доля активно движущихся клеток m определяется как общепопуляционная характеристика. Таким образом, $k_a = k_{free} \cdot m \cdot c$ – это доля популяции, которая перемещается посредством хемотаксиса, где c – нормировочный коэффициент.

Перераспределение клеток в окружающей среде описывается аналогично тому, как это сделано для нутриентов:

$$E_{t+1}^p = E_t^p + F(E_t^p) + D(E_t^p) + A(E_t^p)$$

где $A(E_t^p)$ - это матрица, описывающая изменения размера клеточной популяции в ячейке (i,j) в результате активного движения клеток популяции p . Элементы этой матрицы определяются следующим образом:

$$a_{ij} = p_{in} - p_{out}$$

где p_{in} – это количество клеток, принадлежащих к данной популяции, которые мигрируют в текущую ячейку из соседних ячеек с менее благоприятными условиями; p_{out} – это количество клеток, принадлежащих этой популяции, которые мигрируют из текущей ячейки в соседние ячейки с более благоприятными условиями.

Сначала мы оцениваем «привлекательность» условий окружающей среды в соседних ячейках на основе разницы между аттрактантами и репеллентами в соседней и текущей ячейке. Значения «привлекательности» для всех соседей рассчитываются следующим образом:

$$A_{pq} = \sum (b_k * c_k)$$

где c_k – это концентрация k -го нутриента, а

$$b_k = \begin{cases} 1, & \text{если } \mathbf{k} - \text{й нутриентр является аттрактантом для } \mathbf{p}, \\ -1, & \text{если } \mathbf{k} - \text{й нутриент является репеллентом для } \mathbf{p}, \\ 0, & \text{если } \mathbf{p} \text{ не потребляет } \mathbf{k} - \text{й нутриент.} \end{cases}$$

После этого формируются два списка соседей. В первом списке находятся «отталкивающие» ячейки, которые обладают меньшими значениями привлекательности для клеток, чем текущая, а во втором списке – «притягивающие» ячейки, которые обладают большими значениями привлекательности. Изменение численности популяции в текущей ячейке определяется как сумма входящих (первый список) минус сумма убывших (второй список) клеток. Кроме того, доля активно перемещающихся клеток делится на все направления, а доля направления пропорциональна нормированному весовому коэффициенту, определяемому значениями привлекательности соответствующей соседней ячейки. Отметим, что при нормализации значений $\vec{f}l$ и k_a , учитывается влияние k_a .

Наконец,

$$p_{out} = \sum_{q \in N(i,j)} k_a \cdot \frac{A_q}{A_{ij} + \sum_{t \in N(i,j)} A_t} \cdot e_q$$

где k_a – часть популяции, которая перемещается посредством хемотаксиса; $N(i,j)$ – множество соседних ячеек (по отношению к текущей); e_q – численность (количество клеток) p -й популяции в q -й ячейке ($q \in N(i,j)$); A_q – значение привлекательности q -й ячейки, A_{ij} – привлекательность текущей ячейки. В итоге, p_{in} рассчитывается на основе значений p_{out} в соседних узлах:

$$p_{in} = \sum_{t \in N(i,j)} w_q \cdot p_{out}^t$$

где w_q – часть p_{out} из соседней ячейки q , которая мигрирует в текущую ячейку.

Одно- и трёхмерные случаи формализуются аналогичным образом. Во втором случае матрицы $E_t^S = (e_{ijk})$ являются трёхмерными, а ячейки – шестисторонне связанными.

2.2.5. Факторы пространственного перераспределения и энергетические резервы клеток

В ГЭК энергетический запас клеток представлен потреблёнными молекулами неспецифических субстратов на одну клетку в конкретной микробной популяции (описание потребления питательных веществ см. в следующей подглаве). При этом описанные выше факторы перераспределения, действующие на микробные клетки, влияют на текущий запас энергии в любой конкретной ячейке. Эти факторы можно разделить на две части – пассивный транспорт (ненаправленный, обусловленный протоком) и активная миграция посредством хемотаксиса. Обозначим r_{pass_immigr} и r_{pass_emigr} приход и расход энергетических запасов за счёт пассивного транспорта клеток популяции в текущую ячейку и из неё, соответственно. Тогда

$$r_{pass_emigr} = \frac{r_n[i,j]}{P_n[i,j]} \cdot P_{pass_out}$$

и

$$r_{pass_immigr} = \sum_{q \in N(i,j)} \left(\frac{r_n^{(q)}}{P^{(q)}} \cdot w_{(q)} \cdot P_{pass_out}^{(q)} \right)$$

где $N(i,j)$ – множество соседних ячеек (по отношению к текущей); $P^{(q)}$ – это численность популяции в соседней ячейке q , $P_{pass_out}^{(q)}$ соответствует общему пассивному оттоку клеток из конкретной ячейки q (см. раздел 2.2.3), а w_q – это часть $P_{pass_out}^{(q)}$ из соседней ячейки q , которая поступает в текущую ячейку.

Аналогичным образом обозначим r_{act_immigr} и r_{act_emigr} – приход и расход

энергетических резервов через иммиграцию и эмиграцию клеток популяции в текущую ячейку и из неё посредством хемотаксиса, соответственно:

$$r_{act_{emigr}} = \frac{r_n[i,j]}{P_n[i,j]} \cdot P_{act_{out}}$$

и

$$r_{act_{immigr}} = \sum_{q \in N(i,j)} \left(\frac{r_n^{(q)}}{P^{(q)}} \cdot w_{(q)} \cdot P_{act_{out}}^{(q)} \right)$$

где $N(i,j)$ – множество соседних ячеек (по отношению к текущей); $P^{(q)}$ – это численность популяции в соседней ячейке q , $P_{act_{out}}^{(q)}$ соответствует общему количеству клеток, мигрировавших из конкретной ячейки q посредством хемотаксиса (см. раздел 2.2.4), а w_q – это часть $P_{act_{out}}^{(q)}$ из соседней ячейки q , которая поступает в текущую ячейку.

2.3. Моделирование питания и синтез питательных веществ

2.3.1. Моделирование потребления питательных веществ

Процессы секреции и потребления питательных веществ (нутриентов) описываются в методике и программном комплексе ГЭК базируясь на следующих предположениях:

1. Питание бактерий осуществляется путем переноса питательных веществ, диффундирующих через клеточную стенку, контактирующую с окружающей средой (пинотрофия);
2. Соотношение площади поверхности и объёма клетки имеет значение для поглощения питательных веществ.

Основная идея, лежащая в основе подмодели секреции и потребления питательных веществ ГЭЕ, заключается в том, что существует объём $V_{consumed}$, протекающий через клетку в течение определенного периода времени (одна итерация, которая соответствует среднему времени деления клетки). Типичное значение $V_{consumed}$ для бактерии *E.coli* составляет около $5,4 \cdot 10^{-12}$ литров в полчаса (Sundararaj et al., 2004).

В предположении, что значения $V_{consumed}$ одинаковы для всех клеток

популяции, что в случае среды с полным перемешиванием малого объёма (точечные среды или ячейки ГЭК) вполне оправдано, изменение концентрации нутриента в ячейке ($S_{env,i}$) описывается следующим образом:

$$S_{env,i}(t + 1) = S_{env,i}(t) \cdot \left(1 - P \frac{V_{consumed}}{V_{total}}\right) \quad (2.3.1)$$

где P – размер популяции (количество клеток); V_{total} – объём ячейки.

Количество молекул питательного вещества, потребленного клетками, принадлежащими к рассматриваемой популяции ($S_{pop,i}$), рассчитывается следующим образом:

$$S_{pop,i}(t + 1) = S_{pop,i}(t) + S_{consumed}(t) \quad (2.3.2)$$

где

$$S_{consumed}(t) = P \cdot V_{consumed} \cdot S_{env}(t) \cdot N_A$$

N_A – число Авогадро.

Если в узле слишком много клеток (т.е. верно неравенство 2.3.3), они не смогут отфильтровать весь объём $V_{consumed}$.

$$\sum_{i \in I_{POP}} P_i \cdot V_{consumed} > \eta \cdot V_{total} \quad (2.3.3)$$

где I_{POP} множество популяций; $\eta \leq 1$ коэффициент, отражающий этот эффект «перенаселения».

Определим $\tilde{V}_{consumed}$ как

$$\tilde{V}_{consumed} = \eta \cdot V_{total} / \sum_{i \in I_{POP}} P_i \quad (2.3.4)$$

Эта величина будет описывать скорректированный потреблённый объём. Таким образом, $\tilde{V}_{consumed}$ используется для учёта влияния высокой плотности популяции на снижение потребления нутриентов микробными клетками.

2.3.2. Моделирование синтеза питательных веществ

Процессы синтеза специфических питательных веществ (нутриентов) описываются следующим уравнением, которое рассчитывается для каждого специфического питательного вещества, производимого каждой популяцией в окружающей среде:

$$\Delta S = \left(\sum_{i \in A} av_i \cdot ac_i \right) \cdot cprod \cdot P$$

где ΔS – количество вещества S , синтезируемого клетками популяции за итерацию; A – множество аллелей гена, контролирующего продукцию S клетками популяции; av_i – конкретное значение аллеля гена, контролирующего производство S клетками популяции; ac_i – концентрация аллеля av_i в популяции (доля носителей этого аллеля среди всех клеток популяции); $cprod$ – «норма синтеза» - характерное количество синтезируемых молекул на клетку за одно поколение; P – размер популяции (количество клеток).

2.4. Моделирование генетического разнообразия популяции

В оригинальной методике моделирования ГЭК генетическое разнообразие клетки описывалось с помощью так называемых *генетических спектров* (Лашин, 2010; Лашин et al., 2008). Генетический спектр – это распределение частот аллелей конкретного гена в популяции. Совокупный геном микробной популяции таким образом представляет из себя вектор генетических спектров. Такие генетические процессы, как мутации, горизонтальные перенос и потеря генов описываются с помощью специально разработанной *арифметики генетических спектров*, детально описанной в (Лашин, 2010). Данная арифметика позволяла вычислительно эффективно моделировать эволюционные процессы в генетически гетерогенных микробных сообществах реалистичной численности, что было весьма актуально 10-15 лет назад.

Описание генетического разнообразия с помощью арифметики генетических спектров ГЭК имеет ряд ограничений и особенностей. В ряде случаев это приводит к тому, что возможны ситуации, когда часть клеток микробной популяции, которые вносят незначительный вклад в изменение её численности, могут получить необоснованный прирост за счёт наиболее эффективных клеток, обладающих более приспособленным генотипом, т.е.

численность наименее эффективных особей поддерживается искусственным путем. Такое поведение системы не всегда является вычислительным артефактом (ср., например, с известным статистическим парадоксом Симпсона (Blyth, 1972)).

Однако, в ряде случаев имеет место систематическая ошибка моделирования, связанная с накоплениями погрешности из-за арифметики генетических спектров. В предельном случае это может привести к ситуации, когда в изменении численности могут даже участвовать уже вымершие особи. Например, пусть в первом поколении распределение аллелей для некоторых двух генов А и В описываются соответствующими генетическими спектрами.

a1 (0.5) a2 (0.5) - распределение аллелей для гена А

b1 (0.5) b2 (0.5) - распределение аллелей для гена В

Пусть пары (a_i, b_i) наиболее эффективно утилизируют субстрат из окружающей среды в заданной трофической стратегии. Первое изменение численности популяции даст следующие результаты (группа аллелей; частота встречаемости до изменения численности/после изменения численности):

Гр.1 (a_1, b_1) ; 0.25 / (0.3)

Гр.2 (a_1, b_2) ; 0.25 / (0.2)

Гр.3 (a_2, b_1) ; 0.25 / (0.2)

Гр.4 (a_2, b_2) ; 0.25 / (0.3)

Но, несмотря на то, что особей 1 и 4 группы больше, при возвращении к исходному представлению мы снова получим концентрацию 0.5 для каждого вида аллелей, так как вклад в каждый аллель каждого гена вносят разные сочетания пар особей с количеством 0.2 и 0.3. В результате этого на ранних стадиях моделирования нарушается точность итогового результата, и, со временем, данный эффект приводит к нарушению реального генетического разнообразия системы (в данном примере группы (a_1, b_2) и (b_1, a_2) должны

достигнуть вымирания, но из-за специфики модели представления генетического разнообразия они всегда будут иметь 25% всех особей популяции).

Для исключения подобных ситуаций в рамках данной работы была реализована новая модель описания генетического разнообразия популяции. В новой модели мы перешли к хранению всех аллельных комбинаций, встречающихся в популяции, в явном виде. Их частоты хранятся в специальном списке, что позволяет избавиться от генетических спектров и постоянного построения и сохранения аллельных комбинаций. Таким образом, вымершие аллельных комбинаций аллельные комбинации больше не оказывают влияния на дальнейшую судьбу моделируемого сообщества. Отметим, что новая модель является существенно более требовательной к оперативной памяти и не рекомендована для использования в задачах очень высокого генетического разнообразия (при рассмотрении популяций с числом аллельных комбинаций, превышающем 10^8).

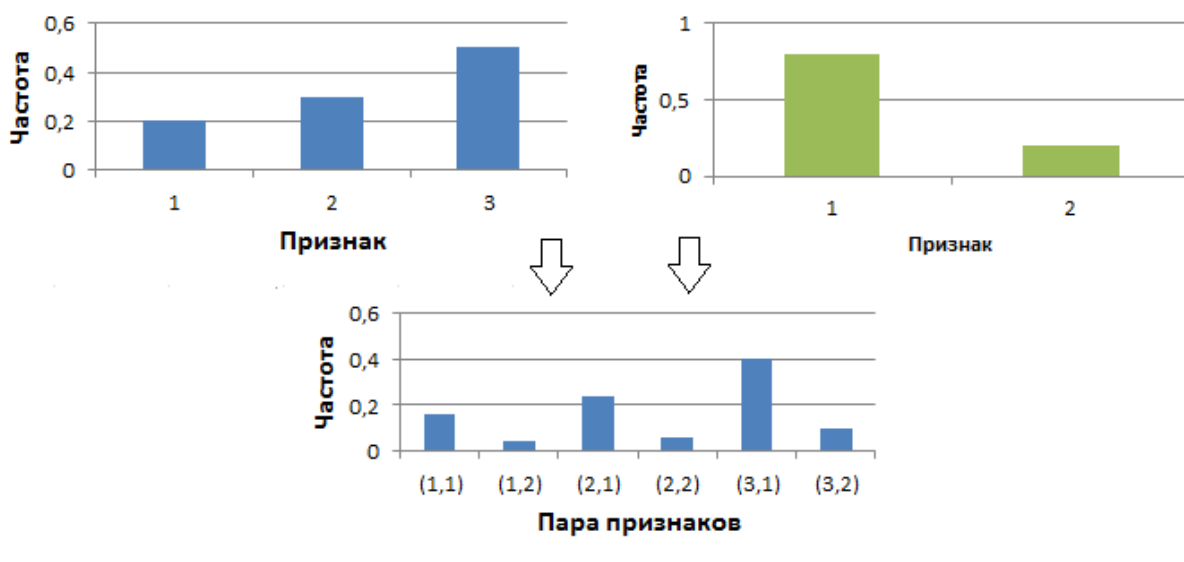


Рисунок 2.4.1. С помощью генетических спектров образуются всевозможные аллельные комбинации.

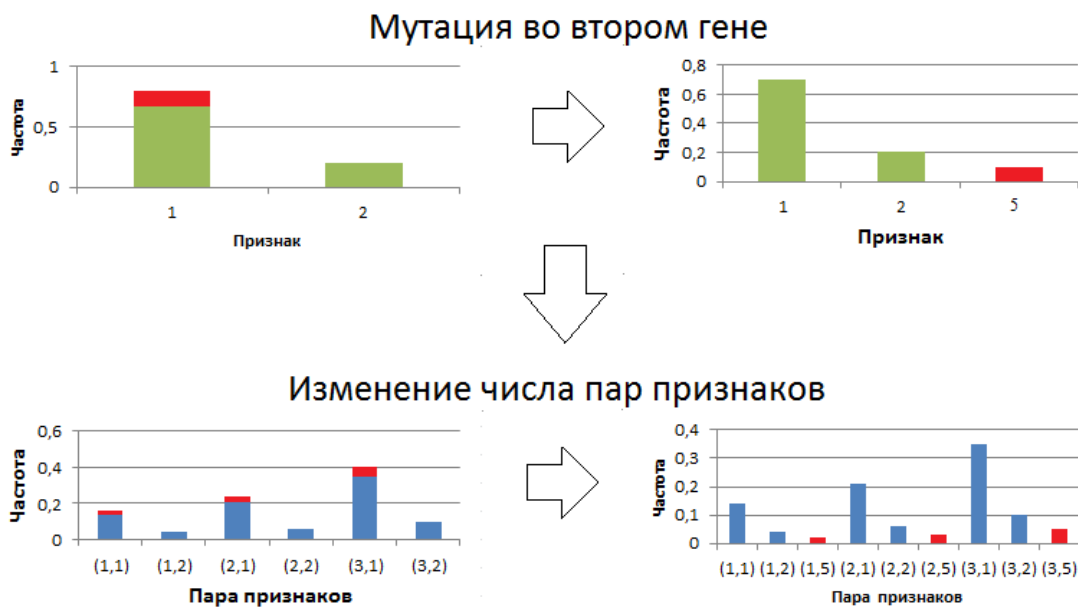


Рисунок 2.4.2. Мутация во втором гене приводит к появлению нового аллельного варианта, и, как следствие, число аллельных комбинаций увеличивается в 1.5 раза.

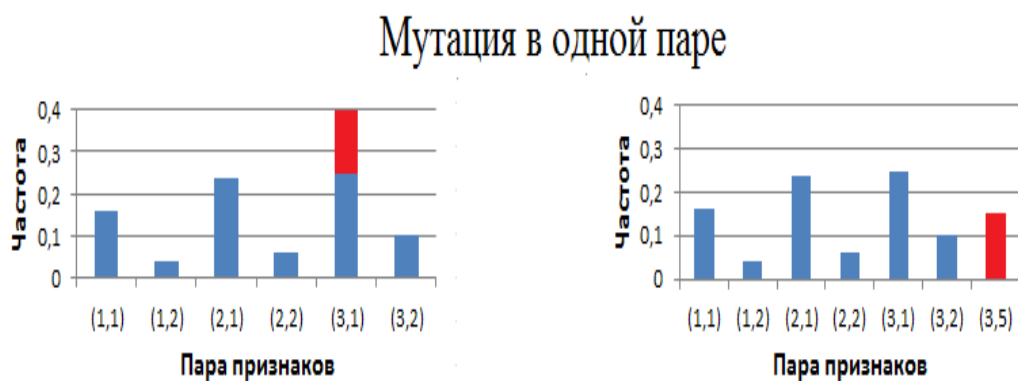


Рисунок 2.4.3. В новом алгоритме моделирования мутаций за одну итерацию мутирует одна аллельная комбинация.

Для описанной модели был реализован новый алгоритм моделирования мутаций. Исходный алгоритм в ГЭК моделирует мутацию в генетическом спектре, тем самым, затрагивая сразу множество аллельных комбинаций (рис. 2.4.1-2.4.2). Новый алгоритм моделирования мутаций осуществляет мутацию

в одной, указанной или случайной, аллельной комбинации (рис. 2.4.3). Данный алгоритм работает в обеих моделях описания генетического разнообразия.

Недостатком новой модели представления генетического разнообразия популяции являются более высокие требования к оперативной памяти компьютера. Они могут достигать нескольких гигабайт при моделировании сообществ высокого генетического разнообразия.

Таблица 2.4.1. Зависимость требуемой оперативной памяти от числа аллельных комбинаций в модели представления генетического разнообразия с непосредственным хранением аллельных комбинаций.

Число генов	Число аллельных вариантов гена	Число мономорфных популяций	Затраты памяти на параллельный процесс	Время выполнения
6	10 для всех генов	10^6	28 Мб	5 минут
12	Половина генов 5, половина 2	10^6	52 Мб	5 минут
7	10 для всех генов	10^7	320 Мб	48 минут
14	Половина генов 5, половина 2	10^7	600 Мб	48 минут
8	10 для всех генов	10^8	3.6 Гб	8 часов
16	Половина генов 5, половина 2	10^8	6.8 Гб	8 часов

Были выставлены приоритеты использования моделей описания генетического разнообразия:

1) Для расчёта моделей с большим количеством аллельных комбинаций ($>10^8$) лучше всего использовать модель с хранением генетического спектра популяции, так как данная модель не потребует больших затрат к оперативной памяти и имеет большой выбор параллельных реализаций.

2) Для расчёта моделей низкого и среднего уровня генетического разнообразия ($<10^8$ аллельных комбинаций) и моделей с высокими требованиями к точности результата лучше использовать модель непосредственного хранения всех аллельных комбинаций.

2.5. Моделирование фаговой инфекции

Для описания фаговой инфекции в методику ГЭК были добавлены дополнительные объекты полиморфных популяций: *полиморфные популяции фагов* и *полиморфные популяции инфицированных* клеток (таким образом «обычные» полиморфные популяции считаются «здоровыми»). Моделирование инфекционного процесса включает следующие фазы:

- заражение здоровых клеток путем проникновения фагов из окружающей среды в некую часть популяции клеток;
- размножение фагов внутри инфицированных клеток;
- выход фагов в окружающую среду после лизиса клеток.

Инфицированные клетки образуют полиморфные популяции, дальнейшее размножение которых может идти по литическому или лизогенному пути.

Литический путь означает гибель (лизис) инфицированных клеток с одновременным образованием фагов и их переносом в окружающую среду.

Лизогенный путь предполагает образование профагов и отсутствие транспорта фагов в окружающую среду. При этом популяция инфицированных клеток размножается как обычная полиморфная популяция в ГЭК (т.е. в соответствии с трофической стратегией), приобретая иммунитет к данному типу фага (см. рис. 2.5.1).

Выбор лизогенного или литического сценария зависит от условий, в которых в момент заражения находятся клетки популяции (рис. 2.5.2). В условиях положительной динамики численности популяции, обеспеченной высокими концентрациями субстратов в среде и/или улучшенной по сравнению с другими генетической приспособленностью (данные параметры можно специфицировать в модели), инфицирование популяции проходит по литическому типу: клетки погибают с образованием фагов, число которых зависит от копияности фага. В менее оптимальных условиях инфицирование

идёт уже по лизогенному типу: в обобщённый геном популяции добавляются гены фага, т.е. образуется профаг. Дальнейшее размножение такой инфицированной популяции совпадает с таким у обычной полиморфной популяции, при этом она приобретает иммунитет к данному типу фага. Впоследствии, при улучшении условий возможен случайный переход части популяции в литическую форму, с последующей гибелью этой части и генерацией фагов.

Полиморфизм фагов, появление новых штаммов вследствие мутаций, и конкуренцию между штаммами также описываются с помощью арифметики генетических спектров. При этом в популяциях фагов, в отличие от прокариотических популяций, гены определяют специфическую вирулентность (возможность заражать те или иные популяции) и копияность (количество копий на лизируемую клетку).

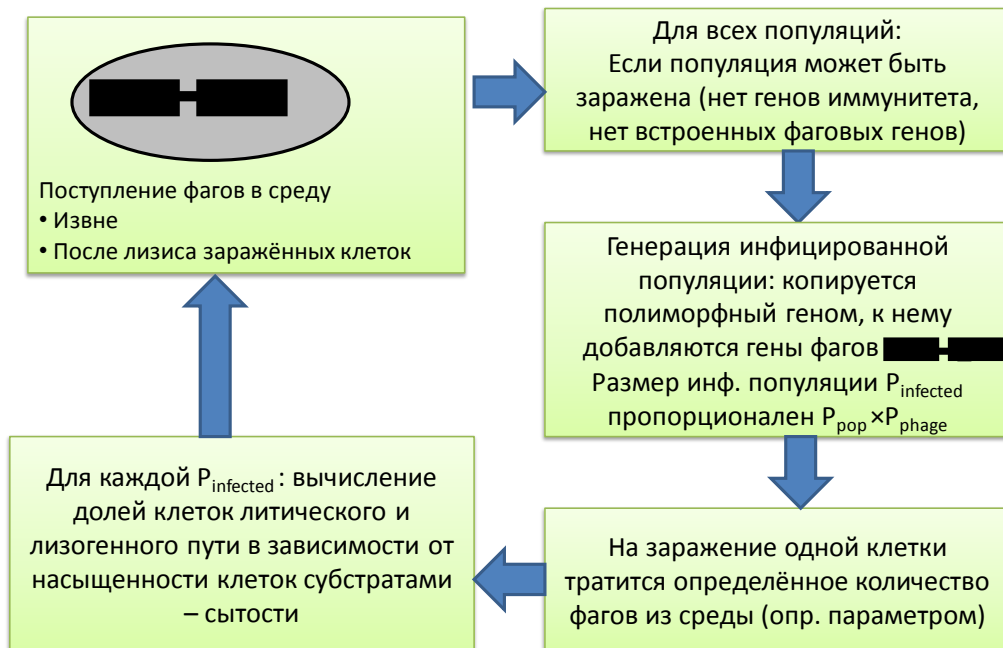


Рисунок 2.5.1. Моделирование заражения клеток микробного сообщества фаговой инфекцией. Принципиальная схема моделирования в ГЭК.

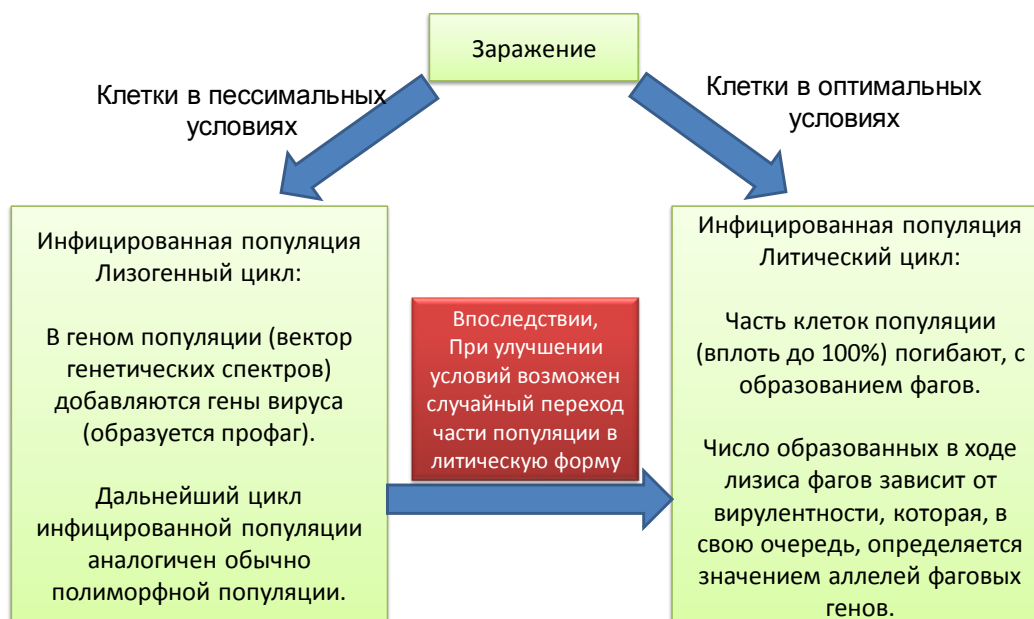


Рисунок 2.5.2. Выбор литического и лизогенного сценариев при моделировании заражения клеток микробного сообщества фаговой инфекцией в ГЭК.

2.6. Программный комплекс «Гаплоидный эволюционный конструктор»

Программный комплекс «Гаплоидный эволюционный конструктор» (ГЭК) был разработан в качестве реализации рассмотренной в первой части данной главы одноимённой методики моделирования. В течение более чем десяти лет в рамках ГЭК были созданы: вычислительные модули в виде консольных приложений, графический интерфейс пользователя, веб-ориентированное приложение, средства обработки результатов моделирования.

2.6.1. Общая архитектура ПК ГЭК

В настоящее время ГЭК содержит следующие основные модули:

- Ядро 0D (основные молекулярно-генетические, клеточные и популяционные процессы в среде с равномерным перемешиванием). Реализовано на C++ с использованием библиотеки Boost.
- Ядро 3D (процессы, связанные с пространственной

гетерогенностью, транспортом веществ и клеток и т.п.). Реализовано на C++ с использованием библиотеки Boost.

- Модули OpenMP, MPI, Cuda, оптимизирующие Ядро 0D в части распараллеливания процессов моделирования генетического разнообразия. Реализованы на C++ с использованием соответствующих API.
- Модули OpenMP, MPI, оптимизирующие Ядро 3D в части распараллеливания процессов моделирования пространственной гетерогенности. Реализованы на C++ с использованием соответствующих API.
- Графический интерфейс пользователя, обеспечивающий полный контроль над процессом моделирования (включая возможность пошагового изменения параметров модели). Реализован на C++ на платформе QT.
- Модуль импорта моделей молекулярно-генетических систем из существующих средств моделирования и баз данных (Biomodels, CellML, SabioRK, MGSMODELSDB). Реализован на C++ с использованием библиотек libSBML, sosLib.
- Web ГЭК 3D, позволяющий создавать модели, ставить их в очередь на выполнение на высокопроизводительных кластерах СО РАН и НГУ и анализировать результаты расчётов. Содержит также библиотеку предустановленных моделей. Реализован на языках JavaScript в рамках библиотеки webix (клиентская часть) и Python (модули dispy, paramiko, Eve – планировщик задач и связь с высокопроизводительными кластерами).

На рисунке 2.6.1 приведена общая схема взаимодействия модулей программного комплекса ГЭК.

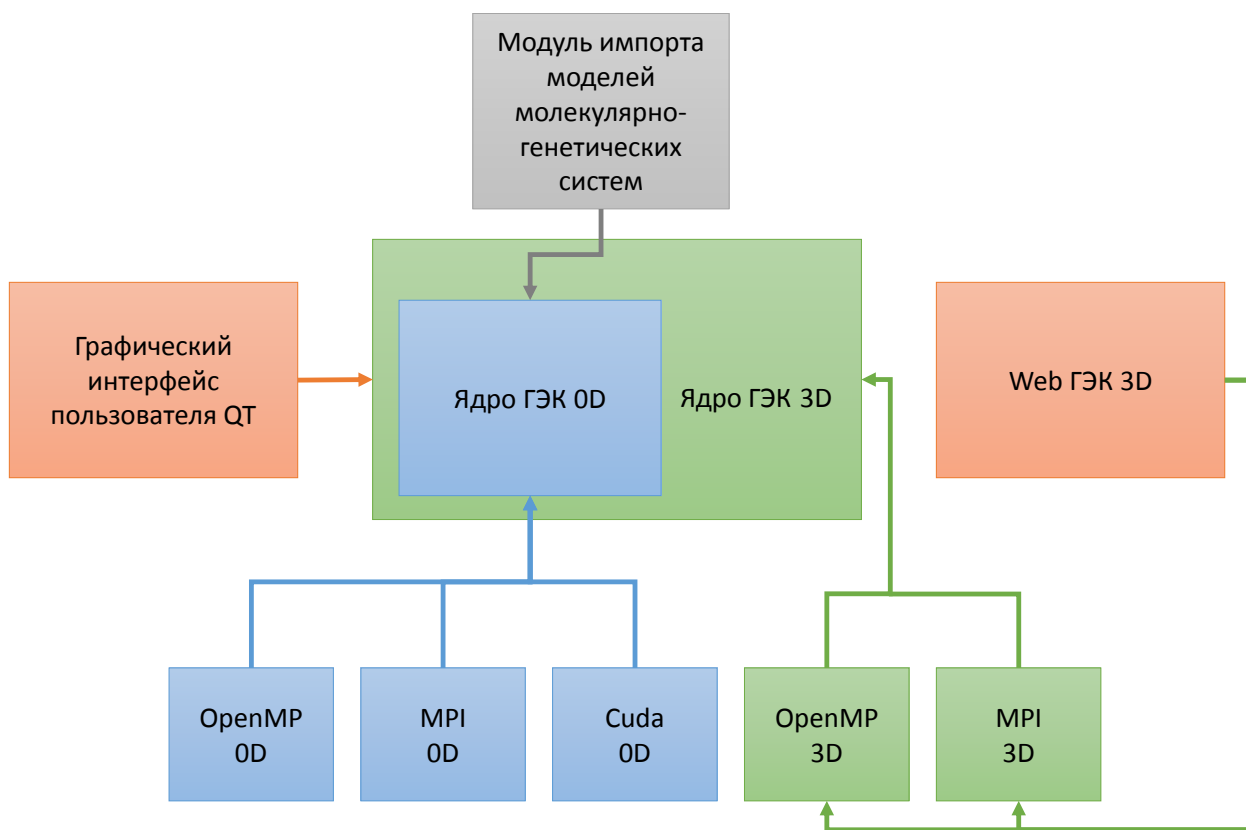


Рисунок 2.6.1. Общая схема взаимодействия модулей ПК ГЭК.

2.6.2. Организация потока вычислений ПК ГЭК

Процесс моделирования в ГЭК разбивается на отдельные итерации – т.н. модельные «тики». При моделировании микробного сообщества в среде с полным перемешиванием – точечной среде (0D-случай) – одна итерация соответствует одному циклу размножения микроорганизмов (характерное время – 30 минут, см. также подробное обоснование это величины в (Лашин, 2010)). При добавлении пространственного распределения – случаи 1D, 2D и 3D – в ходе тех же итераций необходимо учитывать процессы транспорта и миграции.

При изменении количества ячеек рассматриваемой системы, даже при сохранении объёма, можно наблюдать, что при большем количестве ячеек продвижение клеток и веществ происходит медленнее, что соответствует средам с большей вязкостью. Для того, чтобы избежать этого эффекта, в ходе реализации модулей для учёта пространственной гетерогенности среды обитания микробного сообщества, нами был введён параметр учёта *вязкости*,

который показывает сколько раз происходит перераспределение и стандартная итерация без изменения численности популяций на один вызов обработчика размножения. Таким образом, чтобы снизить вязкость среды при сохранении числа ячеек, нужно обсчитывать процессы миграции и транспорта в несколько раз чаще, чем размножение.

Поток вычислений в ПК ГЭК таким образом организован так, как показано на рисунке 2.6.2. Случай точечной среды показан на рисунке 2.6.2а, а модифицированная схема для моделирования пространственно гетерогенных сред – на рис. 2.6.2б. При этом сначала для каждой из N ячеек рассчитывается стандартная итерация (стадии, обведённые зелёной рамкой на Рис. 2.6.2б), а потом отдельным шагом применяется перераспределение веществ и клеток. Процедура размножения клеток моделируется раз в AV (относительная вязкость среды) шагов.

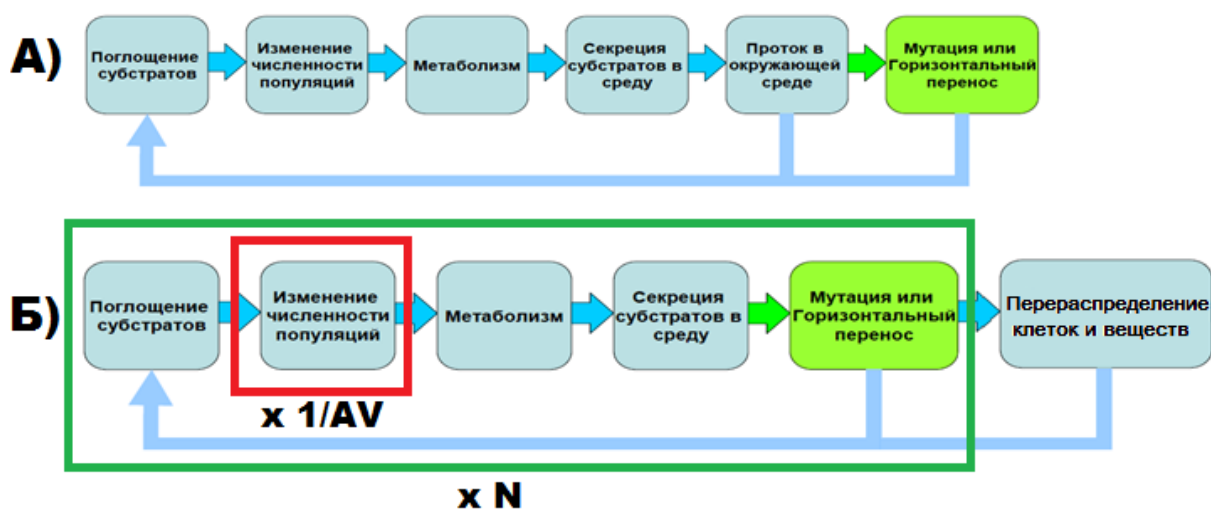


Рисунок 2.6.2. Схема потока вычисления в ПК ГЭК. А) вариант для сред с полным перемешиванием (случай 0D); Б) вариант для пространственно гетерогенных сред (случаи 1D, 2D, 3D). N – количество ячеек, AV – параметр учёта вязкости среды.

2.6.3. Модули высокопроизводительных вычислений с параллелизацией по генетическому разнообразию сообщества

Модель микробного сообщества, реализованная в ПК ГЭК, вообще говоря, подразумевает два направления, по которым может возрастать

вычислительная сложность моделирования – это генетическое разнообразие сообщества (т.е. количество уникальных аллельных комбинаций в сообществе) и пространственная гетерогенность сообщества (т.е. количество пространственных ячеек, каждая из которых характеризуется своим собственным уникальным состоянием). В ходе работы нами были разработаны модули высокопроизводительных вычислений ГЭК для обоих этих случаев. Параллелизация по генетическому разнообразию сообщества описана в данном разделе, по пространственной структуре сообщества – в следующем (см. раздел 2.6.4). Отметим также, что были реализованы высокопроизводительные модули как для расчёта как на суперкомпьютерных кластерах, так и на обычных настольных компьютерах.

Перед тем, как перейти к рассмотрению результатов работы параллельных версий программы, рассмотрим формулы, которые были использованы для подсчета таких характеристик, как *коэффициент ускорения* и *коэффициент эффективности*:

Для расчета коэффициента ускорения использовалась следующая формула:

$$S = T_1 / T_p \quad (2.6.1)$$

T_1 – время выполнения последовательной версии программы,

T_p – время выполнения параллельной версии программы.

Для расчета коэффициента эффективности использовалась следующая формула:

$$E = S / p * 100\% \quad (2.6.2)$$

S – коэффициент ускорения,

p – количество используемых потоков.

Для единообразия мы ввели также единое определение *вычислительной сложности модели ГЭК*:

$$Compl_{HEC} = \sum_{node \in Node_{Env}} (GenDiv_{node}) \quad (2.6.3)$$

где $Node_{Env}$ – множество всех ячеек окружающей среды в модели, а $GenDiv_{Node}$ – генетическое разнообразие сообщества в данной ячейке, определяемое по формуле:

$$GenDiv = \sum_{P_i} AC(P_i) \quad (2.6.4)$$

где, в свою очередь, AC – это количество всех аллельных комбинаций популяции P_i .

Перейдём к описанию высокопроизводительных модулей с параллелизацией по генетическому разнообразию сообщества. В начале работы нами был проведен анализ существующей версии ГЭК при помощи профилировщика Intel Parallel Amplifier (<http://software.intel.com/ru-ru/articles/intel-parallel-studio-home>), и было показано, что при моделировании сообществ с высоким генетическим разнообразием (10^6 - 10^8 уникальных аллельных комбинаций) практически все время выполнения уходит на вычисление единственной функции – функции изменения численности популяции, независимо от типа трофической стратегии (рис. 2.6.3).



Рисунок 2.6.3. Время выполнения функции изменения численности популяции относительно времени работы программы.

На рисунке 2.6.3 показано, что с ростом количества аллельных комбинаций время выполнения программы практически полностью концентрируется на функции изменения численности популяции. Идея параллелизации алгоритма, реализующего изменение численности

популяции, заключается в переборе всех возможных аллельных комбинаций (допустим, всего их n) в одном цикле длины n . Комбинации записываются с помощью специального массива, состоящего из индексов каждого аллельного варианта комбинации. В качестве примера рассмотрим таблицу 2.6.3.

Все аллельные варианты всех генов записываются в один массив (полужирным текстом выделены текущие аллельные варианты каждого гена). С помощью массива индексов (в стартовой комбинации он имеет вид (0, 3, 4, 8)) выписывается первая комбинация, и индекс текущего аллельного варианта в последнем гене увеличивается на 1. Если при этом в последнем гене получен последний аллельный вариант, то значение выбранного аллельного варианта в предыдущем гене увеличивается на 1, а в последнем гене выбирается первый аллельный вариант. Таким образом, с помощью массива индексов осуществляется полный перебор всех аллельных комбинаций в популяции.

Достоинство такого подхода состоит в том, что полученный цикл может быть разбит на любое число параллельных процессов (в работе было протестировано до 900 процессов). Каждому процессу необходимо «знать» массивы значений и концентраций аллельных вариантов (т.н. «развёртки» значений и концентраций), а также стартовую и конечную позиции своего фрагмента цикла. По окончании вычислений в каждом процессе полученные данные суммируются в корневой процесс с помощью функции массового суммирования `MPI_Reduce`, происходит присваивание посчитанных результатов исходному объекту и алгоритм завершает работу (рис. 2.6.4). Все пересылки реализуются с помощью функций `MPI_Bcast` и `MPI_Reduce`, все процессы выполняют приблизительно одинаковый объем работы (несущественные отличия возникают, если число комбинаций не кратно числу процессов).

Таблица 2.6.1. Возможные аллельные комбинации

Аллельные варианты всех генов	Полученные комбинации
(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)	(0, 3, 4, 8)
(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)	(0, 3, 4, 9)
(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)	(0, 3, 5, 8)
...	...
(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)	(2, 3, 7, 9)

Нами было проведено тестирование алгоритма на 6-ти ядерных процессорах X5670 2.93 GHz (Westmere) кластера НКС 30-Т (<http://bioinformatics.bionet.nsc.ru/>). Алгоритм был верифицирован на тестовом наборе сценариев ГЭК (Lashin et al., 2012a; Лашин et al., 2009). Затем были составлены специальные нагрузочные тесты (*GenDiv* от 10^6 до 10^8 аллельных комбинаций). Результаты тестирования алгоритма приведены в таблице 2.6.2.

В ряде вычислительных экспериментов было показано, что эффективность распараллеливания иногда превышает значение 1. Это объясняется тем, что на различных узлах кластера могут быть получены различные значения времени выполнения программы. Расчет параллельной версии на узлах, отличных от узлов, на которых рассчитывалась последовательная версия, может дать увеличение или уменьшение эффективности. Среднеквадратичное отклонение для каждой выборки результатов показано в столбце «дисперсия», с округлением вверх до секунд. Столбец «ускорение» показывает эффективность распараллеливания для различного числа процессов. Получено практически линейное ускорение на протяжении всего тестирования.

Таким образом, время выполнения программы на современных процессорах сократилось с 8 часов до 2 минут, при этом требуя всего 12 узлов

вычислительного комплекса и на каждый параллельный процесс необходимо всего 2 Мб оперативной памяти.

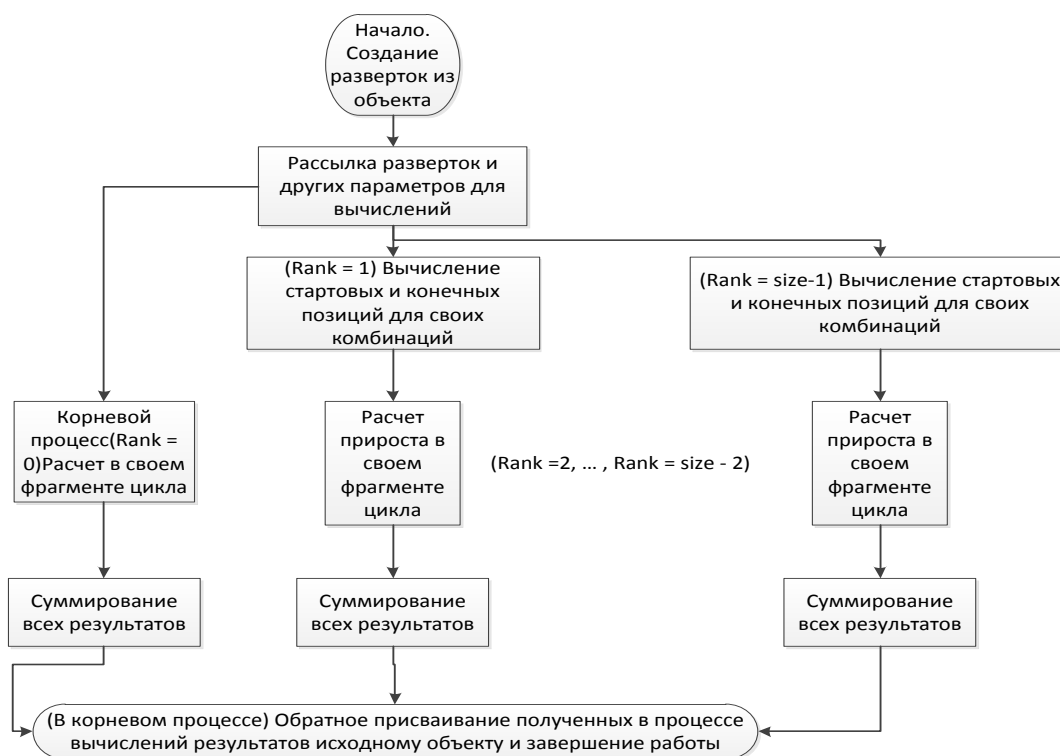


Рисунок 2.6.4. Схема распараллеливания алгоритма.

Таблица 2.6.2. Результаты тестирования на X5670.

Количество процессов	Время работы чч:мм:сс	Эффективность распараллеливания	Ускорение	Дисперсия, сек
1	8:02:26	1	1	249
2	4:10:24	0.9633	1,9266	65
4	2:05:20	0.9623	3,8492	188
8	1:02:03	0.9718	7,7744	4
16	0:31:06	0.9695	15,512	3
24	0:20:48	0.9664	23,1936	3
36	0:13:46	0.9734	35,0424	1
64	0:07:52	0.9582	61,3248	2
96	0:05:13	0.9633	92,4768	1
144	0:03:32	0.9481	136,5264	1
264	0:02:03	0.8914	235,3296	2

Таким образом, чем сложнее структура моделируемого сообщества, чем больше в этом сообществе генетическое разнообразие, тем большую долю в выполнении программы занимает выполнение функции расчёта изменения численности популяции (см. рис. 2.6.3), соответственно, на сложных моделях выигрыш от оптимизации является максимальным. Было показано, что ускорение, близкое к линейному, достигается при моделировании сообществ с большим генетическим разнообразием. Именно такие сообщества и представляют наибольший интерес для исследования, а представленные в данном разделе модули позволяют пользователем увеличивать сложность и разнообразие моделируемых биологических ситуаций.

2.6.4. Модули высокопроизводительных вычислений с параллелизацией по пространственной структуре среды обитания сообщества

В ходе работы нами также были разработаны модули высокопроизводительных вычислений ГЭК с учётом пространственной структуры сообщества, для расчёта как на обычных настольных компьютерах, так и на суперкомпьютерных кластерах. Поддержка высокопроизводительных вычислений представляет собой систему распределения вычислительных ресурсов (вычислительных модулей), встроенную в пространственно-распределённую версию ПК ГЭК.

На первом этапе этого блока работ был проведен анализ времени выполнения последовательной версии ГЭК, который показал, что время расчетов в программе растет при переходе от одной пространственной топологии к другой (рис. 2.6.5). То есть, переходя от 1D версии к 2D и от 2D – к 3D, при постоянном количестве ячеек (добавляется лишь новая пространственная компонента: так, например, было взято 500 ячеек в одномерном случае, 50x10 ячеек в двумерном и 5x10x10 - в трехмерном), время работы увеличивается.

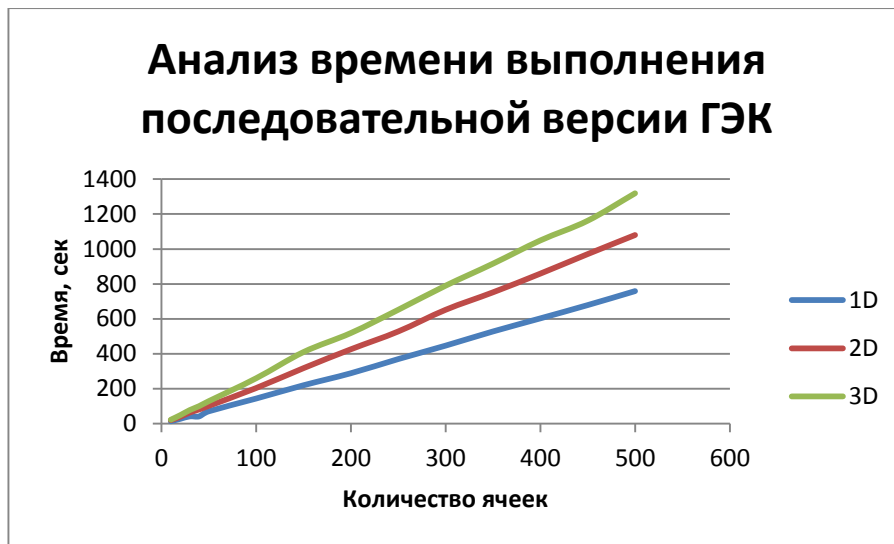


Рисунок 2.6.5. Время выполнения последовательной программы ГЭК.

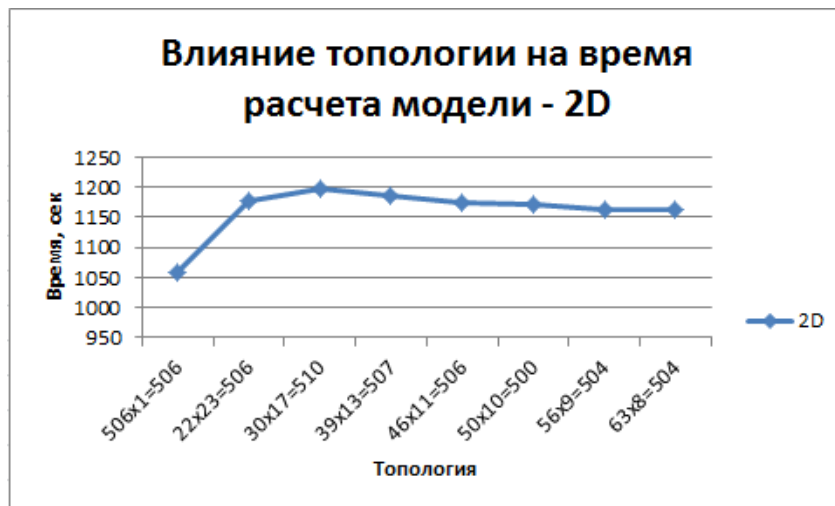


Рисунок 2.6.6. Влияние топологии на время расчета 2D модели.

Анализируя замеры времени работы программы в рамках 2D модели (рис. 2.6.6), можно видеть резкий скачок при переходе от распределения ячеек 506x1, к распределению 22x23, который говорит о влиянии топологии на время расчета внутри конкретной модели. При этом дальнейшие преобразования (30x17, 39x13, 50x10 и т.д.) оставляют время примерно на одном уровне.

В случае с 3D моделью все аналогично. Наблюдались два больших скачка во времени работы программы – при переходе от 500x1x1 к 22x23x1 и при переходе от 63x8x1 к 8x8x8 (рис. 2.6.7).



Рисунок 2.6.7. Влияние топологии на время расчета 3D модели

Полученные выше результаты можно записать в общем виде. Для начала, на примере 2D модели. Пусть есть некоторое число $A = B \times C$. При распределении вида $A \times 1$ времени на расчеты тратится меньше, чем при распределении вида $B \times C$. Это происходит, так как при появлении дополнительных компонент в пространстве возникают дополнительные сложности с расчетами перераспределения клеток и веществ, а также других процессов, которые протекают в пространстве. При этом между всеми видами распределений $B \times C$ (какие бы B и C не были взяты при сохранении общего числа ячеек) существенной разницы во времени не наблюдается.

В случае с 3D моделью все аналогично. Пусть $A = B \times C$; $B = D \times E$. Время работы при распределении вида $A \times 1 \times 1$ меньше времени работы при распределении типа $B \times C \times 1$. Время работы $B \times C \times 1$ меньше, чем время при $D \times E \times C$.

Подводя итог анализа влияния топологии на время работы ГЭК 3D, можно сказать, что время работы программы ГЭК 3D растет лишь с добавлением новой пространственной компоненты, при этом не имеет значения, сколько именно ячеек было добавлено по этой новой компоненте (существенных различий во времени выполнения не наблюдается).

Общая схема работы параллельной версии ГЭК 3D с использованием технологии QtConcurrent

С учетом общей схемы работы и результатов анализа времени выполнения последовательной версии программы ГЭК была предложена и реализована схема работы параллельной версии программы с использованием QtConcurrent (рис. 2.6.8). При старте программы, как и в последовательной версии, идет считывание скрипта. После чего начинается создание потоков с учетом свободных ресурсов компьютера (количество свободных ядер) и каждый созданный поток приступает к инициализации своей области решетки. Далее начинается этап расчетов модели, в который входят расчет стандартной итерации и расчет перераспределения популяций и субстратов. Все эти расчеты также проводятся с использованием потоков, созданных с помощью QtConcurrent, пока не будет достигнуто желаемое количество итераций.

Когда программа завершается, начинается вывод результатов. Вывод, как и расчеты, выполнен в параллельном режиме. Так как многие посчитанные данные выводятся в отдельные файлы для каждой из ячеек, то параллельная реализация представляется возможной. Тот вывод, в котором необходимо вывести общую сводную информацию о системе в целом выполняются в последовательном режиме.

Для тестирования параллельной версии программы, реализованной с использованием QtConcurrent, использовалось следующее оборудование: Настольный компьютер: AMD Phenom™ II X6 1055T (6 ядер), 2800 МГц, 12Гбайт RAM. Результаты работы этой версии программы представлены в таблице ниже (табл. 2.6.3).

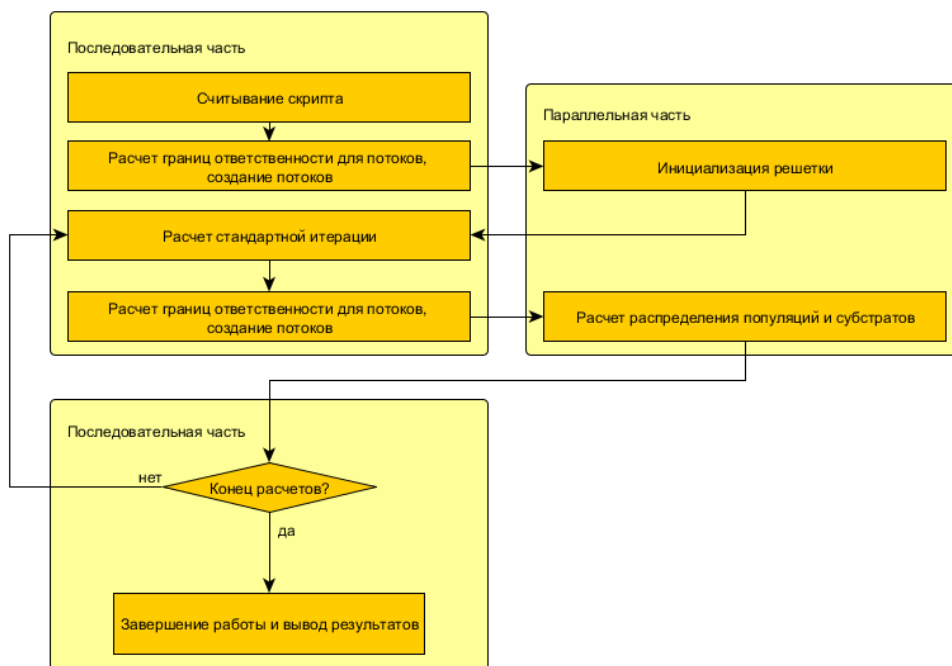


Рисунок 2.6.8. Схема работы параллельной версии ГЭК с использованием QtConcurrent.

Таблица 2.6.3. Время выполнения последовательной и параллельной версии программы с использованием QtConcurrent на моделях с топологией 1D (1000 ячеек), 2D (100x100 ячеек) и 3D (10x10x10 ячеек).

Модель	Инициализация данных (с)	Перераспределение + обновление / итерацию (с)	Время работы программы (с)
1D (1000 ячеек) последовательная	15.9	0.6	74
1D (1000 ячеек) QtConcurrent	13	0.4	53
2D (100x100 ячеек) последовательная	438	4.12	505
2D (100x100 ячеек) QtConcurrent	260	2.5	287

3D (10x10x10 ячеек) последовательная	12.2	0.83	97
3D (10x10x10 ячеек) QtConcurrent	7.9	0.57	68

Проанализировав результаты, представленные в таблице, можно отметить, что с увеличением сложности модель лучше поддается оптимизации посредством параллельных вычислений. Так, в модели 2D (100x100 ячеек) значительно было сокращено время, необходимое для инициализации данных, сокращение которого не столь ощутимо при расчете других моделей и слабее всего проявляется в случае с 1D решеткой.

Средний же показатель ускорения на этапе с перераспределением популяций и субстратов с последующим обновлением данных у моделей держится примерно на одном уровне в 1.5 раза (с небольшими отклонениями в зависимости от количества ячеек и сложности задачи).

Общая схема работы параллельной версии ГЭК с использованием технологии MPI

Общую схему работы последовательной версии ГЭК можно представить в виде блок-схемы (рис. 2.6.9А). В начале работы происходит считывание скрипта с данными о моделируемой среде и необходимом количестве итераций в переменную типа `std::string` (функция `processFileScript`) (в рамках программного комплекса ГЭК число итераций – это число поколений). После чего идет инициализация среды в виде решётки с учетом прочитанных из скрипта данных (функция `processScript`). Все описанные выше операции выполняются в рамках класса `ECProcessor3D`.

Затем в функции `directProcess` класса `ECProcessor3D` идет расчёт стандартной итерации, расчёт распределения популяций и субстратов

(функция `redistribute` класса `Lattice3D`). После чего выполняется обновление данных в решётке при помощи функции `updateCells` класса `Lattice3D`. Затем начинается новая итерация. По завершению расчётов производится вывод результатов в текстовые файлы, и работа программы завершается.

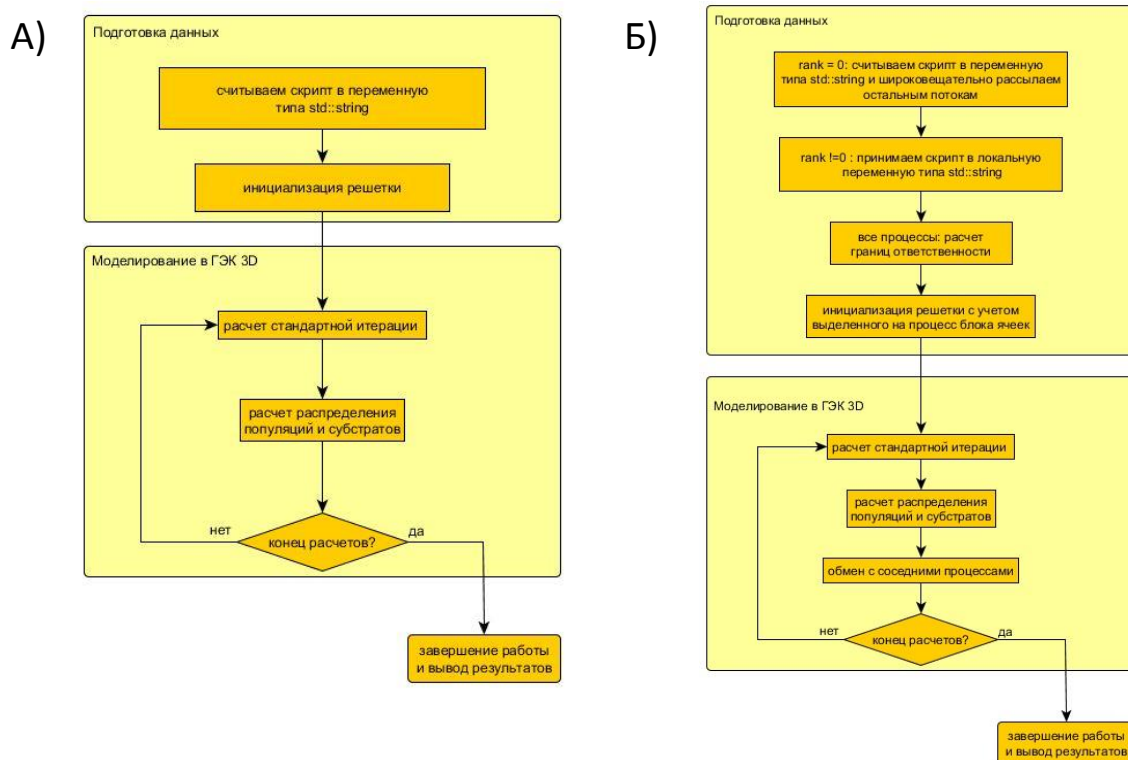


Рисунок 2.6.9. А) Блок-схема работы последовательной версии ГЭК; Б) Блок-схема работы параллельной версии ГЭК.

Перейдём к рассмотрению организации параллельных вычислений в MPI модуле. В каждом процессе должна проходить инициализация своего блока ячеек, необходимая для расчётов в данном процессе. Перед обновлением информации после рассчитанных распределений и стандартной итерации необходимо провести обмен данными между соседними друг для друга процессами. После чего продолжить выполнение программы с новой итерации. (Подробная схема работы параллельной версии ГЭК 3D представлена на блок-схеме рисунка 2.6.9Б).

МРІ версия программного комплекса ГЭК начинает своё выполнение с обработки входных параметров командной строки. Для корректной работы на вход необходимо передать желаемое количество потоков и их распределение по пространству, т.е. при самом общем случае решётки с размером $X \times Y \times Z$ пользователю необходимо будет указать, какое количество потоков следует выделить на каждую из пространственных компонент: X , Y и Z .

Для хранения всех необходимых каждому потоку данных была создана специальная структура, содержащая следующую информацию:

- общее количество потоков (поле `int size`)
- ранг текущего потока (`int rank`)
- границы ответственности текущего потока (`std::vector<int> BoundaryValues + Bounds for each thread` 6 полей)
- заданное пользователем разбиение по каждому из измерений (поля `int widthSize, int heightSize, int lengthSize`)

После инициализации начинаются расчёты модели. В процессе моделирования каждая ячейка решётки рассматривается независимо, но на каждой новой итерации потокам необходимо знать, какие результаты были получены у их «соседей». Для обмена информацией между потоками была организована система обмена новыми полученными данными. Она была реализована с использованием метода `void exchangeThreads`:

```
class Lattice3D {
...
private:
    envVector3D entity;        //length x width(depth) x height
    cellBufVector3D buffer;
    int dimension;
    int bordersCount;
    size_t length;
    size_t width;
    size_t height;
    size_t current[3]; //0 == i, 1 == j, 2 == k
```

```

void exchangeThreads(thread_param threads);
...
}

```

Есть заранее заданные константы, предназначенные для обозначения тегов коммуникаций (запрос на получение информации от соседа, например, LEFT_REQUEST, RIGHT_REQUEST; отправка ответа на полученный запрос, например, LEFT_ANSWER, RIGHT_ANSWER). После обработки запроса на получение информации поток формирует необходимые данные и отправляет их, после чего запрашивает данные для себя. После получения ответа данные внутри потока обновляются, и начинается новая итерация расчёта модели. По окончании расчётов каждый из потоков записывает полученные им данные в выходные файлы и программа завершается.

Общая схема ПК ГЭК по пространству приведена на рисунке 2.6.10.

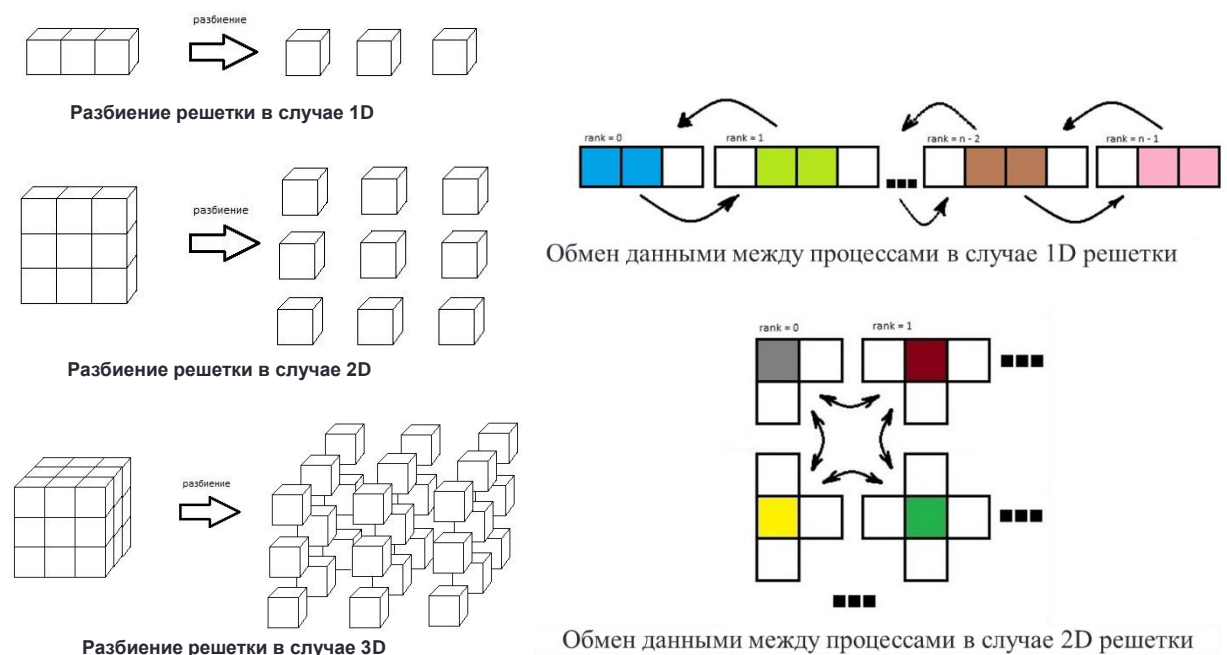


Рисунок 2.6.10. Параллелизация ПК ГЭК по пространству.

Исследования вычислительной эффективности параллельных версий ГЭК велись для задач разной сложности – рассматривались как очень простые модели, так и сложные «нагрузочные». На сложной задаче была достигнута эффективность до 180% с использованием небольшого числа процессов на MPI. Для неё же достигнуто ускорение до 40 раз с использованием большого

числа процессов MPI. С использованием QtConcurrent на обычном настольном компьютере достигнуто ускорение до 3 раз при эффективности распаралеливания до 50%.

В процессе исследований стало понятно, что на задачах с расчётом простых моделей намного эффективнее использовать SMP машину, по сравнению с вычислительным кластером с распределенной памятью. На задачах с расчётом сложных моделей стоит использовать высокопроизводительный кластер с распределенной памятью при наличии не более 50 процессов. Если есть возможность использования 50 процессов на SMP машине, то это будет намного эффективнее по сравнению с высокопроизводительным кластером с распределенной памятью.

Таблица 2.6.4. Результаты работы высокопроизводительных модулей ПК ГЭК.

	ПРОСТАЯ МОДЕЛЬ								
	1D			2D			3D		
	Количество ячеек	Количество популяций	Суммарная сложность	Количество ячеек	Количество популяций	Суммарная сложность	Количество ячеек	Количество популяций	Суммарная сложность
	1000	6	18000 - MPI 9000 - Qt	1000	2	5400 - MPI 9000 - Qt	1000	5	15000 - MPI 9000 - Qt
	эффективность, %		ускорение	эффективность, %		ускорение	эффективность, %		ускорение
MPI 10 потоков	95		9,5	31		3,1	7,3		0,73
MPI 50 потоков	73		36,5	9,2		4,6	2,8		1,4
SMP 10 потоков	110		11	54		5,4	10,6		1,06
SMP 50 потоков	116,8		58,4	21		10,5	6,12		3,06
QT 6 потоков	29,2		1,8	22,5		1,4	22,3		1,3

	СЛОЖНАЯ МОДЕЛЬ								
	1D			2D			3D		
	Количество ячеек	Количество популяций	Суммарная сложность	Количество ячеек	Количество популяций	Суммарная сложность	Количество ячеек	Количество популяций	Суммарная сложность
	10000 - MPI 1000 - Qt	56	33600000 - MPI 3360000 - Qt	10000 - MPI 1000 - Qt	56	33600000 - MPI 3360000 - Qt	10000 - MPI 1000 - Qt	56	33600000 - MPI 3360000 - Qt
	эффективность, %		ускорение	эффективность, %		ускорение	эффективность, %		ускорение
MPI 4 потока	180		7,2	145		5,8	120		4,8
MPI 10 потоков	144		14,4	108		10,8	80		8
MPI 50 потоков	48		24	38,4		19,2	30,8		15,4
SMP 4 потока	75		3	60		2,4	47,5		1,9
SMP 10 потоков	66		6,6	60		6	44		4,4
SMP 50 потоков	75,4		37,7	50,2		25,1	17,6		8,8
QT 6 потоков	46,7		2,8	45,0		2,7	43,3		2,6

Сводные результаты по эффективности и ускорению различных высокопроизводительных модулей ГЭК с параллелизацией по пространству приведены в таблице 2.6.4.

2.6.5. Модули интеграции ПК ГЭК с существующими средствами моделирования генных сетей и базами данных

Внутриклеточный метаболизм, моделируемый ГЭК, как было указано ранее, описывается в рамках парадигмы генных сетей. При моделировании, мы представляем микробные сообщества в виде сети, узлы (популяции) которой содержат генные сети. Эти сети регулируют процессы синтеза и потребления метаболитов, а также размножения бактерий.

В свою очередь, процесс синтеза и потребления субстратов обусловлен генетическим набором. Такой механизм описывается одной или несколькими генными сетями. В рамках ГЭК, такие генные сети называются *генетическими стратегиями* (согласно паттерну проектирования «Стратегия», предложенному Александреску (Александреску, 2019)). Мы выделяем два типа стратегий:

- *Стратегия синтеза (или метаболическая стратегия)* – принимает на вход концентрации внутренних субстратов, в процессе функционирования высчитывает новые концентрации, которые и являются результатом синтеза.
- *Трофическая стратегия* – принимает на вход внутриклеточные концентрации веществ и считает по ним рост или убыль популяции клеток.

В зависимости от того, сколько субстрата было произведено, будет изменяться численность популяции, в большую или меньшую сторону. Генетические параметры популяции обуславливают здесь скорости производства и потребления субстратов (см. подробнее раздел 2.1).

Формальное представление параметров популяции выглядит следующим образом (рис. 2.6.11):

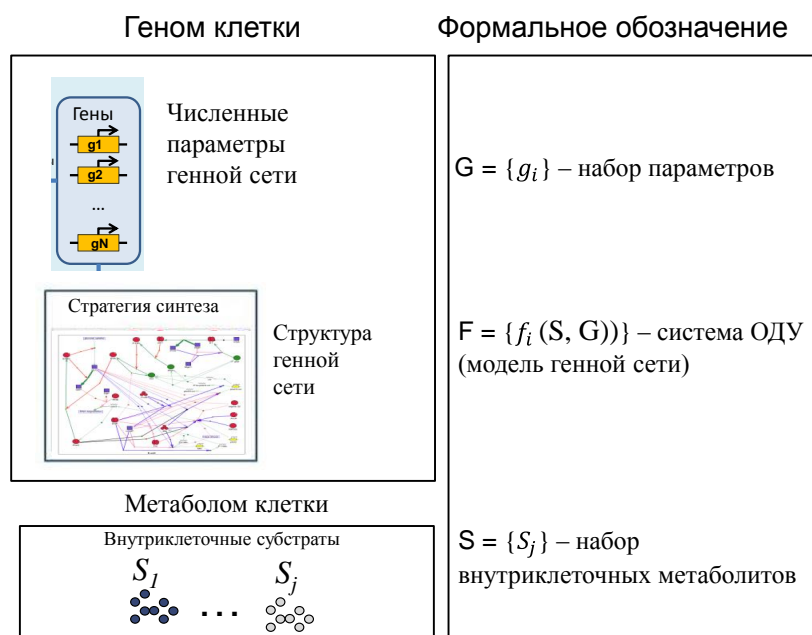


Рисунок 2.6.11. Параметры популяции (гены, генная сеть, внутриклеточные субстраты), используемые при моделировании.

Геном популяции представляется наборами векторов численных параметров генной сети. Стратегия синтеза – системой обыкновенных дифференциальных уравнений (или моделью МГС, записанной в другом формализме, для которой можно генерировать временные ряды, соответствующие изменениям концентраций метаболитов в системе с течением времени). И набор внутриклеточных субстратов представляется аналогично генам, в виде набора векторов численных значений концентраций субстратов.

В разделах 2.1-2.4 были описаны некоторые типовые стратегии, используемые в ГЭК. Архитектура ГЭК, описанная в разделе 2.6.1, позволяет создавать новые стратегии в виде подключаемых программных модулей, т.н. плагинов, в достаточно быстрые сроки, но требует для этого довольно квалифицированных программистов и математических биологов. В то же время научным сообществом были созданы библиотеки готовых моделей, а также базы данных, предоставляющие доступ к этим моделям. Примерами

таких баз являются: Biomodels (Malik-Sheriff et al., 2019), Sabio-RK (Wittig et al., 2012), МАММОТh (Kazantsev et al., 2018) и другие. Универсальный формат представления моделей МГС в этих базах данных – язык SBML (systems biology markup language – язык разметки системной биологии) (Hucka et al., 2015, 2006).

В рамках работы был создан модуль ПК ГЭГ, который позволяет загружать готовые модели в формате SBML из курируемых биологических баз данных, при этом данные модели представляются в виде стратегий синтеза ПК ГЭК. В рамках этого модуля процедура расчёта стандартной итерации для каждой ячейки была модифицирована согласно схемам на рисунках 2.6.12 и 2.6.14.

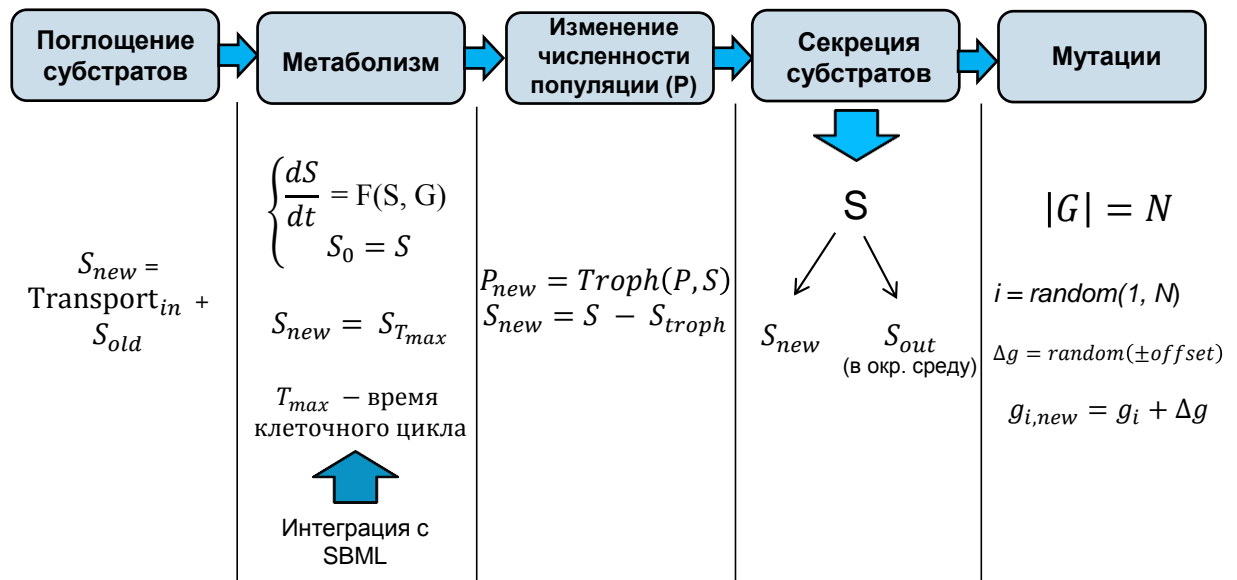


Рисунок 2.6.12. Схема стандартной итерации ГЭК с подробным описанием основных стадий и указанием точки интеграции с внешними моделями, записанными в формате SBML.

Как видно из рис. 2.6.12, на этапе метаболизма происходит выработка новых субстратов, учитывая генетические параметры и начальные значения субстратов. С точки зрения моделирования, происходит генерация временных рядов от T_0 до T_{max} , символизирующих значения концентраций субстратов, на заданном интервале времени.

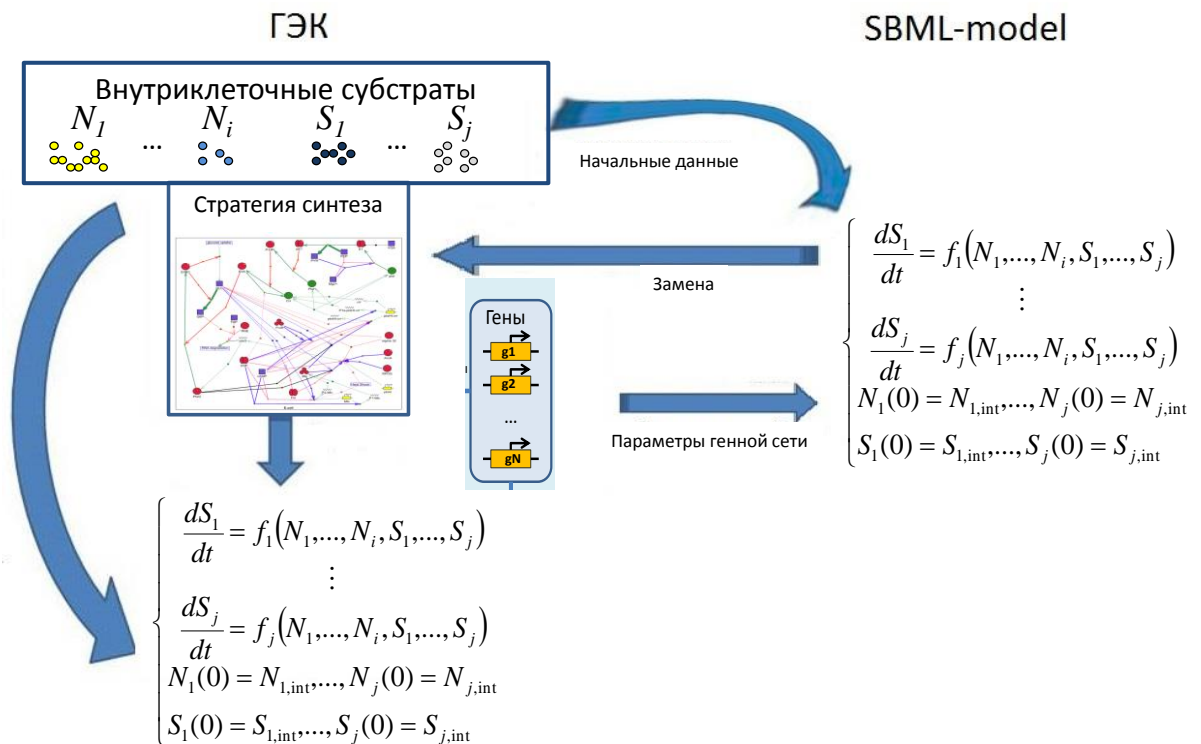


Рисунок 2.6.13. Общая схема интеграции ГЭК и SBML.

Общая схема интеграции SBML-модели в ГЭК выглядит следующим образом. Для выбранной популяции с помощью графического интерфейса пользователя или с помощью специально разработанного языка (см. Приложение 1) устанавливается желаемая SBML-модель с указанием того, какие субстраты будет производить и потреблять популяция (начальные значения концентрации субстратов), а также какие параметры реакций следует использовать (значения генов). Устанавливая производимые и потребляемые субстраты, пользователь ГЭК определяет, с какими популяциями будет взаимодействовать создаваемая популяция, поскольку другие популяции могут производить или потреблять установленный субстрат. В результате будет сконфигурирована новая популяция, использующая стратегию синтеза из загруженной SBML-модели.

Генетические параметры встроенной модели выступают в качестве параметров для SBML-модели, а концентрации субстратов – в качестве начальных данных для дифференциальных уравнений (рис. 2.6.13, 2.6.14).

Подключение модуля интеграции SBML-моделей в ГЭК осуществляется путём замены стандартного этапа метаболизма общего цикла моделирования на внешний этап, реализованный в представленном модуле (рис. 2.6.14). В процессе итерации новая стратегия использует значения субстратов, наработанные с предыдущей итерации, и вычисляет новые значения, которые, в свою очередь, используются при вычислении новых значений численности популяций. Отметим, что числовые значения генетических параметров расширенной модели используются в качестве параметров для SBML-решателя, а концентрации субстратов – в качестве начальных данных (рис. 2.6.14). Все параметры и начальные данные представляются в виде векторов числовых значений.

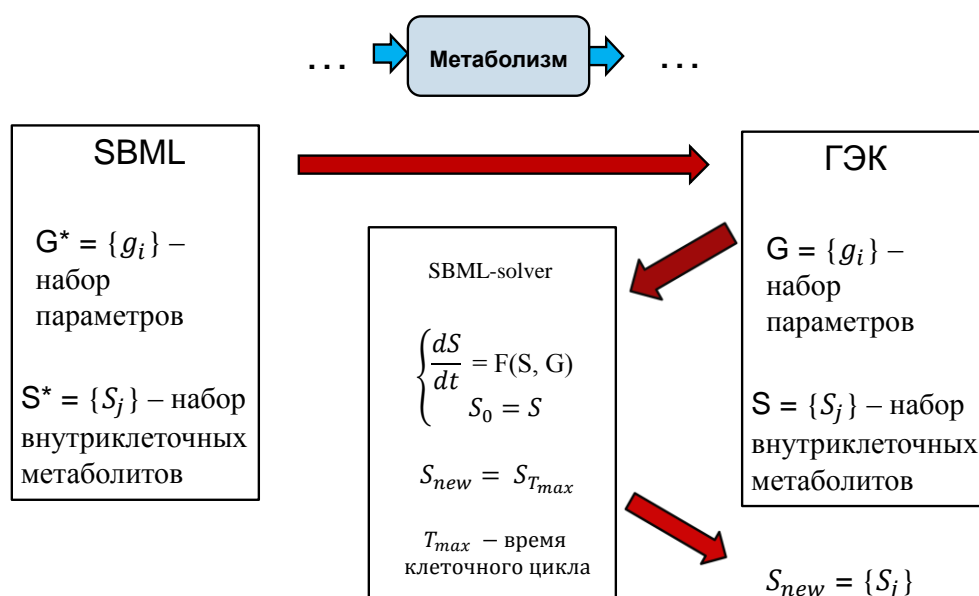


Рисунок 2.6.14. Алгоритм функционирования модуля интеграции.

Архитектура программного модуля интеграции SBML-моделей в ПК ГЭК приведена на рисунке 2.6.15. Общий алгоритм выглядит следующим образом. На вход модулю подаётся файл SBML (он может быть скачан пользователем вручную, либо с помощью программного API). С помощью библиотеки libSBML (Bornstein et al., 2008) из файла извлекаются параметры, которые визуализируются с помощью графического интерфейса, и система дифференциальных уравнений. После того, как пользователь установил все необходимые ему параметры, с помощью библиотеки SOSlib

(<https://www.tbi.univie.ac.at/software/odeSolver/>) происходит расчет системы дифференциальных уравнений, согласно установленным параметрам. Результат работы устанавливается в текущую модель ГЭК и используется при последующих итерациях.3D (с помощью библиотеки libSBML), численный расчёт задач Коши для этих моделей осуществляется с помощью библиотеки SoSLib. Для устранения конфликтов имён при наименовании метаболитов в разных базах данных, было принято решение воспользоваться базой данных наименований химических веществ ChEBI (Hastings et al., 2013), для чего был реализован сервисный модуль интеграции в ГЭК.

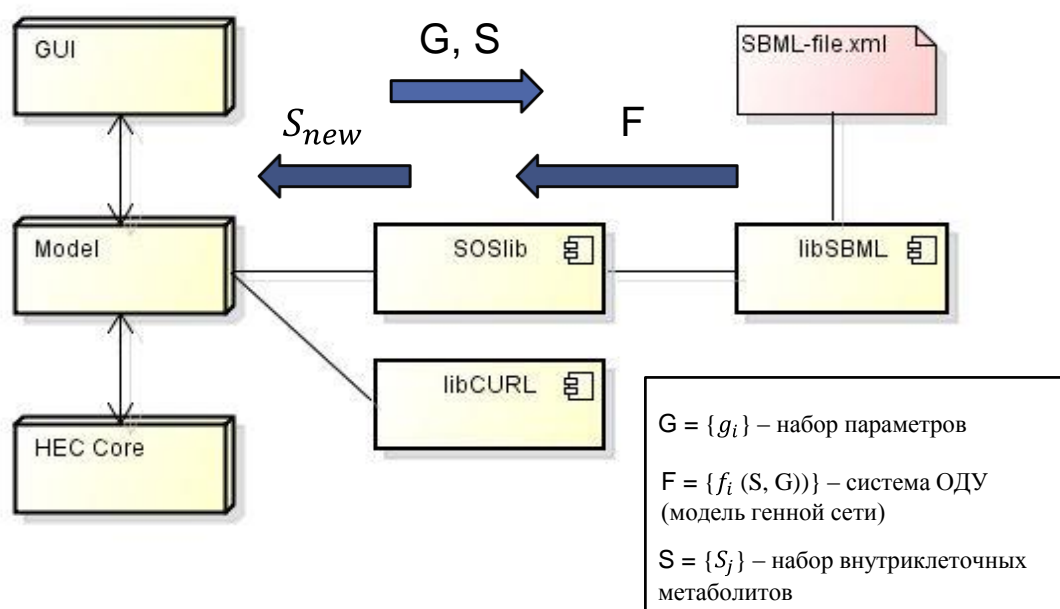


Рисунок 2.6.15. Архитектура модуля интеграции потока данных (пояснения в тексте).

Интеграция с базами данных моделей молекулярно-генетических систем выполнена согласно нижеприведённой схеме:

- 1) Экстракция и подготовка моделей:
 - А. Экстракция моделей производится через предоставляемые программные инструменты доступа:
 - а. МАММОТН: <http://mammoth.biomodelsgroup.ru/api/>

b. Sabio-rk: <http://sabio.h-its.org/layouts/content/docuRESTfulWeb/manual.gsp>

c. Biomodels.net: <https://www.ebi.ac.uk/biomodels-main/webservices>

Модели извлекаются по критерию вхождения в них соответствующих метаболитов.

В. Извлеченные модели разбиваются на серию элементарных подсистем $M_i(S_j, G^*, P)$. Здесь M_i – это i -ая модель элементарной подсистемы описывающей преобразование набора внутриклеточных метаболитов S_j под воздействием серии регуляторных молекул (ферменты, продукты синтеза с генов G) G^* . P – параметры модели. После такого разбиения становится возможным использование двух разных моделей, загруженных из разных источников, без конфликта переменных. Именно способность синтеза G^* является основными характеристиками микробной популяции. Т.е. бактерии фактически могут обмениваться элементарными подсистемами синтеза и утилизации метаболитов.

С. Генерация индексов соответствия множества $G \Rightarrow G^* \Rightarrow M_i(S_j, G^*, P)$

2) Расчет жизнедеятельности популяции:

А. Построение совокупной модели биосинтеза метаболитов (т.н. стратегии синтеза в ГЭК) в формате SBML.

В. Расчет SBML модели средствами интегратора SosLib.

Модуль, описанный в данном подразделе, позволяет интегрировать в ПК ГЭК весь багаж моделей молекулярно-генетических систем, представленных в базах данных в формате SBML – самом популярном на сегодняшний день формате представления моделей МГС.

2.6.6. Графический интерфейс пользователя ПК ГЭК и веб-приложения

За время реализации ПК ГЭК было создано несколько вариантов графического интерфейса пользователя. В частности, в виде отдельного

приложения с использованием технологий QT был реализован графический интерфейс ПК ГЭК (рис. 2.6.16).

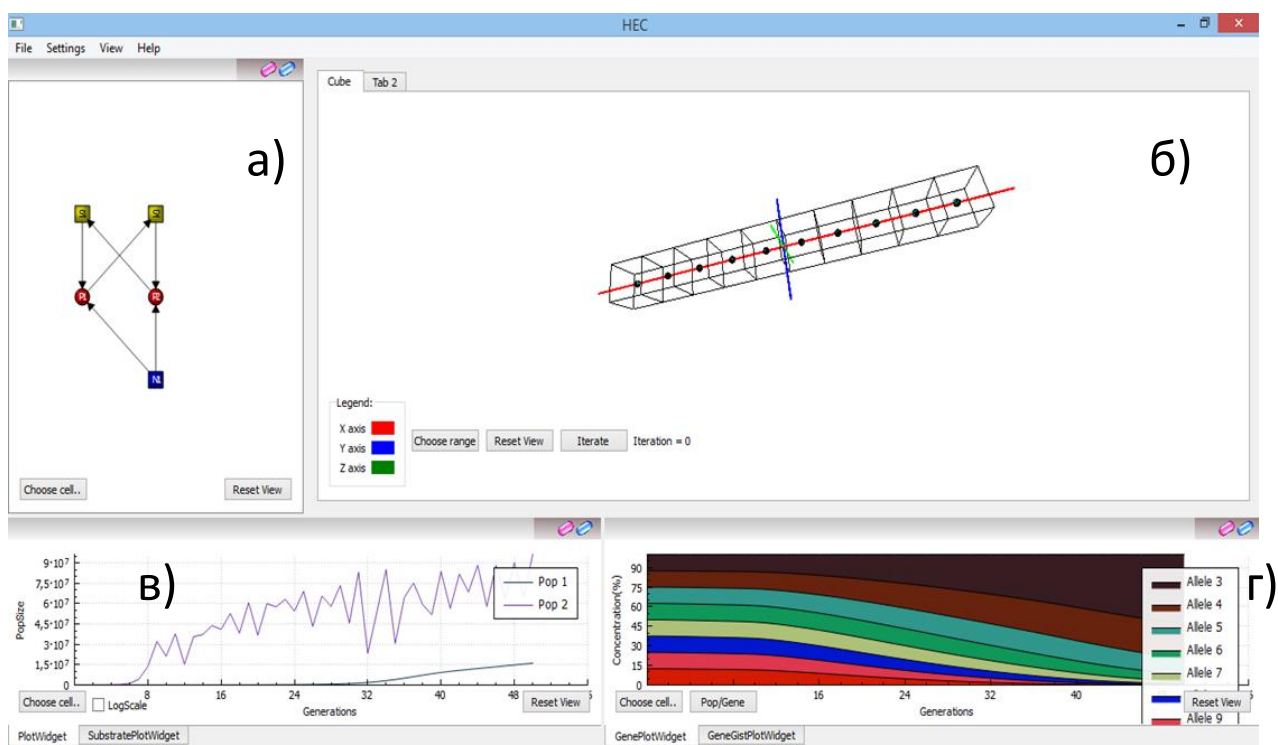


Рисунок 2.6.16. Общий вид графического интерфейса ПК ГЭК, реализованного с помощью QT: а) граф трофических отношений; б) пространственная организация 1D-среды обитания в виде вектора ячеек с населяющими их микробными сообществами; в) график динамики отдельных популяций сообщества; г) динамика генетических спектров выбранной популяции.

В данном графическом интерфейсе возможны следующие визуализации модели:

- 1) Граф трофических взаимоотношений. Виджет, визуализирующий трофическую цепь в загруженной модели. Желтыми квадратами обозначены специфические субстраты, синими квадратами – неспецифические. Красные кружки обозначают популяции. Входящие стрелки отображают потребление субстратов, исходящие – производство.

2) Динамика численности популяций. Виджет визуализирует динамику роста популяций исследуемой модели на протяжении нескольких итераций.

3) Концентрация субстратов. Виджет, предназначенный для визуализации концентрации специфических и неспецифических субстратов.

4) Динамика частот аллелей. Виджет визуализации динамики изменений доли аллелей для выбранной популяции конкретного гена по всему объему на протяжении нескольких итераций.

5) Генетическое разнообразие. В популяции может быть различное генетическое разнообразие по одному или нескольким генам. Данная информация представлена в виде гистограмм, изображающих состояние на текущей итерации. Также можно проследить происходящие изменения при включении анимации по итерациям.

6) 3D визуализация. Данный виджет визуализирует рассматриваемую модель в виде 3D объекта по заданным параметрам. Каждый куб соответствует одной ячейке, а сфера внутри – численности находящихся в ней популяций.

Из-за определённых проблем с совместимостью и корректной работой под разными операционными системами, которые регулярно возникали с разработанными графическими приложениями, мы дополнительно разработали веб-приложение WebHEC, доступное по адресу <http://webhec.sysbio.cytogen.ru>. Основные компоненты приложения показаны на схеме (рис. 2.6.17): это подсистема графического интерфейса и менеджер задач.

Система WebHEC является веб-интерфейсом для серверной версии ядра ГЭЖ (рис. 2.6.18). Её задачи – создавать запросы на исполнение, передавать их на сервер для расчета, информировать о состоянии исполнения задачи и отображать/загружать результаты расчетов. Работа выполняется по

классической схеме «модель/представление/контроллер». На уровне «представления» – веб-приложение, разработанное на языке Java Script в рамках инструментария webix (<https://webix.com/>). Данный инструментарий позволяет разрабатывать пользовательский интерфейс с богатой функциональностью по отображению и фильтрации данных. Библиотека стандартных компонент webix позволяет поместить на страницу не только обычные элементы управления (кнопки, формы ввода, отображение текста), но и сложную графику в виде разнообразных диаграмм.

«Контроллер» представлен приложением на языке Python. Общение с «представлением» происходит посредством реализованного REST API – специально подготовленных запросов в виде URL-путей. Функция «контроллера» обслуживать запросы клиентов на исполнение, распределять по рабочим узлам сервера нагрузку и отслеживать исполнение задач.

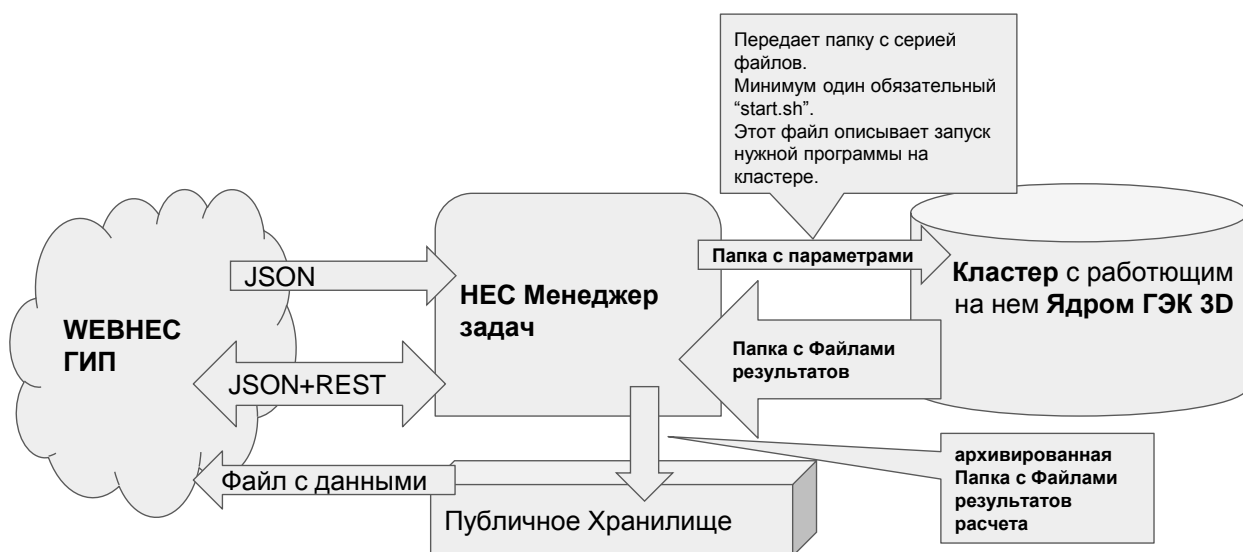


Рисунок 2.6.17. Архитектура веб-приложения WebNES, описывающая общение подсистемы графического интерфейса пользователя (ГИП) с подсистемой «Менеджер задач» посредством REST API.

На стороне «контроллера» также реализована задача по аутентификации пользователей, благодаря чему зарегистрированные пользователи не видят запросы других и быстро находят свои задачи. Поступающие запросы

обрабатываются и переправляются на исполнение высокопроизводительному вычислительному серверу.

The screenshot displays the WebHEC interface. At the top, a navigation menu includes 'About', 'New job', 'Jobs list', and 'Log in form'. The main workspace shows a model configuration for '2D_phage.through.poor_edge.lowvir'. The 'Environment' tab is active, showing parameters: Volume: 1, Dimension: 2D, Lattice: 5,5,1. Below this, a 'sub cycles' table is visible with columns for Location, Substrate, Concentration, Duration, Start_iteration, and Period.

Below the main interface, two pie charts are shown. The left chart, titled 'location:4,4,1', shows the distribution of populations (pop_0, pop_1, pop_2) with values of 40485.2 for each. The right chart shows the distribution of substrates (S_1, S_2, S_3) with values: S_1: 9.99966e-7, S_2: 0.00000100028, and S_3: 0.00000100028. Below the charts is a JSON snippet:

```
{
  "location": [
    4,
    4,
    1
  ],
  "substrates": [
    {
      "sub id": f
```

Рисунок 2.6.18. Графический интерфейс WebHEC. Сверху: Создание новой работы для модели из библиотеки моделей, редактирование параметров среды – объёма, скоростей диффузии и потока, а также топологии среды – 2D:5*5*1; Снизу: распределение долей популяций (слева) и субстратов (справа) в ячейке с координатами (4,4,1).

Планировщик задач реализован на языке программирования Python с использованием модуля `dispy` (<http://dispy.sourceforge.net/>) для Python. С помощью модуля `paramiko` (<http://www.paramiko.org/>) для Python по протоколу SSH реализована возможность постановки вычислительных задач в очередь на высокопроизводительный кластер. Общение между веб-интерфейсом пользователя и планировщиком реализовано через REST-запросы с использованием модуля `Eve` для Python (<http://python-eve.org/>).

В веб-приложение WebHEC был интегрирован ряд содержательных задач, разработанных в течение работы над ПК ГЭК. Скрипты этих моделей доступны также на головном сайте проекта ГЭК (<http://webhec.sysbio.cytogen.ru/>).

2.7. Содержательные биологические результаты, полученные с использованием ПК ГЭК

С помощью ПК ГЭК было построено большое количество содержательных биологических моделей, часть из которых была представлена ранее в диссертационных и дипломных работах, выполненных под руководством автора. В данном разделе приведены некоторые примеры моделей ГЭК и содержательная интерпретация результатов численных экспериментов, полученных с их помощью. Главное преимущество моделей ГЭК состоит в том, что с их помощью можно проследить причинно-следственные связи, возникающие на разных уровнях организации биологической системы.

2.7.1. Компьютерное моделирование эволюции микробной популяции: преодоление локальных минимумов при достижении пика на ландшафте приспособленности

Как уже было отмечено ранее, большинство фенотипических признаков живых организмов определяется аллельными комбинациями более чем одного локуса. В связи с этим для адекватного описания эффекта мутаций при моделировании эволюционной динамики необходимо учитывать такие

важные факторы, как межлокусные взаимодействия (эпистаз), экспрессию признаков и их влияние на приспособленность особей (Moore, Williams, 2015). В результате наблюдаются такие интересные эволюционные эффекты как компенсаторные нейтрализующие мутации, впервые рассмотренные Кимурой в работе (Kimura, 1985). Компьютерная модель, рассматриваемая в данной подглаве, описывает эволюцию двух генов, мутации в каждом из которых по одиночке вредоносны, но в случае если мутированы оба гена, сохраняется приспособленность на уровне дикого типа. Существенным затруднением, с которым сталкиваются модели такого типа, оказывается способность популяции преодолевать промежуточные эволюционные стадии низкой приспособленности после приобретения отдельных вредных мутаций. В более широких терминах эта проблема может быть сформулирована как проблема смещающегося баланса (анг. shifting balance) в эволюционной модели, описываемой ландшафтом приспособленности (Whitlock et al., 1995; Wright, 1984; Алтухов, 1983). С практической точки зрения такие модели оказываются полезны для направленной эволюции ферментов в сторону улучшенной стабильности и активности (Kuchner, Arnold, 1997), терапии онкологических заболеваний (Gatenby, Vincent, 2003), процессов видообразования (Turelli et al., 2001), а также для понимания механизмов устойчивости и выживаемости вирусов (Domingo, Holland, 1997). Интерес к данной проблеме постоянно растёт на фоне использования различных методов математического и компьютерного моделирования, позволяющих исследователям подробно изучать различные аспекты таких моделей (Ivan G Szendro et al., 2013). В настоящее время этот интерес также подогревается попытками с помощью моделей объяснить большие объёмы данных, собранных недавно в результате экспериментов по высокопроизводительному секвенированию геномов (Maskay, 2013; Meer et al., 2010), что, в свою очередь, позволяет пересмотреть классические и относительно новые модели генетики и эволюции популяций.

Эти эволюционные модели предназначены для ответа на следующий основной вопрос: какие факторы позволяют популяциям пересекать

«долины», соответствующие пониженной приспособленности на ландшафте приспособленности, при их переходе от одного пика к другому? Ранее было показано, что наиболее важными факторами, ответственными за успешное продвижение популяции по ландшафтными долинам, являются: размер популяции (Handel, Rozen, 2009; I. G. Szendro et al., 2013; Whitlock et al., 2003), скорость фиксации мутаций (Gokhale et al., 2009; Handel, Rozen, 2009), частота рекомбинации в популяциях, использующих половое размножение (Weissman et al., 2010), а также форма ландшафта приспособленности (Weissman et al., 2009). Большинство из приведённых выше исследований рассматривают генотип особи как набор аллелей из дискретного пространства аллелей (см., например, (Naarsma et al., 2016)).

Мы проанализировали механизмы пересечения долин ландшафта приспособленности популяцией гаплоидных микроорганизмов, приспособленность которых зависела от аллельных значений в двух локусах и определялась сложным ландшафтом, форма которого может быть описана как «окружённая рвом гора в поле» (рис. 2.7.2, в дальнейшем мы будем считать, что ров – это частный случай более общего термина «долина»). Эта форма является расширением ландшафта «Гора Фудзи» (Aita, Nusimi, 1996), использованного для аппроксимации приспособленности мутантных ферментов. На молекулярном уровне эта двухлокусная модель описывает взаимодействие двух разных субъединиц, образующих фермент, способность которого расщеплять неспецифический субстрат, являющийся источником энергии микроорганизма, определяет его репродуктивную эффективность. Ширина рва между областями с высокой приспособленностью на ландшафте достаточно велика: требуется, как минимум, несколько последовательных мутаций, приводящих к сниженной приспособленности, чтобы преодолеть его.

Мы рассмотрели влияние различных генетических и экологических факторов на эволюционную перспективу микробной популяции, состоящей из клеток, чей генотип определяется двумя взаимодействующими локусами. Эти

факторы включали в себя как молекулярно-генетическую (частота мутаций, афинность субъединиц), популяционную (функция ландшафта приспособленности) и экологическую (концентрация доступного субстрата в среде) составляющие.

С помощью ПК ГЭК была построена модель эволюционной динамики в популяции гаплоидных микроорганизмов, состоящей из клеток одного вида, населяющих резервуар с постоянным притоком неспецифического субстрата S , являющегося источником энергии для живущих в нём клеток (Рис. 2.7.1а). Приспособленность клеток определяется размером прироста популяции в единицу времени и зависит от эффективности утилизации субстрата, доступности субстрата и смертности.

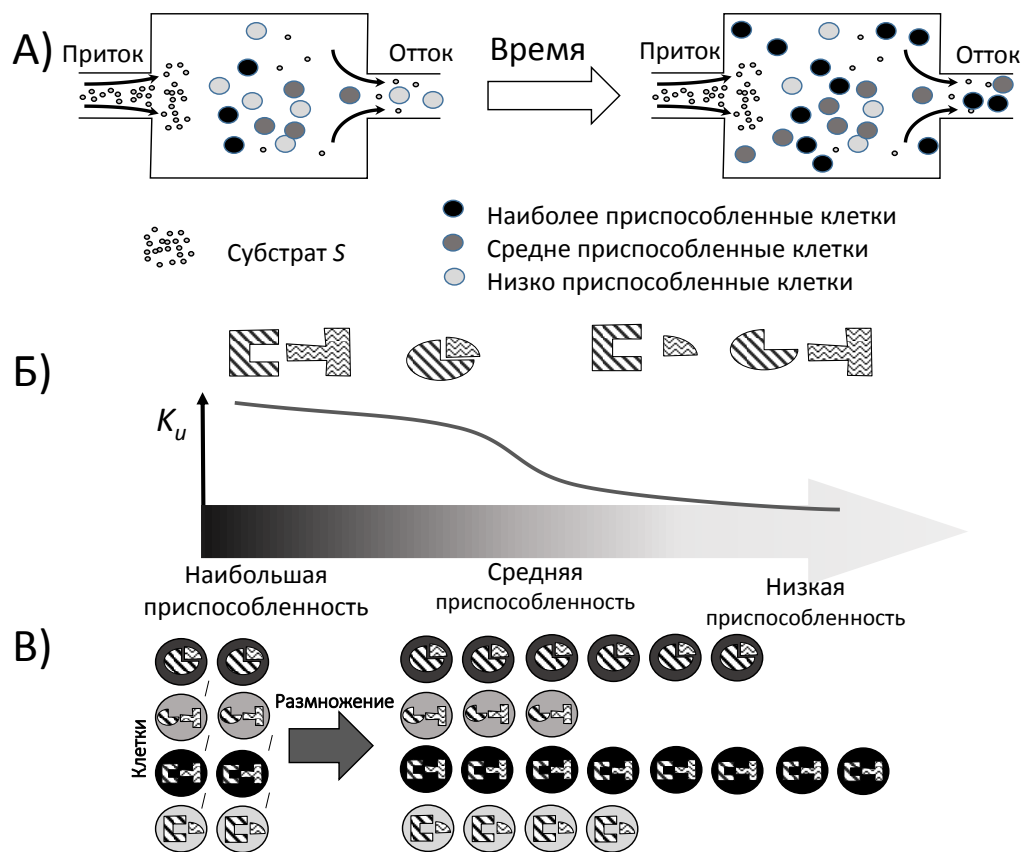


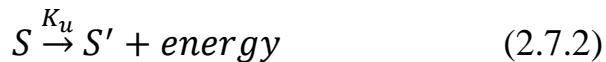
Рисунок 2.7.1. А) Схема моделируемой системы: проточный резервуар, на вход которого подаётся неспецифический субстрат – источник энергии. Резервуар населяет гетерогенная микробная популяция (клетки разных генотипов показаны разным цветом). С течением времени меняется как размер популяции, так и соотношения между представленностями клеток различных

генотипов. Б) Схема функции K_u , демонстрирующая изменение приспособленности в зависимости от комбинации аллельных вариантов. В) Более приспособленные клетки дают больше потомства за единицу времени и в долгосрочной перспективе вытесняют менее приспособленных.

Популяционная динамика (ΔP) описывается уравнением 2.7.1:

$$\Delta P = P \cdot K_u \cdot S_{uptaken} - k_{flow} \cdot P - k_{dens} \cdot P^2 \quad (2.7.1),$$

где P – это размер популяции, $S_{uptaken}$ – количество субстрата, поглощённого клетками из окружающей среды, k_{flow} и k_{dens} – коэффициенты протока и плотности соответственно. В данном уравнении лишь один член зависит от генотипа, а именно способность клеток утилизировать субстраты, добывая энергию, что определяется параметром утилизации K_u :



Эффективность утилизации субстрата K_u определяется некоторым ферментом E , состоящим из двух субъединиц E_1 и E_2 , кодируемых генами, расположенными в двух различных локусах E_1 и E_2 (см. рис. 2.7.1б, 2.7.1в). Таким образом, K_u зависит от аллельных вариантов обоих локусов; обозначим эти варианты e_1 и e_2 :

$$K_u = F(e_1, e_2) \quad (2.7.3)$$

В данной модели мы берём значения аллелей из заданных отрезков $[M_{i,min}; M_{i,max}]$ ($i = 1, 2$). На молекулярном уровне это соответствует различным значениям физико-химических характеристик аминокислотных остатков, участвующих во взаимодействиях между белковыми субъединицами (например, заряд, объём или полярность). Мы можем рассматривать генетические последовательности, кодирующие аминокислоты в тех позициях белка, которые формируют поверхность взаимодействия субъединиц, как генетические локусы и представить их вышеописанным образом. С другой стороны, эти локусы можно рассматривать и более общо – как множества позиций аминокислотных последовательностей в субъединицах фермента, которые образуют поверхность взаимодействия субъединиц. В последнем

случае, аллельное значение может соответствовать определённой интегральной характеристике аминокислотных остатков, формирующих поверхность взаимодействия, например, их совокупный заряд или объём (Afonnikov et al., 2001).

Ландшафты приспособленности

Функция F описывает зависимость эффективности утилизации субстрата от значений аллелей (ур. 2.7.3). Хотя её можно определить параметрически, в данном исследовании мы используем таблично-заданные функции для значений K_u . Мы задаём квадратную сетку соответствующих значений аллелей, e_1, e_2 , ($1 \leq e_1 \leq 9$; $1 \leq e_2 \leq 9$) с шагом сетки, равным 0.2. Общий размер сетки равен $41 \times 41 = 1681$ – числу возможных аллельных комбинаций. На данной сетке задаётся поверхность, высота которой определяется значениями из таблицы K_u , аппроксимирующей F – поверхность произвольной формы. Однако следует заметить, что мы не закладываем в нашу модель никаких предзаданных отношений между значениями K_u для каждой пары аллелей и приспособленностью организма, обладающего данным генотипом. Реальная приспособленность дополнительно обуславливается состоянием популяции (находится она в фазе роста или в стационарной фазе), её плотностью, а также доступностью субстрата и влиянием скорости протока.

В данном исследовании мы рассмотрели множество $F(e_1, e_2)$ поверхностей с пиком и долиной различной формы, соответствующих паттерну «окружённая ровом гора в поле» (см. пример на рис. 2.7.2). Эти сложные поверхности были сгенерированы как комбинация нескольких элементарных фрагментов поверхностей. Первым фрагментом выступает горизонтальная плоскость $K_u = 1$ (константный базовый уровень утилизации). Второй – это фрагмент поверхности конуса, вершина которого соответствует значению $K_u = 2$, а фокальный радиус (r_c) равен 1. Область пика соответствует комбинациям аллелей, приводящим к наиболее эффективной утилизации субстрата. Она отделена от остальной поверхности окольцовывающей

долиной. Обозначим её внутренний радиус как r_{inner} , а внешний радиус – как r_{outer} . Этот центральный пик похож на ландшафт типа «Гора Фудзи» в модели, предложенной Aita и Husimi (Aita, Husimi, 1996). Область долины («ров») соответствует пониженным значениям параметра K_u ($K_u < 1$). В данной работе мы проанализировали влияние формы ландшафта приспособленности на способность популяции достигать максимальных значений K_u . Мы не меняли положение пика приспособленности и его форму, а также внутренний радиус долины, r_{inner} . Только внешний радиус долины (параметр r_{outer}), определяющий её ширину $w = r_{outer} - r_{inner}$, и глубина долины варьировались в нашем исследовании. Мы проанализировали несколько типов поверхностей ландшафта приспособленности, табличное описание которых можно найти на рис. 2.7.2 и в таблице 2.7.1. Три примера таких поверхностей можно найти на рисунке 2.7.2.

Чтобы сгенерировать различные поверхности K_u мы использовали два значения ширины долины, $w = 1$ и $w = 2$. Кроме того, мы варьировали минимальные значения K_u ($K_{u,min}$) для долины и наклона её поверхности. Чтобы охарактеризовать соотношение между высокими и низкими значениями K_u , соответствующими пику и долине, мы посчитали среднее \bar{K}_u для узлов сетки в области между пиком и долиной (центр окружности радиуса r_{outer} совпадает с пиком). Чтобы упростить обозначение поверхностей мы ввели следующую нотацию, отражающую значения параметров, использованных для их генерации: S_A_B, где А – это значение $K_{u,min}$, а В – значение \bar{K}_u . Таким образом, мы изучили ландшафты с узкими долинами (S_0.44_0.81), широкими и неглубокими (S_0.44_0.86), а также широкими и глубокими (S_0.28_0.67) (см. таблицу 2.7.1).

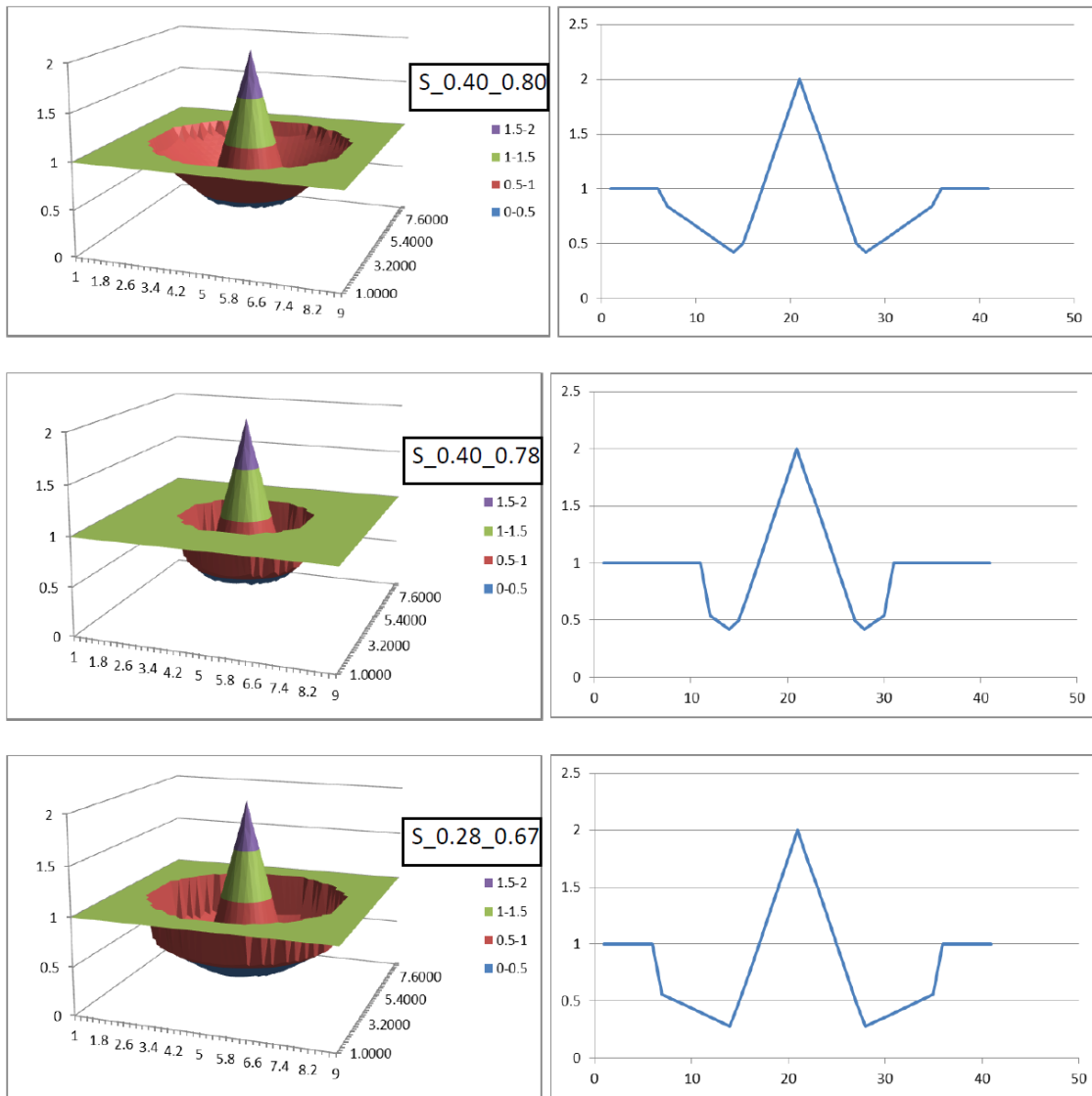


Рисунок 2.7.2. Варианты ландшафта, задаваемые функцией K_u . Слева приведены пространственные изображения ландшафтов (оси Ox , Oy соответствуют аллельным состояниям локусов E_1 и E_2 , соответственно). Справа приведены сечения плоскостью, перпендикулярной плоскости xu и параллельной оси Ox . Мнемонические обозначения ландшафтов вида S_A_B расшифровываются в таблице 2.7.1.

Таблица 2.7.1. Типы и характеристики поверхностей K_u , использованных в работе.

Тип поверхности K_u	Ширина долины, w	Максимальная глубина долины, $K_{u,min}$	\bar{K}_u	Описание
S_0.44_0.86	2	0.44	0.86	широкая мелкая долина
S_0.4_0.8	2	0.4	0.8	широкая неглубокая долина
S_0.34_0.73	2	0.34	0.73	широкая глубокая долина
S_0.28_0.67	2	0.28	0.67	широкая глубочайшая долина
S_0.44_0.81	1	0.44	0.81	узкая мелкая долина
S_0.4_0.78	1	0.4	0.78	узкая неглубокая долина
S_0.34_0.75	1	0.34	0.75	узкая глубокая долина
S_0.28_0.72	1	0.28	0.72	узкая глубочайшая долина

Изменение аллельных значений

Известно, что частоты замен аминокислот в белках распределены неравномерно: чем меньше физико-химические различия, тем более часто одна аминокислота заменяется на другую (полярная на полярную, заряженная на заряженную и т.д.) (Dayhoff et al., 1978). Мы учли этот эффект в своей модели, когда описывали изменение значений аллелей в локусах. Вероятность изменения аллельного значения описывается с помощью дискретной аппроксимации бета-распределения, учитывая четыре возможных комбинации параметров (рис. 2.7.3): в первом, втором и третьем случаях малые изменения значений аллелей более вероятны, чем существенные; в

четвёртом случае все изменения равновероятны. Затем мы дискретизируем случайные числа, полученные из этих распределений шаг за шагом согласно модели конечных сайтов (Ewens, 2004). Форма распределения определяет размер шага для изменения аллельного значения вдоль поверхности K_u . Следует отметить, что параметры были выбраны таким образом, чтобы долина могла быть преодолена лишь спустя, как минимум, несколько последовательных шагов.

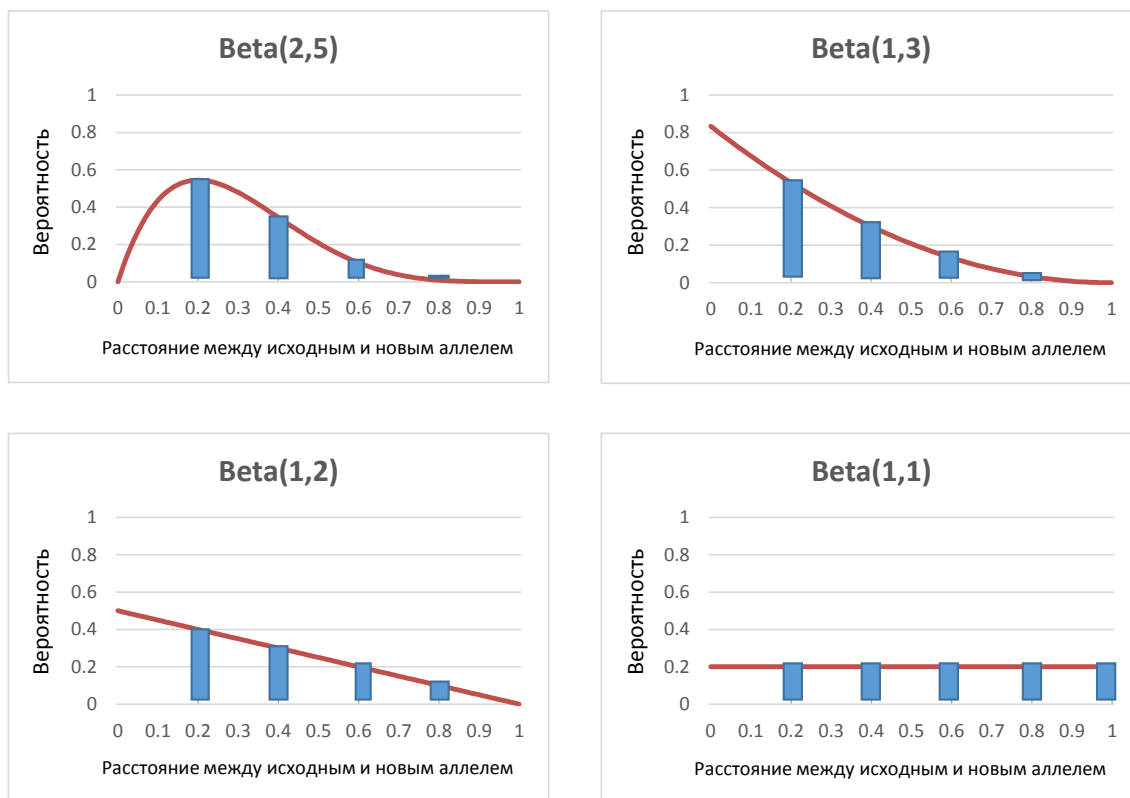


Рисунок 2.7.3. Функции плотности распределения вероятности изменения аллельного состояния в результате мутации. Столбцы – дискретные аппроксимации бета-распределения $Beta(a, b)$. В каждом случае возможны переходы аллеля в одно из пяти новых состояний (шаг дискретизации равен 0.2).

Частота мутаций

Для того чтобы иметь возможность варьировать вероятность мутаций в модели, мы ввели параметр *частоты мутаций*. Этот параметр определяет вероятность того, что аллель материнской клетки изменит своё значение у

дочерних клеток. Мы оценили этот параметр на основе имеющихся литературных данных: прежде всего, известно, что для прокариот он составляет порядка 0.003-0.005 мутаций на геном на поколение (МГП) (Drake et al., 1998; Lynch, 2010), затем, размер генома (РГ) прокариот, как правило, находится в диапазоне ~0.5-5 Мб, в то время как типичная длина фермента составляет порядка 300-400 аминокислот (т.е. соответствующие гены имеют длину (ДГ – длина гена) около 900-1200 нуклеотидов). Таким образом, вероятность того, что такой локус окажется мутирован, можно оценить с помощью произведения $\text{МГП} \times (\text{ДГ}/\text{РГ})$: $5.4 \cdot 10^{-7} - 1.2 \cdot 10^{-5}$ на локус на клетку на поколение. Мутации в разных локусах моделируются независимо и равновероятно. В литературе отмечался тот факт, что частота мутаций и вероятность рекомбинации может существенно повышаться в условиях стресса (Foster, 2005; Hoffmann, Hercus, 2000; Markel, 2008). Так, Корогодин с коллегами показали, что стрессовые условия (в частности, у ауксотрофных по аденину дрожжей) могут приводить к тысячекратному повышению мутабельности генов, относящихся к метаболизму аденина (Korogodin et al., 1991; Ильина et al., 1987). Поэтому в данной работе мы рассматриваем частоту мутаций в пределах типичного диапазона в 0.000001-0.15 на клетку на поколение.

Схема численного эксперимента

Начальная популяция состояла из однородной группы клеток, "расположенных" в нижнем левом углу ландшафта приспособленности, соответствующего аллельному значению $(e_1, e_2) = (1, 1)$. Начальный размер популяции составлял 1×10^7 клеток. Дальнейший рост популяции обеспечивал постепенное достижение максимального размера, равного 1.5×10^7 (для стартового генотипа) и обусловленного параметрами среды обитания. В ходе дальнейшей эволюции и движения по ландшафту приспособленности происходит рост средней приспособленности клеток популяции, что приводит

к дальнейшему росту размера популяции, достигая соответствующего пика максимума в 2.1×10^7 клеток.

Мы проанализировали данную модель на протяжении 10 000 поколений, за каждое поколение выполняется стандартная итерация ГЭК, состоящая из следующих шагов:

1. Клетки поглощают субстраты из окружающей среды;
2. Клетки метаболизируют поглощённые субстраты, используя их для:
 - a. Синтеза и секреции некоторых метаболитов;
 - b. Размножения, в ходе которого могут произойти мутации;
3. Состояние окружающей среды обновляется, в т.ч. обрабатывается приток и отток субстрата, а также отток клеток из системы.

В результате мутаций могут появиться новые аллельные варианты. Таким образом, мутации с каждым поколением повышают генетическое разнообразие популяции и приводят к «движению» потомков клетки по ландшафту приспособленности с течением времени, измеряемом в поколениях. Некоторые аллельные комбинации могут оказаться более адаптивными, чем другие, поэтому несущие их мутантные клетки будут расти быстрее и постепенно вытеснят менее адаптивных клеток из популяции. В зависимости от глубины «рва», окружающего центральный пик на ландшафте приспособленности (рис. 2.7.2), клеткам потребуется более или менее интенсивный поток генов, «поставляемых» мутационным процессом, чтобы достичь наиболее приспособленного состояния.

Корреляции между скоростью утилизации субстрата и приспособленностью клеток

Как мы отметили ранее, в данной модели нет явной зависимости между приспособленностью клеток и их генотипом. Эта взаимосвязь определяется *a posteriori* через моделирование действия естественного отбора на популяционном уровне. Она зависит не только от эффективности утилизации субстрата, но также и от доступности субстрата в среде (которая может быть пространственно неоднородна), а также от скорости гибели клеток. Проверка

этого соответствия с использованием данных моделирования представляла особый интерес. Ради этой цели мы посчитали среднюю скорость размножения клеток, т.е. число потомков в единицу времени для представителей каждого генотипа, определяемого как аллельная комбинация. Затем мы построили зависимость этой скорости от значений $K_u = F(e_{1,i}, e_{2,j})$ для всех узлов поверхности. Результаты представлены на (рис. 2.7.4) и демонстрируют линейную зависимость между значением K_u и средней скоростью размножения клеток: $Fitness = a \cdot K_u + b$, где $a = 0.0035$; $b = 10^9$, коэффициент корреляции Пирсона равен 0.9999 ($p < 10^9$). Следовательно, можно сделать вывод, что в нашей модели поверхность приспособленности изоморфна поверхности, задаваемой функцией $F(e_1, e_2)$, т.е. её можно использовать в качестве заместителя для исследования ландшафта приспособленности. Схожие результаты были получены для всех вариантов поверхности. Таким образом, было показано, что в рассматриваемых условиях поверхность K_u полностью определяет относительную приспособленность клеток как функцию их генотипа в нашей модели.

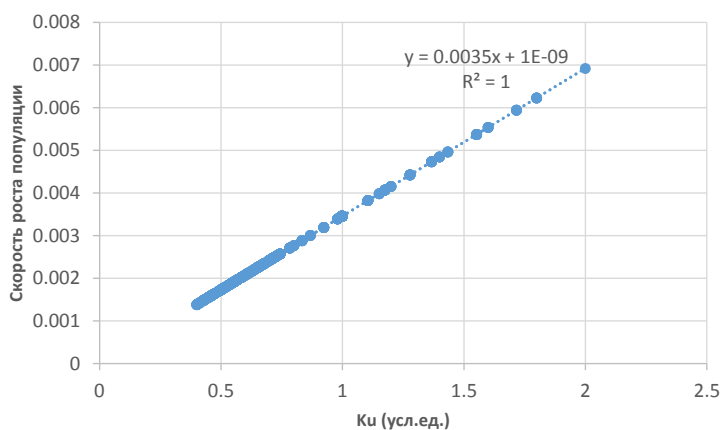


Рисунок 2.7.4. Зависимость скорости роста популяции (число потомков на клетку за единицу времени) от значения K_u .

Достижение максимума средней приспособленности популяции на разных ландшафтах

Главный вопрос, на который мы стремились ответить, звучит следующим образом: как взаимосвязаны ландшафт приспособленности, частота мутаций и

распределение аллелей среди мутантов? Как эти факторы способствуют или препятствуют появлению наиболее приспособленных клеток? Размер наиболее приспособленной субпопуляции (НПС) напрямую зависит от частоты мутаций. Когда она низкая, в системе появляется недостаточно мутантов для того, чтобы популяция в целом могла преодолеть долину приспособленности. Когда частота мутаций высока, то даже наиболее приспособленные клетки продолжают мутировать, и НПС не может достигнуть критической доли во всей популяции. Если мутационный шаг мал, то маловероятно, что даже интенсивный мутационный процесс произведёт наиболее приспособленные клетки.

В таблице 2.7.2 и на рисунке 2.7.5А представлена итоговая доля НПС (т.е. клеток, генотип которых соответствует области пика приспособленности). В случае малого мутационного шага при использовании распределений Beta(2,5) и Beta(1,3), клетки гибнут в глубине долины, не доходя до области наиболее приспособленного генотипа. Мы повышали параметр частоты мутации, чтобы найти размер получившейся НПС. В случаях распределений Beta(2,5) и Beta(1,3) это даёт возможность клеткам достигать зоны наиболее приспособленных генотипов. Однако, размер НПС снизился во всех прочих вычислительных экспериментах из-за постоянной потери клеток в результате приобретения ими вредных мутаций (табл. 2.7.3, рис. 2.7.5В).

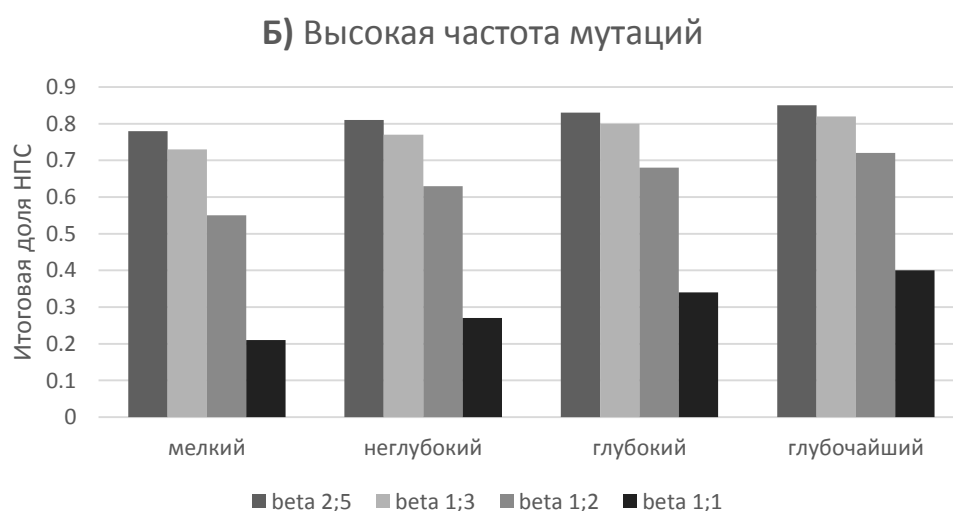
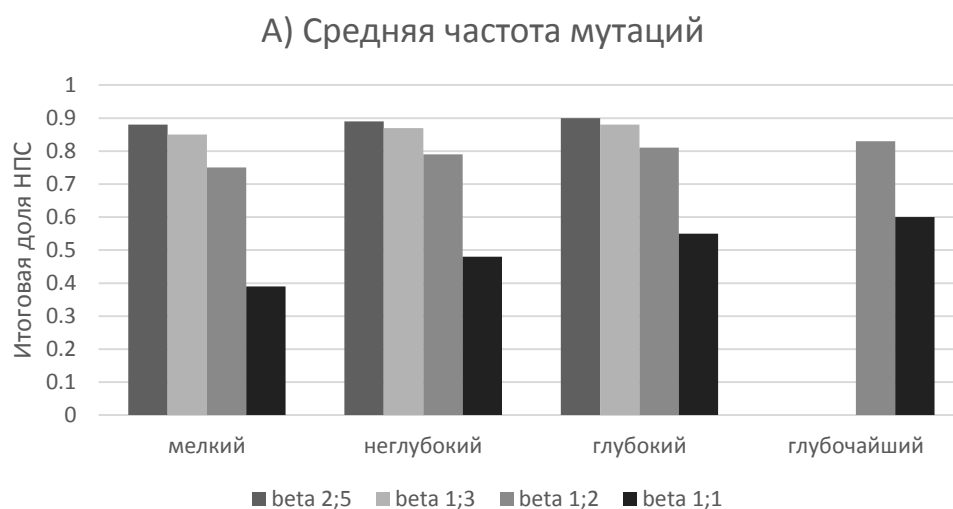


Рисунок 2.7.5. Итоговая доля НПС – клеток наиболее приспособленной субпопуляции, полученная по окончании процесса моделирования для разных вариантов ландшафтов и распределений мутантов. А) Средняя частота мутаций; Б) Высокая частота мутаций.

Таблица 2.7.2. Итоговая доля наиболее приспособленной субпопуляции в случае средней частоты мутаций.

Ландшафт распределение	S_0.44_0.86	S_0.4_0.8	S_0.34_0.73	S_0.28_0.67
Beta(2,5)	0.88	0.89	0.9	0
Beta(1,3)	0.85	0.87	0.88	0

Beta(1,2)	0.75	0.79	0.81	0.83
Beta(1,1)	0.39	0.48	0.55	0.6

Из рисунка 2.7.5 видно, что если изменения значений аллелей малы (Beta(2,5) и Beta(1,3) согласно рис. 2.7.3), то вероятность преодоления долины низкой приспособленности теряется при средних частотах мутаций. Мы также установили зависимость между шириной долины и распределением генотипов в том случае, когда частота мутаций была низкой. В таблице 2.7.3 показано существование наиболее приспособленной популяции в разных вариантах расчёта моделей.

Таблица 2.7.3. Существование наиболее приспособленной субпопуляции (НПС) в конце расчётов модели в зависимости от ширины долины ландшафта приспособленности и распределения мутаций. Знаком «+» обозначено существование НПС, «-» - отсутствие НПС.

Ширина долины / Распределение	0	0.2	0.4	0.6	0.8
Beta(2;5)	+	+	-	-	-
Beta(1;3)	+	+	-	-	-
Beta(1;2)	+	+	+	-	-
Beta(1;1)	+	+	+	+	-

Влияние популяционных параметров и параметров окружающей среды

Мы рассмотрели влияние других параметров модели на достижимость пика приспособленности. Коэффициент k_{dens} из уравнения 2.7.1 отражает долю особей, которые погибают во время процесса изменения численности популяции по квадратичному закону (плотностно-зависимым образом). Коэффициент протока k_{flow} из уравнения 2.7.1 описывает линейное вымывание части клеток из популяции. Наконец, объём окружающей среды (*volume*)

оказывает прямое влияние на размер популяции путём сдерживания её роста (методические детали, связанные с описанием объёма окружающей среды в моделях ГЭК приведены в нашей предыдущей работе (Lashin et al., 2014)). Если в окружающей среде достаточно субстрата, и в популяции не действуют никакие ингибирующие законы, уменьшающие размер популяции при её росте, то увеличение *volume* гарантированно даёт увеличение максимально возможного размера популяции.

Сочетание трёх перечисленных параметров определяет, будет ли преодолена долина заданного ландшафта при заданной скорости мутационного процесса и заданном распределении значений новых, мутантных аллелей. В таблицах 2.7.4 и 2.7.5 можно видеть на каком поколении достигается ситуация, когда доля НПС достигает 99%.

В таблицах 2.7.4 и 2.7.5 указан номер поколения, на котором достигнуто 99% присутствие у особей сочетания аллелей, соответствующего вершине пика на ландшафте приспособленности. *x* означает, что такого сочетания за время моделирования (500 000 поколений) получено не было (то есть, возможны два варианта: либо в популяции образовалась другая стационарная точка (как правило, формируется пик недалеко от начала долины), и долина не смогла быть преодолена, либо долина могла бы быть преодолена позже, но 500 000 оказалось недостаточно для этого). Приведены результаты для высокой частоты мутаций.

Таблица 2.7.4. Номер поколения фиксации 99% доли НПС в зависимости от k_{dens} и *volume* (при $k_{flow} = 0.01$).

$\text{Log}_{10}(k_{dens})$	-16	-17	-18	-19
$\text{Log}_{10}(\text{volume})$				
4	x	x	x	x
5	x	x	453276	445499

6	x	418906	313627	284643
7	x	399870	265096	204690

Таблица 2.7.5. Номер поколения фиксации 99% доли в зависимости от k_{dens} и k_{flow} (при $volume=10^6$ мл).

$\text{Log}_{10}(k_{dens})$ k_{flow}	-16	-17	-18	-19
0.001	x	x	454093	446415
0.01	x	418906	313627	284643
0.1	x	399824	264252	202600

Необходимо отметить, что в основе нашей модели лежит представление о физико-химических взаимодействиях между субъединицами фермента, отвечающего за утилизацию субстрата. Эти субъединицы представлены двумя различными локусами в геноме. Таким образом, различные аллельные значения не определяют приспособленность организма напрямую, как в большинстве подходов, использующихся в этой области (Gokhale et al., 2009; Weissman et al., 2010, 2009). Другие факторы, такие как пространственная доступность субстрата, например, могут также оказывать влияние на размножение клеток. Влияние этих дополнительных факторов стохастично по своей природе и варьирует от клетки к клетке, однако, мы показали, что в нашей модели активность фермента хорошо коррелирует со способностью клетки произвести потомка (т.е. с приспособленностью) (рис. 2.7.4). Эти результаты показывают применимость моделей, опирающихся на молекулярный фенотип как предиктор приспособленности при описании эволюционных процессов в прокариотической популяции.

В основе нашей модели лежит статический ландшафт приспособленности. Эта концепция не может отражать динамические изменения параметров окружающей среды. Однако, она применима в тех случаях, когда, например, окружающая среда меняется медленно по сравнению со временем, которое требуется для того, чтобы популяция достигла пика приспособленности. С другой стороны, если изменения в ландшафте характеристик происходят быстро, то эволюционный процесс можно представить как движение по поверхности приспособленности, усреднённой по времени (Wilke, 2001). В этих случаях статический ландшафт может выступать в качестве оправданного приближения.

Другим результатом этого раздела является то, что форма распределения вероятности приобретения аллельных значений в результате мутационного процесса имеет важное значение для способности популяции преодолевать долину. Этот параметр может служить в качестве посредника при рассмотрении изменчивости физико-химических свойств аминокислот при мутациях. Он отражает сложность соответствия между мутациями нуклеотидов в кодоне и изменений в физико-химических свойствах аминокислоты, что очень важно для ферментной активности. Значения этих свойств распределены не равномерно между 20 аминокислотами, входящими в состав белка, но напротив, для большинства свойств аминокислоты кластеризуются в независимые группы (полярные/неполярные, малые/большие, заряженные/незаряженные) (Kawashima et al., 2007). Как мы показали в данной модели, это соответствие может быть важно для эволюции популяции.

В целом, наши результаты раскрывают характер взаимосвязей между мутационными изменениями значений аллелей (заместителей физико-химических свойств аминокислот), частотой мутаций и формой ландшафта приспособленности, зависящей от параметров активности фермента, при эволюционном сценарии, когда популяция гаплоидных организмов пересекает долину низкой приспособленности на ландшафте приспособленности.

На основании данных моделирования можно предложить следующий ряд биологических интерпретаций:

- 1) Чем более однородным является распределение аллелей, тем более вероятно, что глобальный максимум будет достигнут (см. результаты моделирования для распределения $Beta(1,1)$).
- 2) Чем менее однородным является распределение аллелей, тем выше вероятность достичь локальных максимумов (см. результаты моделирования для распределения $Beta(2,5)$).
- 3) Локальные максимумы формируются в областях с наибольшим числом высокоприспособленных соседей (зоны, располагающиеся вблизи долины и начинающиеся в каждом из углов ландшафта приспособленности).
- 4) Чем выше частота мутаций, тем выше шанс достичь глобального максимума, но ниже число клеток, относящихся к наиболее приспособленной субпопуляции (НПС).
- 5) Чем ниже частота мутаций, тем ниже шанс достичь глобального максимума, но выше число клеток НПС (зона локального максимума или глобального, если он существует).
- 6) Чем меньше глубина долины, тем выше вероятность достичь глобального максимума.

2.7.2. Моделирование формирования и конкуренции популяций подвижных микроорганизмов с различными адаптационными стратегиями в периодически изменяющейся среде

Подвижность является одним из ключевых факторов адаптации микроорганизмов к разнообразным условиям внешней среды (Grossart et al., 2001), обеспечивая бактериям преимущество в ограниченных ресурсами физически структурированных местообитаниях даже в случае неориентированной подвижности (Wei et al., 2011). В то время, как процессы адаптации, опосредованные хемотаксисом, относительно хорошо изучены на

микроскопическом уровне (Adler, 1976; Asakura, Honda, 1984; Bren, Eisenbach, 2000; Falke et al., 1997; Macnab, Koshland, 1972; Mitchell, Kogure, 2006; Tindall et al., 2008; Wadhams, Armitage, 2004), существует недостаточное понимание того, как способность к адаптивным миграциям влияет на эволюционный успех популяции. В целом, адаптивные миграции микроорганизмов выступают ключевыми факторами, определяющими борьбу за существование в микробных экосистемах (Bridier et al., 2017; Niehus et al., 2015). В борьбе за существование существуют две различные стратегии, которые применяют микроорганизмы, столкнувшись с изменением условий среды – либо адаптироваться к ним, в частности, подстраивая свой метаболизм к альтернативным источникам энергии или более эффективно расходуя имеющиеся запасы, либо мигрировать в след за оптимальными условиями среды, вторгаясь в новые биотопы и вступая в конкуренцию с видами, принадлежащими локальным сообществам (Callahan et al., 2014; Chew et al., 2014; Liu et al., 2016). Обе стратегии характерны для морских бактерий, которые могут либо прикрепляться к источникам питательных веществ и колонизировать их путём выращивания биопленки, либо быстро обнаруживать недолговечные источники питательных веществ и постоянно двигаться к ним, как это делают мобильные виды (Yawata et al., 2014). Влияние такого индивидуального адаптивного поведения на динамику уровня сообщества представляет фундаментальный научный интерес и освещается с точки зрения конкретных аспектов экологии и эволюции поискового поведения (Hein et al., 2016).

Проблема взаимосвязи миграционной активности и приспособленности популяции была исследована у многих биологических видов (Roff, 1988; Waite et al., 2018; Wong-Ng et al., 2016; Xu et al., 2017). Подвижность не только улучшает кормовое поведение и повышает шансы на успешное нахождение партнёров для размножения, но и может способствовать развитию кооперативного поведения в мобильных группах (Aktipis, 2011). Однако миграционная активность организмов несет определенные энергетические

затраты, которые могли бы вместо этого быть затрачены на воспроизводство (Ni et al., 2020). Различные группы по-разному управляют этим возникающим компромиссом между репродуктивными инвестициями и миграцией. В то время как, например, чавыча (чинукский лосось) прибегает к модификации развития, приводящей к увеличению отношения числа икринок к размеру икринки в сочетании с большим расстоянием миграции (Kinnison et al., 2001), одноклеточные организмы не могут обладать такой гибкостью на уровне тканей и органов и вынуждены решать эту задачу, принимая определенную жизненную стратегию на видовом уровне.

Подвижность и хемотаксис широко распространены среди бактерий, обитающих в морской среде (Grossart et al., 2001; Mitchell et al., 1995; Stocker, Seymour, 2012), хотя среди них встречаются и неподвижные микроорганизмы, такие как SAR11 (Giovannoni, 2017, 2005) и *Prochlorococcus* (Liu et al., 1997). Современные оценки энергетических затрат, затрачиваемых на подвижность морских бактерий, показывают, что она колеблется от 2% до 50% от их общего энергетического бюджета (Stocker, Seymour, 2012) для видов с различной подвижностью.

Затраты на подвижность бактерий, с точки зрения изменения их приспособленности к среде, связаны не только с прямыми затратами энергии на движение (например, через вращение жгутиков), но и с поддержанием соответствующей «молекулярной инфраструктур», используемой, например, для биосинтеза жгутиков (Ni et al., 2017). Таким образом, миграция через хемотаксис может быть довольно дорогостоящей для энергетических бюджетов подвижных бактерий, что создает компромисс между вложением энергии в оптимальные стратегии добывания корма и распределением ресурсов на прямое воспроизводство. Изучение влияния такого компромисса на приспособленность популяции в различных экосистемах представляет собой большую проблему для экологического моделирования, и именно этот вопрос и рассмотрен в настоящем разделе.

Компьютерная модель конкуренции «оседлой» и «кочевой» популяций (неспособных и способных к хемотаксису, соответственно) в среде обитания с периодическим во времени и пространстве субстратом была построена с помощью ПК ГЭК чтобы оценить, в каких условиях более приспособленной будет более мобильная популяция, а в каких – популяция с более эффективным метаболизмом. Мы провели исследование модели сообщества, состоящего из двух популяций подвижных прокариотических организмов, эволюционирующих в пространственно-распределённой среде с диффузией и периодическим во времени и пространстве субстратом, являющейся аналогом застойного водоёма (см. Рис. 2.7.6 а). Мы рассмотрели 2D среду обитания в виде квадрата размером 10x10 ячеек с замкнутыми границами, в которую периодически с двух противоположных углов подавался питательный субстрат (рис. 2.7.6). Назовем «стартовым источником» ячейку, в которую осуществляется подача неспецифического субстрата (НС) в начальные моменты времени. Эту же ячейку исходно населяют обе популяции. В других ячейках в начальный момент времени клеток нет. Назовём «дальним источником» – исходное положение НС, расположенное в противоположном углу среды обитания (рис. 2.7.6 б). Первоначально питательное вещество распределялось пространственно однородно, но в последующем ходе итераций гетерогенность возрастает за счет локализованного периодического источника питательных веществ и кормовой активности клеток (рис. 2.7.6 в). Диффузия идёт по всем направлениям, распространяя по ячейкам молекулы субстрата и клетки. Последние также способны активно передвигаться в более благоприятные условия обитания с помощью хемотаксиса (клетки «оседлой» популяции – с меньшей интенсивностью, чем клетки популяции «кочевника») (рис. 2.7.6а).

В каждой пространственной ячейке численность популяции подчиняется следующему уравнению:

$$P_{n+1} = P_n * \left(1 + \frac{x/K}{1 + x/(C*K+N)} \right) - mort(P) \quad (2.7.4)$$

где P_{n+1} – численность популяции на (n+1)-ом поколении; x – число молекул нутриента на клетку; N – генетически определённая эффективность утилизации нутриента; C , K – настраиваемые параметры, описывающие кривую насыщения популяции нутриентом (в нашем случае, $K=5$, $C=0.8$); $mort(P)$ описывает естественную смертность популяции. В данном исследовании мы рассмотрели два типа смертности: квадратичную (плотностно-зависимую) и линейную (пропорциональную).

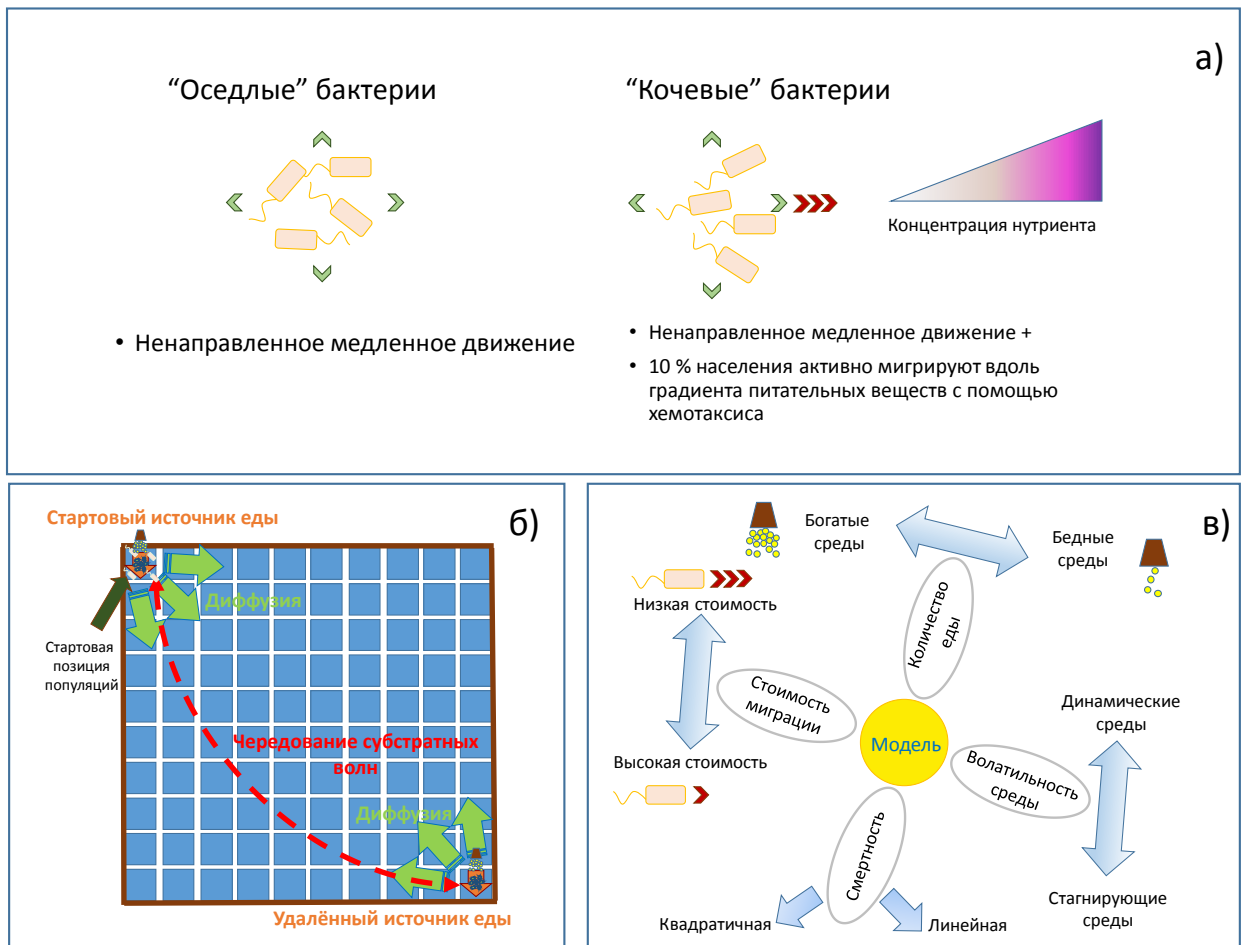


Рисунок 2.7.6. Схема модели конкуренции оседлой и кочевой популяции: а) трофическая структура сообщества; б) схема пространственной организации среды обитания сообщества; в) различные сценарии, рассмотренные в модели.

В первом случае смертность описывается следующим образом:

$$mort(P) = death_{coeff} * P_n^2 \quad (2.7.5)$$

во втором:

$$mort(P) = death_{coeff} * P_n \quad (2.7.6)$$

где $death_{coef}$ – это коэффициент смертности популяции.

Размер популяций в каждом конкретном узле зависит не только от размножения и гибели клеток, как описано в (уравнении 2.7.4), но и от миграционных потоков. Мы учитываем адаптивную миграцию клеток через хемотаксис, подробное описание используемого алгоритма можно найти в разделах выше, но здесь стоит упомянуть некоторые из его ключевых моментов. Свободноплавающая часть активно мигрирующих клеток (мы берем 10% популяции, что в целом согласуется с наблюдаемыми значениями для образцов прибрежной морской воды (Mitchell et al., 1995)) делится между благоприятными направлениями в соответствии со значениями притяжения соответствующих узлов решетки. Мы оцениваем притяжение условий окружающей среды в соседних (связных) узлах на основе разницы между аттрактантами в соседних и текущих узлах. После этого соседние узлы разделяются на два списка: непривлекательные узлы с более низкими значениями притяжения, чем в текущем узле, и привлекательные узлы, обладающие более высокими значениями притяжения. Сумма числа ячеек, вошедших из первого списка, за вычетом ячеек, вышедших в узлы из второго списка, определяет изменение размера популяции в текущем узле. Следует отметить, что часть активно движущихся ячеек в текущем узле делится между всеми благоприятными направлениями, и доля этого направления пропорциональна значениям притяжения соответствующих соседних узлов. Таким образом, мы предполагаем, что подвижные клетки, способные к хемотаксису, распространяются через окружающую среду как за счет случайного плавания, так и за счет градиентов питательных веществ, тогда как клетки без хемотаксической способности распространяются только случайным образом без какого-либо определенного направления.

Для учета миграционных издержек нами была предложена подмодель штрафа за миграцию с двумя параметрами. Первый параметр определяет степень нелинейности функции миграционных издержек (см. уравнение 2.7.7) и он зафиксирован во всех сценариях моделирования, рассмотренных в

данном разделе. Второй параметр соответствует базовому генетическому фону, определяющему затраты на миграцию, и он варьируется в различных сценариях моделирования, приводя к появлению *кочевых* популяций, характеризующихся различными энергетическими затратами на миграцию. Оба параметра могут быть отнесены к затратам, связанным с поддержанием соответствующих генетических регуляторных сетей, участвующих в механизмах, контролирующих механизм хемотаксиса в любом конкретном микроорганизме. В случае, если запас энергии клетки меньше, чем необходимо для «уплаты штрафа» за миграцию, миграция не производится, а энергия не расходуется.

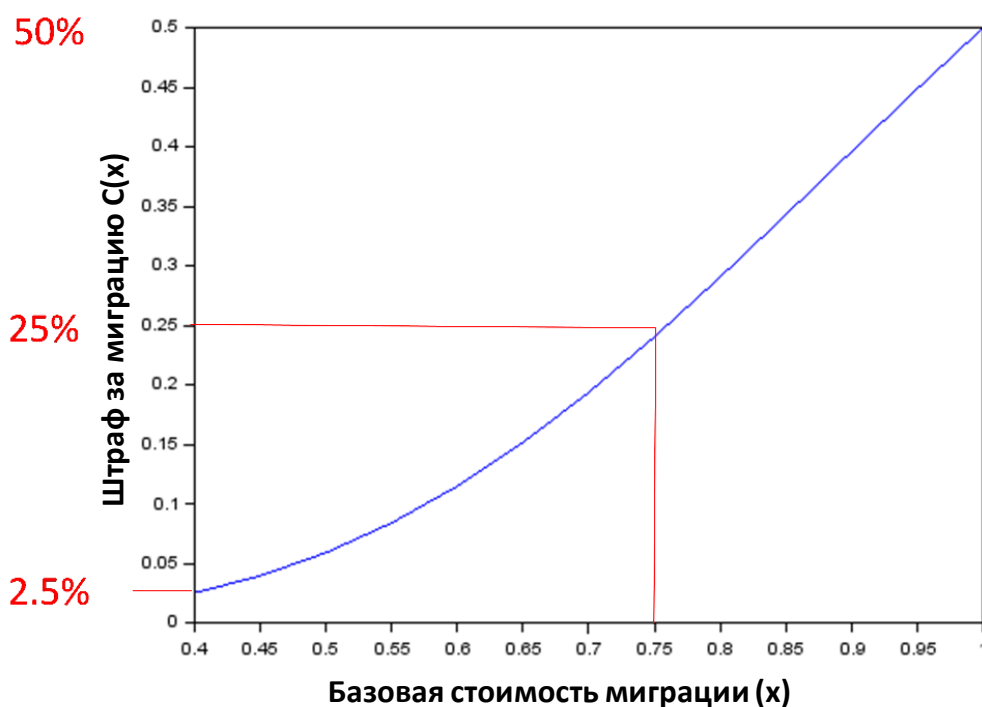


Рисунок 2.7.7. Зависимость штрафа за миграцию при значении параметра нелинейности $h = 4$.

Таким образом, мы определяем $C(x)$ как *функцию стоимости миграции* (рис. 2.7.7), следующее уравнение связывает энергетическую стоимость миграции и организм-специфические параметры, контролирующие затраты энергии на передвижение:

$$C(x) = \frac{x^h}{1 + x^h} \quad (2.7.7)$$

x – параметр, контролирующий базовую стоимость миграции, h – параметр нелинейности между энергетической стоимостью миграции и базовой стоимостью.

Таким образом, мы оцениваем штраф за миграцию как часть максимального значения запаса энергии клетки, который расходуется, если клетка мигрирует. В этой модели мы рассматриваем потреблённый неспецифический субстрат (НС) как энергетический резерв клетки. Таким образом, штраф за миграцию представляет собой фактические расходы накопленного НС на активное движение.

Проясним некоторые биологические предпосылки, лежащие в основе алгоритма моделирования хемотаксиса. Поскольку нас интересует динамика численности населения, и временная шкала ГЭК калибруется соответственно на основе среднего времени одного поколения (т.е. среднего времени деления клетки), то эффекты хемотаксиса проявляются в той же шкале, и мы рассматриваем его интегральное влияние на популяционном уровне. По этой причине мы не описываем подробно неравномерности движения отдельных клеток на коротких временных интервалах (таких как секунды или доли секунды), поскольку они в любом случае усредняются в этом масштабе моделирования. В свою очередь, на первый план выходят такие факторы, как адаптация к исходному сигналу в текущей ячейке и учёт градиента аттрактанта в широком масштабе (т. е. с учетом различий концентраций в текущей ячейке и её соседях). Первое достигается единичным характером акта хемотаксиса в течение поколения – клетки перемещаются в соседние ячейки, но физические параметры движения (импульс, инерция) не учитываются, роль играет только хемотактический сигнал (это вполне соответствует литературным данным по сигнальным путям бактериального хемотаксиса (Wadhams, Armitage, 2004)). Второе достигается с помощью алгоритма, оценивающего привлекательность ячеек (см. раздел 2.2.4 «Моделирование хемотаксиса»). Клетка,двигающаяся с посредством хемотаксиса, расходует энергию. Те клетки, у которых заканчиваются запасы энергии, останавливают вращение жгутиковых

двигателей до тех пор, пока они снова не наберут больше энергии (Куо, Koshland, 1989). Таким образом, ограничение активного движения для клеток, которым не хватает энергии, имеет ясный физический смысл.

Мы рассмотрели две группы модельных сценариев: (1) сценарии, в которых популяции не имеют миграционного штрафа; (2) сценарии, в которых популяции имеют энергетические затраты на миграцию. Для тех и других сценариев мы варьировали такие параметры, как концентрация питательных веществ (нутриентов), поступающих в систему с подкормкой, период подкормки и тип смертности (либо уравнение 2.7.5, либо уравнение 2.7.6). Для модели, учитывающей затраты энергии на миграцию, мы также варьировали величину миграционного штрафа. В таблице ниже приведены рассмотренные варианты модельных сценариев:

Таблица 2.7.6. Параметры, использованные в модельных сценариях «оседлый-кочевник».

		Возможные значения									
Параметры	<i>Концентрация нутриента в подкормке</i> <i>Log10(моль/л)</i>	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10
	<i>Период подкормки (поколения)</i>	1	25	50	100	<i>От 100 до 500, в отдельных случаях</i>					
	<i>Стоимость миграции (%)</i>	0	2	5	10	15	20	25	33.33	50	99.99
	<i>Смертность</i>	Квадратичная					Линейная				

Малые значения периода подкормки соответствуют более динамичным средам, тогда как большие значения – более застойным средам. Будем

называть среды обитания, характеризующиеся высокой концентрацией питательных веществ в порции подкормки (10^{-1} моль/л), «богатыми» средами, а среды, характеризующиеся низкой концентрацией (10^{-5} моль/л) – «бедной». Среда, несущая ещё более низкую концентрацию НС в подкормке, в дальнейшем будут называться «чрезвычайно дефицитными». Сравнение двух типов смертности позволяет сделать вывод о том, является ли плотностно-зависимая смертность причиной эффектов, наблюдаемых в модели, или же у этих эффектов существует более глубокая причина.

Все модельные сценарии выполнялись в течение 9000 итераций (поколений), что дало результаты моделирования динамики популяционной численности и других характеристик в системе.

В данном исследовании исследовалась модель конкуренции подвижных и оседлых видов и влияние способности к адаптивной миграции на приспособленность популяции в различных условиях. Существует целый ряд факторов – как популяционных, так и экологических, – которые формируют контекст конкуренции между подвижными и малоподвижными микроорганизмами. Концентрация питательных веществ в порции и период порции относятся к числу факторов окружающей среды, в то время как миграционные энергетические затраты и смертность, характеризующие естественную убыль популяции, можно рассматривать как неотъемлемые черты популяций в целом. Отметим, что два типа смертности, рассмотренные выше – либо в виде квадратичного (ур. 2.7.5), либо в виде линейного (ур. 2.7.6) слагаемого, – соответствуют классической ферхюльстовской модели плотностно-зависимой смертности и пропорциональной смертности, вызванной постоянными неблагоприятными факторами, соответственно. В качестве таких факторов, например, может выступать среднее давление внешних аменсалов, которое не учитывается в моделируемой экосистеме. Сравнение результатов моделирования двух типов модели – учитывающей энергетические затраты на миграцию и не учитывающей, позволяет ответить на следующие вопросы:

- Каковы критические значения штрафов за миграцию, которые «кочевник» может выдерживать, сохраняя свое доминирование в системе?
- Зависит ли критическая значения от других экологических и демографических факторов?

Влияние способности к хемотаксису на приспособленность

Прежде всего мы провели численное исследование модели, которая не учитывает энергетические затраты на миграцию. Это наиболее крайний случай, биологически соответствующий ситуации, когда механизм, лежащий в основе подвижности клеток, устроен достаточно эффективно, настолько, что энергетические затраты на миграцию ничтожны. Полученные результаты представлены в таблице 2.7.7.

Таблица 2.7.7. Сводная таблица результатов численного исследования модели, в которой не учитываются затраты на миграцию. В ячейках таблицы приведены популяции, являющиеся доминирующими в системе к финальному моменту времени. В «крайне дефицитной» среде «кочевник» вытесняет «оседлого» вне зависимости от периода подкормки (период варьировался от 50 до 1000 поколений) при обоих типах смертности.

<u>Смертность</u>	Линейная	период 50	период 1000	период 1	период 1000
		Кочевник	Кочевник	Оседлый (почти всегда в паритете с Кочевником)	Оседлый
	Квадратичная	период ≤ 127	период ≥ 128	период 1	период 1000
		Кочевник	Оседлый	Оседлый (Кочевник вымирает)	Оседлый (Кочевник вымирает)
		«Бедная» среда		«Богатая» среда	
		<u>Условия среды</u>			

Как видно из таблицы 2.7.7, тип смертности, период подкормки и концентрация в ней питательных веществ влияют на то, какая популяция будет

доминировать в экосистеме. Стоит отметить, что даже без каких-либо миграционных штрафов «кочевник» способен извлечь выгоду из своего адаптивного поведения только при определенных условиях, а именно, в бедных питательными веществами и динамичных средах, где пространственная локализация источника питательных веществ быстро меняется, определяя динамическое неравенство в доступности питательных веществ. Анализ пространственного распределения популяций по ячейкам (см. рис. 2.7.8) позволяет нам разобраться, как именно «оседлый» и «кочевник» разделяют общую среду обитания. При линейной смертности в «бедной» среде обитания «кочевник» вытесняет «оседлого» вне зависимости от периода подкормки. Это достигается за счёт того, что «кочевник» свободно перемещается между обоими источниками питательных веществ и занимает пространство, которое является более выгодным в каждый момент времени. При квадратичной смертности в «бедной» среде действие плотностно-зависимых факторов приводит к тому, что в застойных средах, характеризующихся длительным периодом подкормки (1000 поколений), доминирование «оседлого» в системе является скорее номинальным: популяции поляризуются по источникам питательных веществ – «кочевник» занимает дальний источник, а поселенец контролирует стартовый источник питательных веществ.

В «богатых» средах при линейной смертности «оседлый» контролирует стартовый источник питательных веществ независимо от периода подкормки. Однако период подкормки влияет на ситуацию вокруг дальнего источника питательных веществ – в более динамичных средах «кочевнику» удастся занять его, в то время как в более застойных, по мере неуклонного увеличения периода подкормки, «оседлый» оттесняет «кочевника» на окраину ареала обитания. В случае квадратичной смертности в «богатых» средах, «кочевник» вымирает, а «оседлый» процветает независимо от значения периода партии.

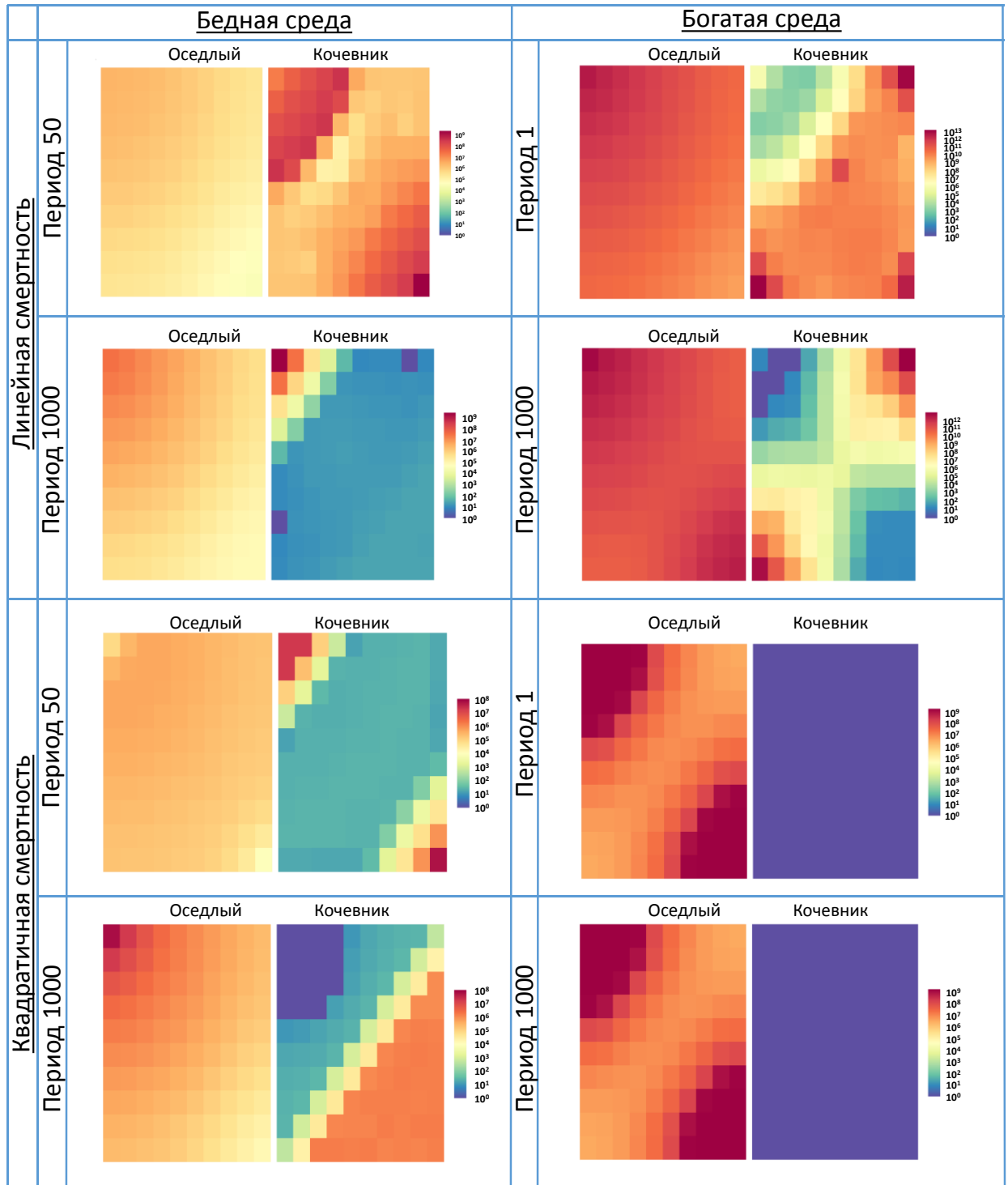


Рисунок 2.7.8. Тепловые карты пространственных распределений временных снимков клеток «оседлых» и «кочевников» в среде обитания, сделанные в конце численных экспериментов, выполненных по различным модельным сценариям. Шкала тепловой карты представляет собой численность популяции (количество клеток). Стартовый источник питательных веществ расположен в верхнем левом углу, а дальний – в противоположном нижнем правом углу.

Влияние энергетической стоимости миграции

Рассматривая миграционный штраф, важно выяснить, каковы его пороговые значения, определяющие доминирование «кочевника» или «оседлого» в среде обитания. Мы тщательно изучили влияние величины миграционного штрафа на исход конкуренции в данной модели при различных типах смертности и в условиях, характеризующихся различной концентрацией питательных веществ в подкормке, но с одним и тем же фиксированным периодом, равным 50 поколениям. Сводные результаты представлены в таблицах 2.7.8 и 2.7.9.

Таблица 2.7.8. Доля популяции «кочевника» в финальном сообществе в зависимости от величины миграционного штрафа в средах, варьирующихся по концентрации питательных веществ в подкормке от крайне бедных до богатых. Период подкормки равен 50 поколениям, смертность – квадратичная.

		Штраф за миграцию (%)									
		0%	2%	5%	10%	15%	20%	25%	33.33%	50%	99.99%
Концентрация нутриента в подкормке (Log ₁₀ (моль/л))	-1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-3	66%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-4	99%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-5	83%	56%	1%	1%	1%	1%	1%	1%	1%	1%
	-6	41%	67%	65%	11%	11%	11%	11%	11%	11%	11%
	-7	14%	50%	50%	50%	50%	50%	50%	50%	50%	50%
	-8	61%	50%	50%	50%	50%	50%	50%	50%	50%	50%
	-9	77%	50%	50%	50%	50%	50%	50%	50%	50%	50%
	-10	78%	50%	50%	50%	50%	50%	50%	50%	50%	50%

Таблица 2.7.9. Доля популяции «кочевника» в финальном сообществе в зависимости от величины миграционного штрафа в средах, варьирующихся по концентрации питательных веществ в подкормке от крайне бедных до богатых. Период подкормки равен 50 поколениям, смертность – линейная.

		Штраф за миграцию (%)									
		0%	2%	5%	10%	15%	20%	25%	33.33%	50%	99.99%
Концентрация нутриента в подкормке (Log10(моль/л))	-1	49%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-2	49%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-3	50%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-4	52%	51%	1%	1%	1%	1%	1%	1%	1%	1%
	-5	99%	65%	65%	64%	6%	6%	6%	6%	6%	6%
	-6	96%	81%	80%	20%	20%	20%	20%	20%	20%	20%
	-7	61%	50%	50%	50%	50%	50%	50%	50%	50%	50%
	-8	71%	50%	50%	50%	50%	50%	50%	50%	50%	50%
	-9	72%	50%	50%	50%	50%	50%	50%	50%	50%	50%
	-10	72%	50%	50%	50%	50%	50%	50%	50%	50%	50%

Наблюдаемые паттерны не описываются простыми линейными зависимостями. Однако, по мере увеличения величины миграционного штрафа, доля «кочевника» уменьшается до тех пор, пока не будет достигнут определенный предел, который различен для сред, характеризующихся различной концентрацией питательных веществ в подкормке. Этот факт может быть объяснен либо отсутствием у кочевников энергетических резервов для перемещения куда-либо по мере увеличения миграционного штрафа и снижения доступности питательных веществ, либо расходом этих ограниченных ресурсов на обустройство уже занятых ячеек, а не на рекультивацию новых территорий.

В относительно изобильных условиях, характеризующих промежуточные среды между богатыми и бедными, введение миграционного штрафа приводит к тому, что доля «кочевников» становится ничтожной, вплоть до полного вымирания «кочевников» (см. таблицу 2.7.8). Примечательно, что тип смертности не оказывает существенного влияния на эту тенденцию, хотя учёт плотностно-зависимых факторов, безусловно, ухудшает последствия для «кочевников», тогда как линейная смертность смягчает их.

Стоит отметить, что существует определенная критическая точка в значении штрафа за миграцию, которая определяет достаточно узкую область пространства параметров, где стратегия «кочевника» оказывается адаптивной и положение этой точки зависит от условий окружающей среды, в частности – от концентрации питательных веществ в подкормке. Пространственное распределение клеток разных популяций проясняет причины, лежащие в основе этого «переключения доминирования» (см. рис. 2.7.9).

Рисунок 2.7.9 показывает, что, если нет явного доминирования ни «оседлого», ни «кочевника» (см. рис. 2.7.9 а, в), то обе популяции пространственно разделяются и разделяют контроль над различными источниками питательных веществ (что хорошо согласуется с предыдущими исследованиями (Yawata et al., 2014)) – «кочевник» заселяет дальний источник, а «оседлый» – начальный. Если смена доминирования происходит внезапно, то это обычно приводит к захвату обоих источников питательных веществ «оседлым», которому удается оттеснить «кочевника» назад, в то время как последний теряет энергию на перемещение (рис. 2.7.9 б, г).

Детальное исследование влияния периода поступления питательных веществ на все возможные эколого-эволюционные сценарии в данной системе является трудной задачей, которая может быть предметом отдельного исследования. Однако, грубая оценка для граничных случаев, богатых и крайне дефицитных сред, полученная для коротких и длинных периодов подкормки, показывает, что концентрация питательных веществ в подкормке и энергетические затраты на миграцию играют более значимую роль в определении исхода конкуренции в модели «оседлый – кочевник», чем длительность периода подкормки нутриентами (см. таблицу 2.7.10).

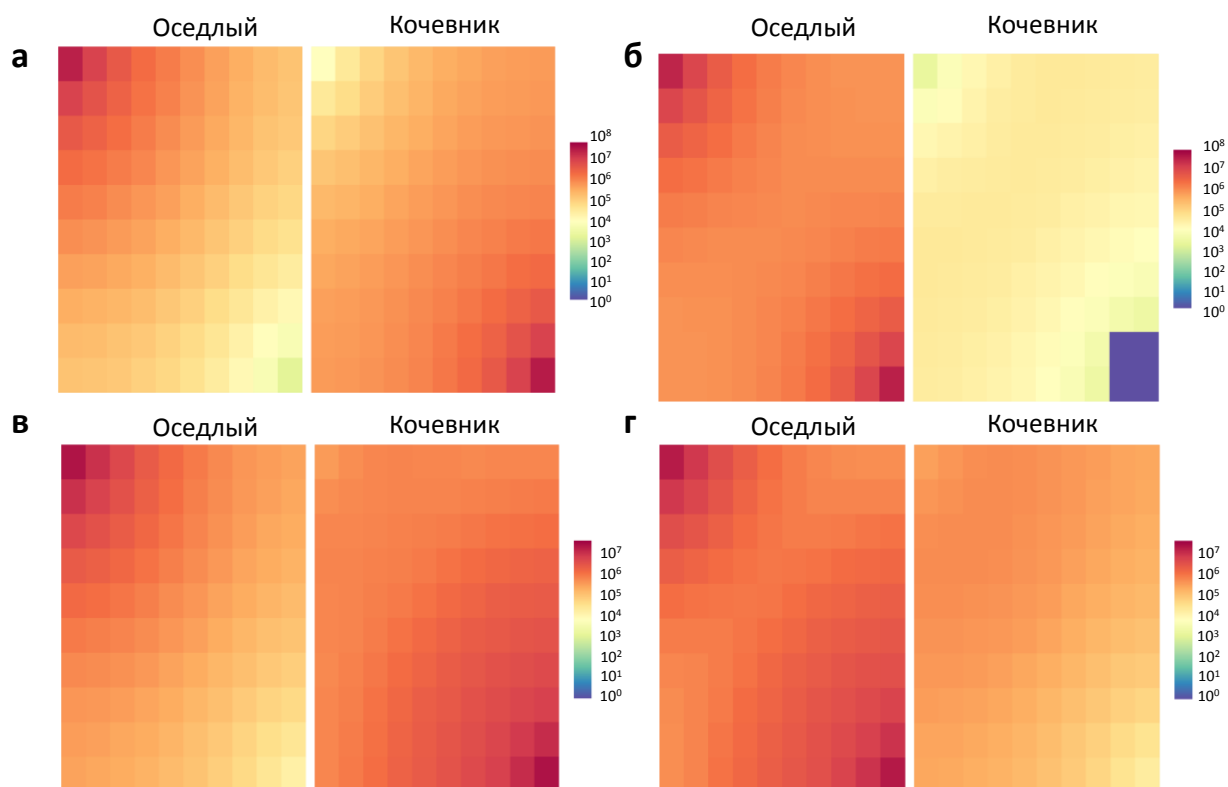


Рис. 2.7.9. Тепловые карты пространственных распределений временных снимков клеток «оседлых» и «кочевников» в среде обитания, сделанные в конце численных экспериментов, выполненных по различным модельным сценариям. Приведены результаты для сценариев с квадратичной смертностью (для линейной смертности результаты практически не отличаются) Шкала тепловой карты представляет собой численность популяции (количество клеток) в конечный момент времени моделирования. Стартовый источник питательных веществ расположен в верхнем левом углу, а дальний – в противоположном нижнем правом углу. Можно видеть, что по мере увеличения энергетических затрат на миграцию происходит переключение доминирования в следующих случаях: переход от сценария с концентрацией питательных веществ в подкормке, равной 10^{-5} моль/л, и величиной штрафа, равной 2% (а), к сценарию с той же концентрацией питательных веществ в подкормке, но более высокой величиной штрафа, равной 5% (б); и переход от сценария с концентрацией питательных веществ в партии, равной 10^{-6} моль/л и величиной штрафа, равной 5% (в), к сценарию

с той же концентрацией питательных веществ в подкормке, но более высокой величиной штрафа, равной 10% (г).

Таблица 2.7.10. Сводная таблица результатов численного исследования модели для разных наборов параметров. В ячейках таблицы приведены популяции, являющиеся доминирующими в системе к финальному моменту времени. Варианты с квадратичной и линейной смертностью не отличаются друг от друга за исключением случая, выделенного звездочкой (*), для которого в линейном случае получен полный паритет (равные численности популяций) «оседлого» и «кочевника». Период подкормки указан в поколениях.

Штраф за миграцию	Штраф $\leq 0.3\%$	период 50	период 1000	период 1	период 1000
		Кочевник	Кочевник	Оседлый *	Оседлый
	Штраф $\geq 0.4\%$	период 50	период 1000	период 1	период 1000
		Оседлый / Кочевник (полный паритет).	Оседлый / Кочевник (полный паритет).	Оседлый	Оседлый
		Крайне бедная среда		Богатая среда	
		<u>Условия среды</u>			

Кроме того, поскольку критическое значение миграционного штрафа слишком близко к нулю, это означает, что на самом деле в крайне бедной среде стратегия «кочевников» не может принести реальной выгоды, если энергетические затраты на миграцию являются сколь угодно значительными (начина с 0.4% от максимального запаса энергии клетки и выше). Однако, как было показано выше, существует определенный интервал параметров между крайне бедными и богатыми средами, где возможно, чтобы «кочевые»

популяции получали преимущество от своей подвижности для достижения доминирующего положения в экосистеме.

Данные результаты моделирования, показывающие, что стратегия «оседлых» может иметь преимущество перед стратегией «кочевников» в различных условиях, даже в отсутствие штрафов на миграции, выглядят весьма контринтуитивно. Тот факт, что в богатых питательными веществами средах даже при линейной смертности «кочевник» уступает первенство «оседлому», нуждается в объяснении. Мы связываем этот эффект с появлением «биотической пустыни» вокруг «оседлого», занимающего стартовый источник нутриентов (см. рис. 2.7.10). Поскольку клетки «оседлой» популяции уменьшают локальную доступность питательных веществ в непосредственной близости от источника, клетки популяции «кочевника» не может прорваться через эту самоорганизовавшуюся «зону отчуждения» к источнику питательных веществ из-за низкой концентрации питательных веществ в промежуточных ячейках. Таким образом, подвижные популяции, полагающиеся на локальную оптимизирующую стратегию, как правило, гоняющиеся за лучшими условиями (иногда прирост может быть совсем небольшим), терпят неудачу при длительном стрессе, связанном с пересечением долин субоптимальной доступности питательных веществ.

Интересно сравнить эти результаты с результатами моделирования преодоления локальных минимумов на ландшафте приспособленности (см. предыдущую подглаву диссертации 2.7.1). *Схожее влияние локальных минимумов, относящихся в совершенно разным иерархическим уровням биологической организации, полученное не за счёт использования одинаковых или похожих уравнений (как, например, в случае использования уравнения Вольтера для популяционно-экологических задач и задач химической кинетики), а за счёт построения механистической компьютерной модели (т.е. основанное на так называемых первых принципах), свидетельствует о существовании фундаментальных паттернов развития биологических систем.*

Результаты же численного исследования, изложенные в данном разделе, показывают, что существует переломный момент в величине миграционного штрафа, определяющий, является ли «кочевая» стратегия подвижной популяции адаптивной или нет. Однако эта величина в значительной степени зависит от наличия питательных веществ, а общий спектр условий, благоприятствующих стратегии «кочевников», оценивается как довольно узкий, что накладывает ограничения на компромисс «кочевников» между расходом энергии на миграцию в более благоприятные условия и «инвестициями» ресурсов в воспроизводство. Оказалось, что более динамичные и дефицитные среды благоприятствуют доминированию подвижных популяций, тогда как энергетическая стоимость передвижения превышает получаемую от движения выгоду в средах, богатых питательными веществами и застойных средах. Такие среды, как правило, способствуют доминированию «оседлых» микроорганизмов. Кроме того, плотностно-зависимая квадратичная смертность более губительна для подвижной популяции, чем линейная, поскольку хемотаксис способствует организации её клеток организовываться в сообщества высокой плотности. Хотя это и дает более адекватное описание, учёт энергетических затрат на миграцию не меняет этих основных тенденций в конкуренции между двумя типами микроорганизмов, полагающихся на различные стратегии с точки зрения подвижности.

Следует, однако, отметить, что мы не можем рассматривать «оседлую» популяцию как полностью неподвижную, поскольку коэффициент пассивного переноса клеток может быть интерпретирован как неориентированное движение, подобное диффузии для веществ. С этой точки зрения соперничество «оседлых» и «кочевников» может быть интерпретировано как соперничество между пассивно движущимися неповоротливыми клетками и клетками, обладающими сенсорным и сигнальным аппаратом, связанным с хемотаксисом, способными адаптивно двигаться в наиболее благоприятном направлении.

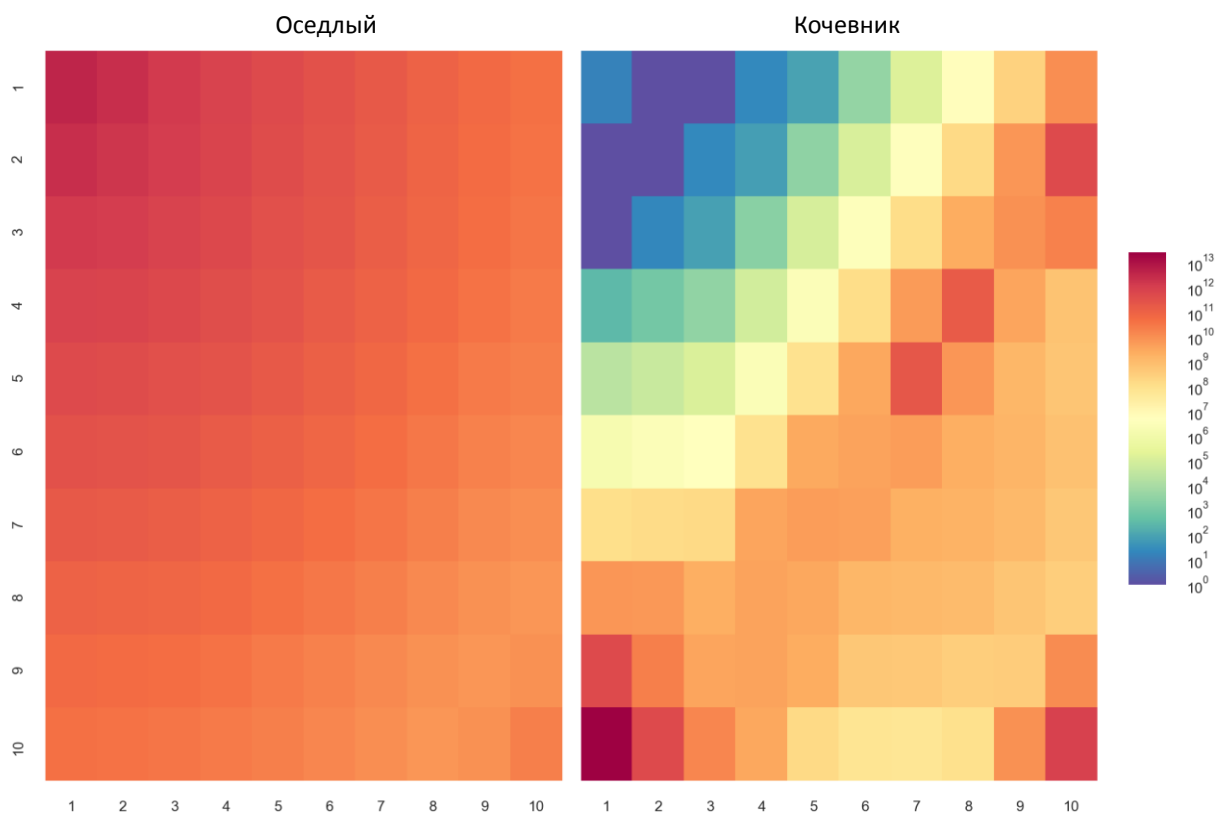


Рисунок 2.7.10. Тепловые карты пространственных распределений временных снимков клеток «оседлых» и «кочевников» в среде обитания, сделанные в конце численного эксперимента для богатой нутриентами среды обитания. Шкала тепловой карты представляет собой численность популяции (количество клеток) в конечный момент времени моделирования (9000 поколений).

Подвижные формы получают преимущество в условиях недостатка питательных веществ, соответствующих типичным средам обитания морских бактерий. Однако, поскольку в природе мы можем наблюдать процветание обеих стратегий, при этом ни одна из них полностью не вытесняет другую, вопрос о том, как такое разнообразие сохраняется, становится одной из интригующих проблем для дальнейших исследований в этой области. Динамическая и микрогетерогенная природа водных сред может быть одним из возможных факторов, приводящих к постоянно меняющемуся направлению отбора, препятствующему возникновению единственно верного направления

отбора. Другая важная проблема заключается в том, как генетическое и негенетическое разнообразие популяций по признакам подвижности влияет на приспособленность популяции в целом. Существует целый ряд исследований, посвященных этому вопросу (Frankel et al., 2014; Waite et al., 2018, 2016), хотя всеобъемлющий взгляд на эти вопросы ещё только предстоит сформировать. Несмотря на то, что современные исследования проясняют некоторые аспекты, механизмы и ограничения конкуренции между различными стратегиями с точки зрения подвижности бактерий, взаимосвязь между адаптивными миграциями микроорганизмов и их приспособленностью в общем случае остается фундаментальной проблемой, которая еще не исследована. И математическое и компьютерное моделирование, наряду с экспериментальными подходами, является ведущим инструментом для её решения.

2.7.3. Моделирование эволюции в популяциях подвижных микроорганизмов с различными адаптационными стратегиями в периодически изменяющейся среде

Рассмотренная в разделе 2.7.2 модель «Оседлый-кочевник» была дополнительно исследована с точки зрения эволюции признаков эффективности утилизации нутриентов и эффективности миграции. В то время, как в предыдущем разделе данные параметры не менялись в течение отдельного численного эксперимента, в модельных сценариях, рассмотренных в этом разделе, мы допускали возможность варьирования аллельного разнообразия по генам, ответственным за эти признаки.

По-прежнему, рассматривалась модель сообщества, состоящего из двух популяций подвижных прокариотических организмов, эволюционирующих в пространственно-распределённой среде с диффузией и периодической во времени и пространстве подкормкой нутриентом. Чтобы оценить, в каких условиях более приспособленной будет более мобильная популяция, а в каких – популяция с более эффективным метаболизмом, было проведено исследование модели сообщества, состоящего из двух популяций подвижных прокариотических организмов, эволюционирующих в пространственно-

распределённой среде с диффузией и периодическим во времени и пространстве субстратом, являющейся аналогом застойного водоёма (см. Рис. 2.7.6).

Была исследована эволюция данной системы на 25000 итерациях, что покрывает 4 полных цикла подачи субстрата. В ходе исследования варьировались такие параметры как количество субстрата в подкормке, а также следующие популяционные параметры – эффективность утилизации субстрата и интенсивность миграции (мобильность). Исследовались следующие вопросы:

- Как соотносятся по степени влияния на приспособленность популяций такие факторы как эффективность утилизации субстрата и мобильность популяции?
- В каких условиях больший выигрыш для приспособленности популяции будет давать более эффективная утилизация субстрата, а в каких – более высокая интенсивность миграции?

Сравнение преимуществ эффективного метаболизма и мобильности в условиях нормальной подкормки

Рисунок 2.7.11 показывает влияние таких параметров как эффективность утилизации субстрата и мобильность на приспособленность соответствующей популяции в условиях нормальной подкормки (при увеличении концентрации субстрата подкормкой на 10^{-3} моль/л).

При одинаковой мобильности, но разной эффективности утилизации субстрата (см. рис. 2.7.11 а) численность оседлого в среднем в 2.376 раз выше, чем у кочевника, в данном случае не имеющего никаких преимуществ. При одинаковой эффективности утилизации, но разной мобильности (разница в интенсивности миграции составляет 1% – см. рис. 2.7.11 б) в системе преобладает кочевник, извлекая пользу из своей способности к более интенсивной миграции.

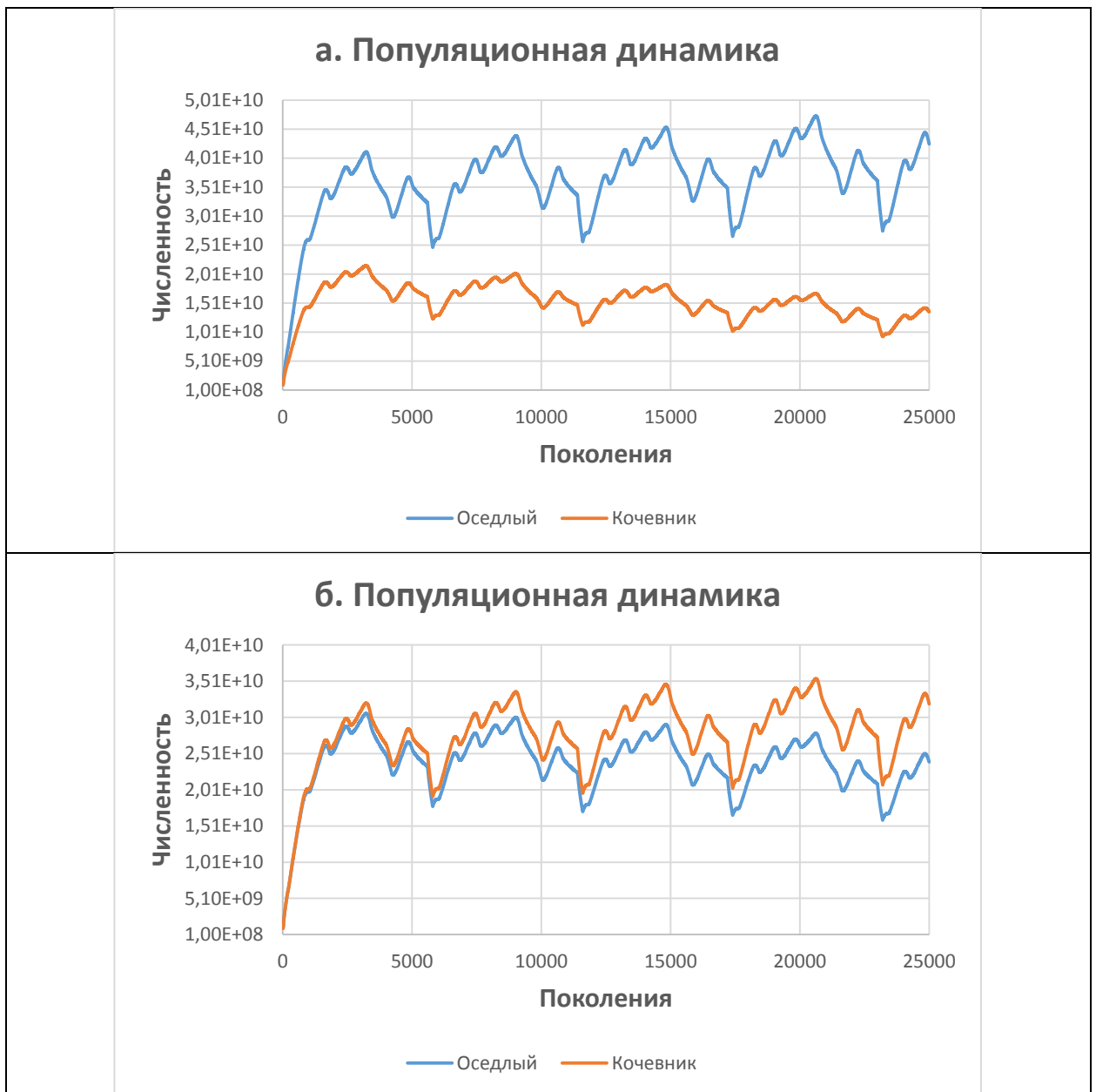


Рисунок 2.7.11. Популяционные динамики. а – одинаковые мобильности, разные эффективности утилизации (1 у кочевника и 10 у оседлого). б – разные мобильности (0.01 у оседлого и 0.0101 у кочевника), одинаковая эффективность утилизации ($N = 1$).

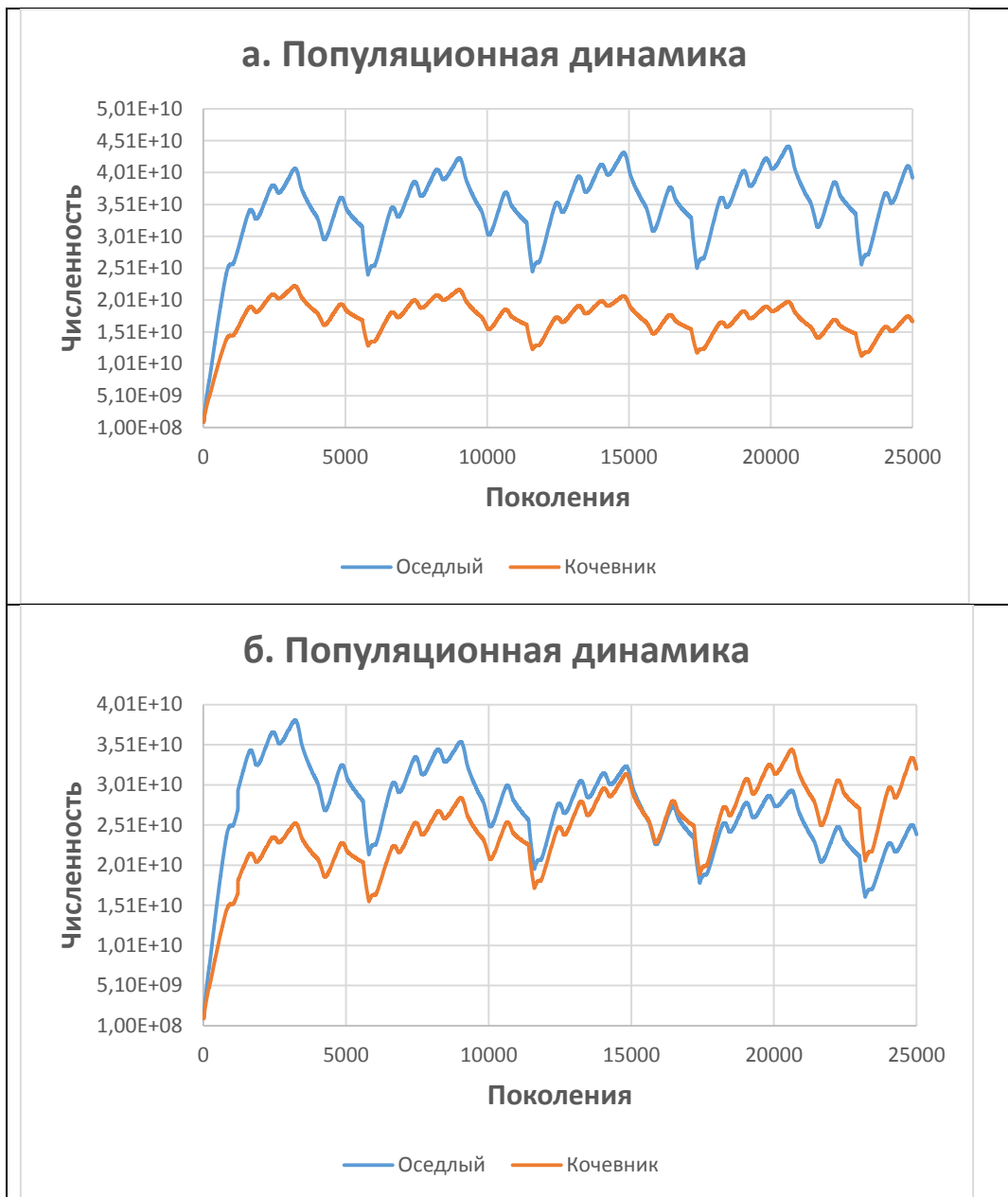


Рисунок 2.7.12. Популяционная динамика для случаев: а – интенсивность миграции у кочевника на 1% выше, чем у оседлого, а эффективность утилизации субстрата N составляет 1 у кочевника и 10 у оседлого; б – интенсивность миграции у кочевника на 5% выше, чем у оседлого, а эффективность утилизации субстрата N составляет 1 у кочевника и 10 у оседлого.

Если мы повышаем эффективность утилизации субстрата у оседлой популяции до $N=10$, то мы снова наблюдаем преобладание оседлых в системе, несмотря на миграционное преимущество кочевника (см. рис. 2.7.12 а). Однако если мы увеличим долю активно мигрирующей популяции у

кочевника с 0.0101 до 0.0105, к 15000 итерации численности обеих популяций сравниваются, а затем кочевник начинает преобладать в системе (рис. 2.7.12 б).

Дальнейшее увеличение параметра эффективности утилизации субстрата способно замедлить процесс захвата кочевником лидирующей позиции в системе, но мобильность кочевника влияет на приспособленность значительно сильнее.

Сравнение преимуществ эффективного метаболизма и мобильности в условиях скудной подкормки

Аналогичным образом был проведён анализ для пессимальных условий со скудным поступлением неспецифического субстрата в среду (при увеличении концентрации субстрата подкормкой на $3 \cdot 10^{-5}$ моль/л).

При одинаковой мобильности, но разной эффективности утилизации субстрата (см. рис. 2.7.13 а), численность оседлого в среднем в 1.156 раз выше, чем у кочевника, в данном случае не имеющего никаких преимуществ. При более ещё скудной подкормке, увеличивающей концентрацию субстрата на 10^{-5} моль/л и менее, в случае одинаковой мобильности более высокое значение параметра эффективности утилизации субстрата не даёт каких-либо преимуществ – графики численностей обеих популяций неразличимы.

При одинаковой эффективности утилизации, но разной мобильности (разница в интенсивности миграции составляет 1% – см. рис. 2.7.13 б), начиная с некоторого момента, в системе преобладает кочевник, извлекая пользу из своей способности к более интенсивной миграции. Интересно, что феноменологически эта популяционно-динамическая картина практически симметрична случаю, представленному на рис. 2.7.13 а.

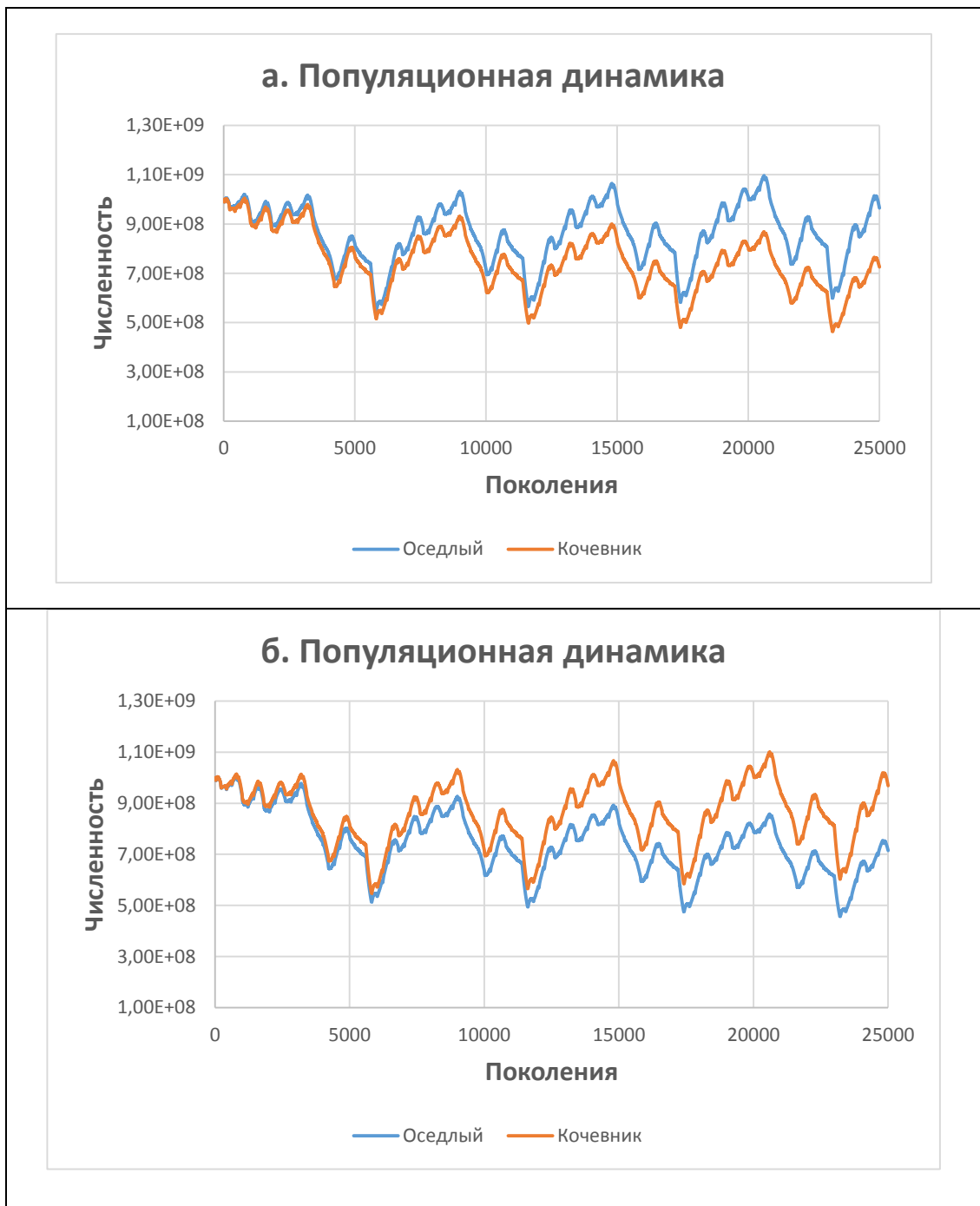


Рисунок 2.7.13. Популяционные динамики. а – одинаковые мобильности, разные эффективности утилизации (1 и 10 соответственно). б – разные мобильности (0.01 и 0.0101 соответственно), одинаковая эффективность утилизации ($N = 1$).

Если мы повышаем эффективность утилизации субстрата у оседлой популяции до $N=10$, то графики численностей обеих популяций становятся практически неразличимыми (см. рис. 2.7.14 а), т.е. преимущество повышенной эффективности утилизации у клеток оседлой популяции

уравновешивает преимущество кочевника в мобильности. Однако если мы увеличим долю активно мигрирующей популяции у кочевника с 0.0101 до 0.0105, это приведёт, начиная с некоторого момента, к уверенному преобладанию кочевника в системе (см. Рис. 2.7.14 б).

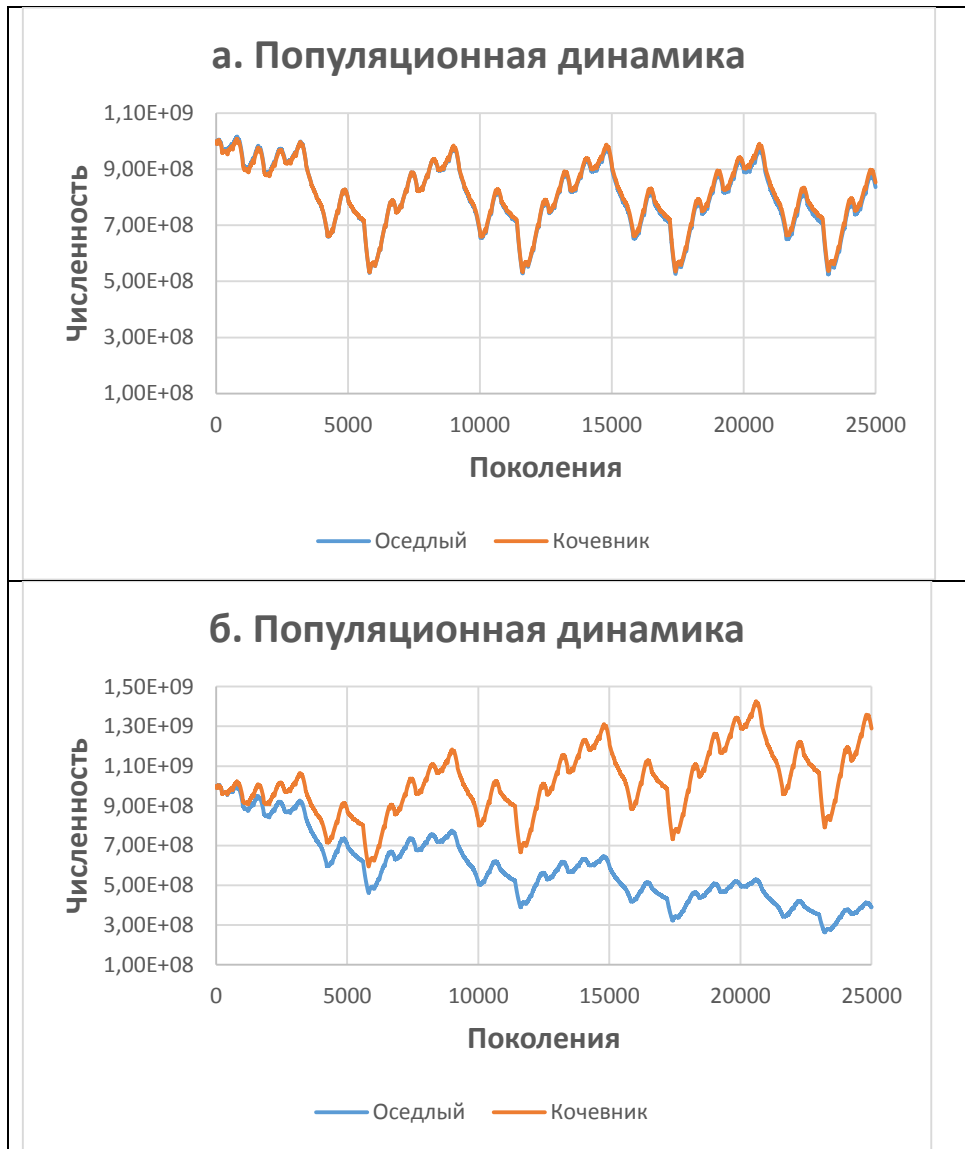


Рисунок 2.7.14. а) Популяционная динамика для случая, когда интенсивность миграции у кочевника на 1% выше, чем у оседлого, а эффективность утилизации субстрата N составляет 1 у кочевника и 10 у оседлого. б) Популяционная динамика для случая, когда интенсивность миграции у кочевника на 5% выше, чем у оседлого, а эффективность утилизации субстрата N составляет 1 у кочевника и 10 у оседлого.

Таким образом в условиях дефицита ресурса повышенная мобильность кочевника даёт более значимое преимущество в приспособленности, чем высокая эффективность утилизации субстрата у оседлого, по сравнению с более ресурсно-богатыми средами, где кочевнику требуется значительно большее время, чтобы занять лидирующее положение в системе (см. Таблицу 2.7.11).

Таблица 2.7.11. Сравнение соотношения средних по времени приспособленностей: $\frac{\overline{fitness}_{оседлого}}{\overline{fitness}_{кочевника}}$.

	Нормальная подкормка	Скудная подкормка
Одинаковая мобильность, преимущество оседлого в эффективности метаболизма ($\frac{N_{оседлого}}{N_{кочевника}} = 10$)	1.035466065	1.0109135
Одинаковая эффективность метаболизма, преимущество кочевника в мобильности	0.997856244	0.98892584
Преимущество оседлого в эффективности метаболизма ($\frac{N_{оседлого}}{N_{кочевника}} = 10$), преимущество кочевника в мобильности (1%)	1.031018076	1.002195675
Преимущество оседлого в эффективности метаболизма ($\frac{N_{оседлого}}{N_{кочевника}} = 10$), серьёзное (5%) преимущество кочевника в мобильности	1.026115176	0.994033142

Таким образом, мы показали, что даже при 5% увеличении доли активно мигрирующей субпопуляции (0.0105 у кочевника против 0.01 у оседлого) наблюдается преобладание численности кочевника и постепенное вымирание оседлого. Однако, этот процесс идёт значительно медленнее в том случае, если оседлый эффективнее утилизирует субстрат, чем кочевник. Преимущество

кочевой популяции выше в условиях ограниченности субстрата, в то время как преимущество эффективной утилизации субстрата, которым обладают клетки оседлой популяции, более явно проявляет себя в условиях обилия субстрата.

Кроме того, в данном исследовании была также рассмотрена модель формирования популяций подвижных микроорганизмов с различными адаптационными стратегиями в периодически изменяющейся среде.

По оценкам различные морские бактерии затрачивают на активное передвижение ~2-50% своих энергетических бюджетов (Stocker, Seymour, 2012). При этом различные виды сильно варьируют по степени подвижности. Однако остаётся не совсем ясным, почему сохраняется такой разброс, и как это может соотноситься с эффективностью метаболизма. В данном разделе работы мы исследовали, какие различные адаптационные стратегии вытекают из при эволюции признаков, связанных с миграцией.

Была рассмотрена следующая модель: одна генетически полиморфная популяция обитает в протяжённой стоячей 1D-среде длиной 1000 ячеек (рис. 2.7.15). По среде бежит субстратная волна, которая отражается, доходя до конца, и бежит в обратном направлении (аналог приливов и отливов или прибрежных волн, приносящих питательные вещества).

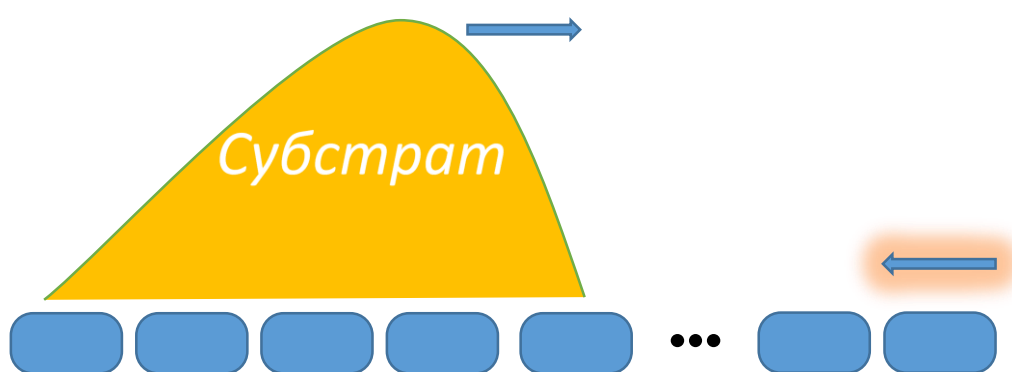


Рисунок 2.7.15. Схема моделируемой системы.

Мы рассмотрели однородное начальное распределение клеток, соответствующее появлению периодического источника питательных веществ

в среде обитания сложившейся колонии клеток. Пик «субстратной волны» проходил 1 ячейку за 1 поколение.

Для учёта энергозатрат на активное перемещение мы использовали модель штрафа за миграцию, использующую два гена. Аллель гена миграционной ригидности выступала в качестве фиксированного параметра, а энергетическая стоимость миграции варьировалась так, чтобы при разных её значениях получались субпопуляции, характеризующиеся разным значением штрафа за миграцию. Если запас энергии клетки оказывался ниже, чем требуется для «уплаты штрафа», перемещения не происходило.

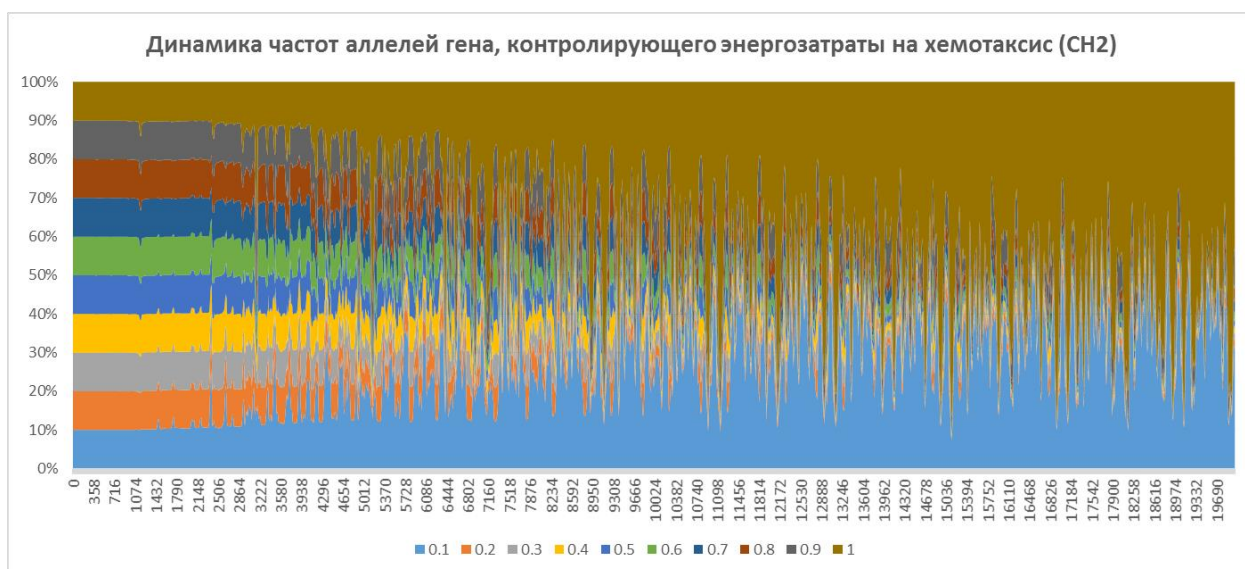


Рисунок 2.7.16. Динамика частот аллелей гена, контролирующего энергозатраты на хемотаксис (энергетическая стоимость миграции). Быстрая волна и однородное распределение клеток. Промежуточные формы уступают место крайним значениям, соответствующих сверхмобильным мигрантам и «домоседам», стоимость акта хемотаксиса у которых наибольшая.

Исследуя модели с полиморфизмом по гену стоимости миграции, из глобальной динамики аллелей мы установили, что при однородном начальном распределении клеток существуют две адаптивные стратегии: либо снижать стоимость хемотаксиса и бегать за субстратной волной, либо повышать её и сидеть, не тратя энергию на миграцию (рис. 2.7.16).

Таким образом, в условиях изменчивой по субстрату среды обитания и однородного начального распределения клеток действует дизруптивный отбор по признаку энергетической стоимости миграции.

Проведенное исследование конкуренции популяций подвижных гаплоидных организмов с различными адаптационными стратегиями в периодически изменяющейся среде показало, что даже при 5% увеличении доли активно мигрирующей субпопуляции (0.0105 у кочевника против 0.01 у оседлого) наблюдается преобладание численности кочевника и постепенное вымирание оседлого. Однако, этот процесс идёт значительно медленнее в том случае, если оседлый эффективнее утилизирует субстрат, чем кочевник.

Преимущество кочевой популяции выше в условиях ограниченности субстрата, в то время как преимущество эффективной утилизации субстрата, которым обладают клетки оседлой популяции, более явно проявляет себя в условиях обилия субстрата.

В модели формирования популяций подвижных микроорганизмов с различными адаптационными стратегиями, что в условиях изменчивой по субстрату среды обитания и однородного начального распределения клеток действует дизруптивный отбор по признаку энергетической стоимости миграции.

2.7.4. Моделирование эволюции в системе «прокариотическое сообщество – бактериофаг»

В данной подглаве приведены результаты моделирования изменения сложности геномов нескольких популяций прокариот, объединённых трофическими связями в сообщество в условиях повторяющегося ухудшения/улучшения условий внешней среды под влиянием фаговой инфекции. Принято считать, что именно сообщества прокариот, в которых различные виды объединены трофическими связями, составляли (и составляют до сих пор) основу жизни на Земле (Заварзин, 2004). Кроме продуктов метаболизма клетки в таких сообществах обмениваются генами за счёт таких механизмов горизонтального переноса генов (ГП), как

трансдукция, конъюгация и трансформация (Levin, Bergstrom, 2000). Все три механизма связаны с факультативной ДНК (плазмиды, фаги, эписомы): трансдукция напрямую связана с фагами, а конъюгация и трансформация – косвенно: концентрация свободной ДНК, поглощаемой при трансформации, может резко возрасти из-за фагового лизиса части популяции/сообщества, а конъюгация обусловлена эписомой. Истинные масштабы распространения ГП среди прокариот стало возможным оценить только после реализации множества полногеномных проектов (Levin, Bergstrom, 2000). Выявлены случаи, когда именно заимствование экзогенов позволяло выживать в условиях, в норме неперемесивных (Nesbo et al., 2001). О систематически неслучайном характере ГП в таксонах прокариот свидетельствует неслучайное расположение экзогенов в геномах (Tamames, 2001) и недавно открытые интегроны и суперинтегроны – естественные векторы для клонирования экзогенов в геномах (Puina, 2006). Считается, что в настоящее время неизвестны сколь-нибудь значительные таксоны прокариот, не затронутые ГП (Amann et al., 1995). В результате этого филогенез прокариот принято изображать в виде сложной сети (Gogarten, Townsend, 2005; Soucy et al., 2015) или «кольца жизни» (Rivera, Lake, 2004), от которого ветками отходят как обособленные таксоны прокариот, так и эукариоты.

Уникальная особенность эукариот – резкое «раздувание» геномов (как по числу генов, так и по количеству некодирующей ДНК). В эволюции всех прочих филумов геном оставался «экономным»: его размеры уменьшались или колебались в определенных рамках (Carroll, 2001; Taft et al., 2007). Эволюционные преимущества «экономного генома» были теоретически обоснованы еще в 70-х годов XX века (Эйген, Шустер, 1982) и в настоящее время имеют многочисленные экспериментальные подтверждения, в том числе на полногеномном уровне (Колчанов et al., 2008). Однако значительная часть этого экспериментального материала была получена на прокариотах, культивируемых в лабораторных условиях. При этом культивируемость

прокариот зачастую обусловлена их способностью существовать в сравнительно простых микробных сообществах (или даже в виде отдельных популяций). То есть, фактически эволюционные преимущества «экономного генома» обоснованы только на популяционном уровне. Вместе с тем, подавляющее большинство прокариот некультивируемо (Amann et al., 1995), в том числе и потому, что в природе они существуют в составе сообществ сложной структуры (биоплёнки, маты, строматолиты и т.д.) (Заварзин, 2004). Данная подглава посвящена моделированию эволюции микробных сообществ в условиях нестабильного поступления ресурса на фоне фаговой инфекции и ГП.

Исходная модель описывала эволюцию сообщества, которое в начале расчётов представляло собой трофическое кольцо, где каждая из трёх популяций потребляла ровно один специфический субстрат и производила ровно один специфический продукт. Пример такого сообщества приведён на рисунке 2.7.17 (слева). В ходе эволюции с вероятностью 10^{-7} на поколение на клетку, моделировались процессы как переноса генов между клетками разных популяций, так и потерь генов. В результате появлялись новые популяции, комбинирующие разные сочетания утилизируемых и синтезируемых субстратов (рис. 2.7.17 справа).

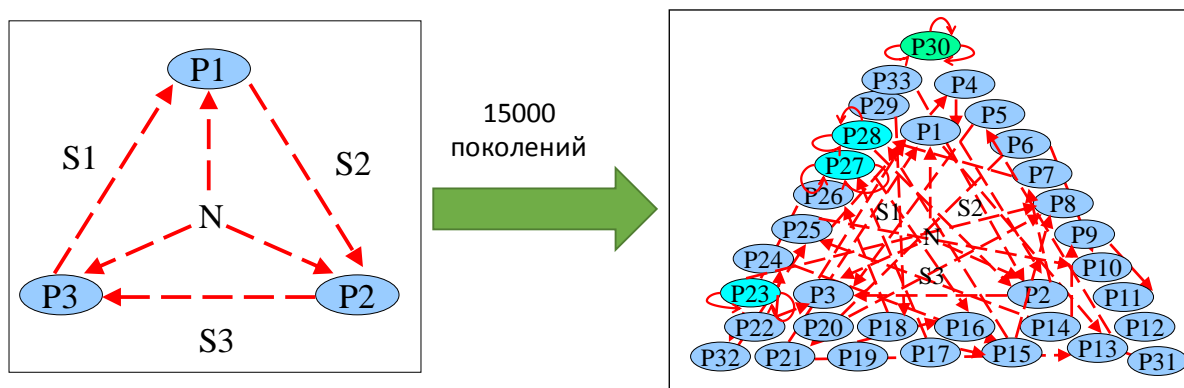


Рисунок 2.7.17. Схема трофического кольца, состоящего из трёх популяций, в начальный момент времени (слева) и через 15000 поколений (справа).

Формула изменения численности популяций сообщества представляла собой модифицированную формулу трофической стратегии:

$$F(n_0, \vec{S}, r_0, \vec{C}, P, G) = T(n_0, \vec{S}, r_0, \vec{C}, P) - \frac{(G/k_1)^n}{1 + (G^n/(k_1^{-n} - 1))} P \quad (2.7.8)$$

где $T(n_0, \vec{S}, r_0, \vec{C}, P)$ – стандартная трофическая стратегия популяции (в данном случае компенсаторная стратегия (ур. 2.1.1)); \vec{S} – вектор количественных значений потребляемых специфических субстратов; n_0 – количество неспецифического субстрата, потребляемого клетками популяции; r_0 – генетически предопределённая скорость утилизации неспецифического субстрата; P – численность популяции; G – количество генов в геноме; k_1 – пороговое значение, определяющее возрастание штрафа за длину генома; n – степень нелинейности штрафа за длину генома.

Моделирование показало, что при умеренных значениях штрафа на размер генома (значения от 0.01 до 0.05), а также в пессимальных условиях среды (невысокая концентрация неспецифического субстрата в притоке и низкая эффективность синтеза специфического субстрата клетками) подавляющее преимущество по биомассе получают популяции с метаболически полным или почти полным геномом. Такие популяции потребляют и продуцируют максимальное (или почти максимальное) количество специфических субстратов в данной трофической системе. Метаболически бедные популяции в долгосрочной перспективе они либо вытесняли прочие популяции из трофической системы, либо выживали за счет богатых, становясь их комменсалами. Их малая численность не позволяла существенно снизить жизнеспособность хозяев: предельная численность каждой из метаболически бедных популяций колебалась в районе нескольких сотен клеток в среде.

Во второй модели в среде изначально находилась единственная метаболически полная популяция, но штраф на размер генома был более

высокий (0.1-0.25). После появления метаболически неполных популяций делеции шли параллельно с горизонтальным переносом генов. Перенос мог происходить многократно, возмещая делеции. Частоты делеций и горизонтального переноса задавались аналогично предыдущей модели. Протоком в среду моделировались субкомфортные условия (10^{-3} mM концентрация неспецифического субстрата в притоке) и комфортные условия (дополнительно вводились в среду извне специфические субстраты до концентрации 10^{-3} mM для каждого субстрата). Было показано, что и в субкомфортных, и комфортных условиях среды наблюдается стойкая тенденция к редукции генома. Наиболее примитивные популяции, имеющие всего два гена утилизации: один для неспецифического, другой для специфического субстрата, вытесняли остальные популяции. При этом в субкомфортных условиях примитивная популяция, вытеснив все остальные популяции, погибает сама, израсходовав необходимый ей специфический субстрат (рис. 2.7.18а). В отличие от этого, в комфортных условиях примитивная популяция продолжает жить после вытеснения всех остальных популяций: уровень специфического субстрата поддерживается искусственно (рис. 2.7.18б).

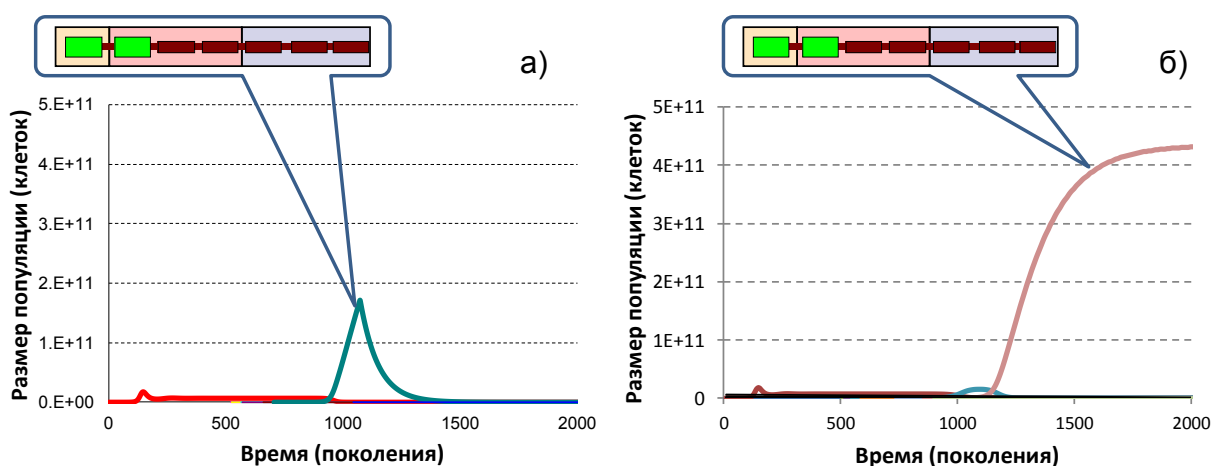


Рисунок 2.7.18. Редукционная эволюция сообщества, исходно состоявшего из одной популяции клеток с полным метаболизмом: а) в субкомфортных условиях среды, примитивная популяция гибнет спустя

примерно 500 поколений после вытеснения всех остальных членов сообщества; б) в комфортных условиях среды, примитивная популяция выживает после вытеснения остальных членов сообщества.

Влияние фаговой инфекции на эволюционные тренды

Для исследования влияния фаговой инфекции на возможные эволюционные тенденции были построены модели заражения прокариотического сообщества. В качестве базовой модели была взята первая модель, упомянутая в подглаве, описывающая эволюцию сообщества, исходно представляющего собой трофическое кольцо из трёх популяций. В процессе расчёта модели также стохастически моделировались процессы горизонтального переноса и потери генов. Добавление в сообщество фаговой популяции приводило к заражению всех популяций, при этом доля заражённых клеток зависела как от концентрации фагов в среде, так и от концентрации прокариотических клеток. Вероятность того, что заражённая популяция или её часть будет развиваться по литическому сценарию, зависела от того, в каких условиях на момент заражения находилась её родительская популяция: если она устойчиво росла, то все заражённые клетки этой популяции лизировались с образованием новой порции фагов; и наоборот, если популяция находилась в пессимальных условиях и снижала свою численность, то заражённые клетки (все или часть) переключались на лизогенный цикл, становясь, с одной стороны носителями генов фага, а с другой стороны, иммунными к заражению этим видом фага. В результате этого, заражение существенно изменяло эволюционную динамику сообщества, сдерживая рост или даже уничтожая динамично растущие популяции, и тем самым поддерживая менее конкурентоспособные в этих условиях популяции. Например, в ряде численных экспериментов было показано, что в пессимальных условиях среды могут выживать популяции, далёкие от метаболической полноты генома (рис. 2.7.19). Это контрастирует с

отмеченной нами ранее тенденцией к усложнению генома в подобных условиях (Lashin et al., 2010).

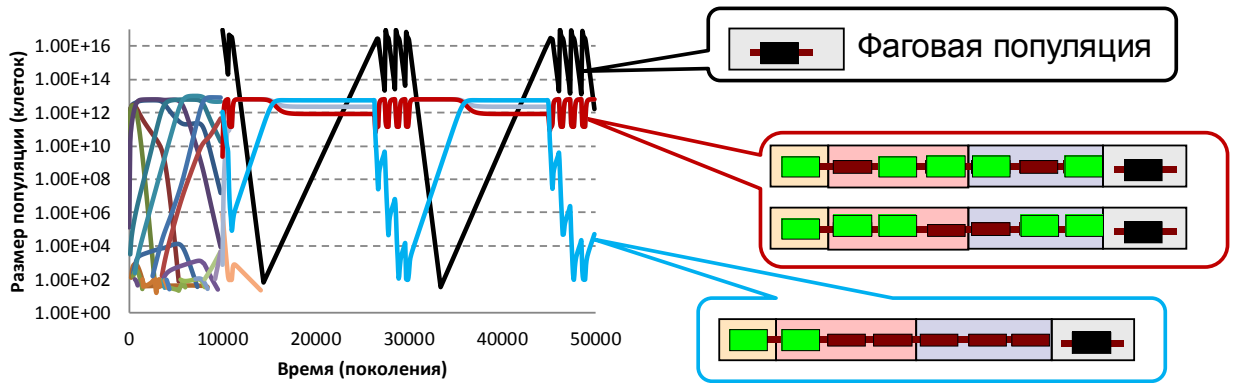


Рисунок 2.7.19. Динамика численности популяций сообщества в пессимальных условиях среды. Заражение популяции происходит на 10000-м поколении. До этого времени в системе происходит появление видов благодаря горизонтальному переносу и потере генов. Спустя короткий промежуток времени после заражения вымирает большая часть популяций. Оставшиеся популяции не обладают метаболически полным геномом, более того, в числе выживших даже популяция с крайне примитивным геномом (нижняя схема на рисунке). На схемах геномов черный прямоугольник соответствует фаговому гену, т.е. все выжившие популяции являются носителями фага.

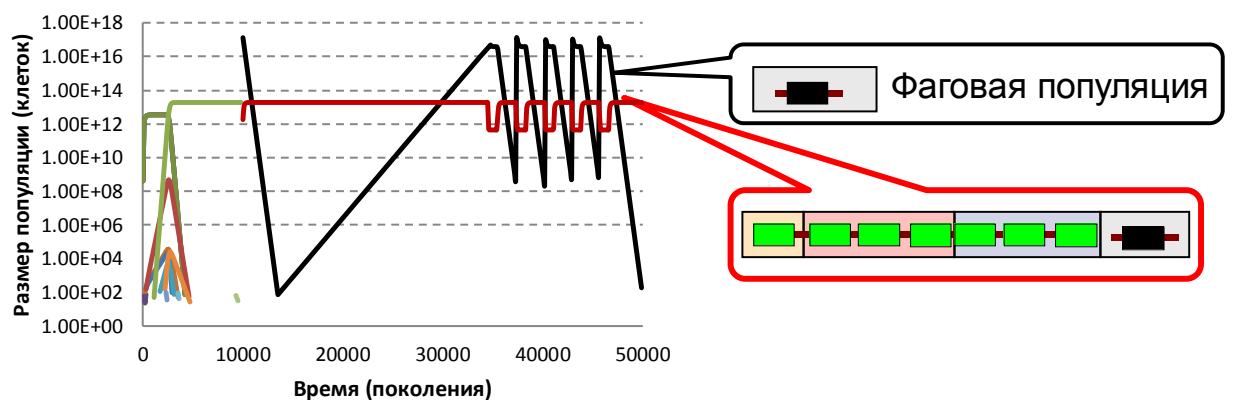


Рисунок 2.7.20. Динамика численности популяций сообщества в оптимальных условиях среды. Заражение популяции происходит на 10000-м поколении. До этого времени в системе происходит появление видов

благодаря горизонтальному переносу и потере генов. Спустя короткий промежуток времени после заражения вымирают все популяции, кроме популяции с метаболически полным геномом, инфицированной фагом.

Возможная смена эволюционной тенденции в случае заражения фагом была обнаружена также и для сообществ в оптимальных условиях среды. На рисунке 2.7.20 показано выживание популяции с метаболически полным геномом, наименее приспособленной популяции сообщества до заражения. Периодический характер кривых численности популяций, как прокариотических, так и фаговых, объясняется спонтанным переходом части заражённых клеток в литическую форму, с последующей гибелью и появлением в результате лизиса новых фагов.

Отметим, что данные результаты численного моделирования носят стохастический характер. В ряде численных экспериментов заражение приводило к гибели всего сообщества, в других случаях сообщество гибло до заражения из-за быстрого роста излишне примитивных популяций (ср. с рис. 2.7.18а). Смена эволюционной тенденции также происходила не всегда. Таким образом, данные результаты показывают, что заражение сообщества фагом может изменить эволюционную тенденцию сообщества.

Влияние пространственной гетерогенности среды обитания на эволюцию системы «прокариотическое сообщество – бактериофаг»

Описанная выше модель с полным перемешиванием была расширена на 2-мерный случай с целью изучения влияния факторов пространственной неоднородности на эволюцию подобных систем. Организация пространственно-распределённой среды приведена на рисунке 2.7.21.

В различных численных экспериментах мы варьировали время и место инвазии фагов в сообщество, а также возможность клеток хемотактировать. Моделирование показало, что фаговая инфекция часто замедляет процессы видообразования и стабилизирует структуру сообщество (рис. 2.7.22). Из

рисунка видно, что инфицирование фагами останавливают (или как минимум замедляет) эволюцию сообщества.

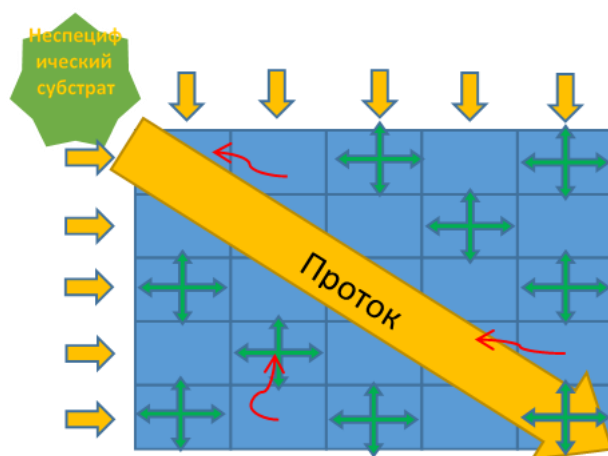


Рисунок 2.7.21. Схема пространственной организации среды обитания в модели «прокариотическое сообщество – бактериофаг» (адаптировано из [xxx]).

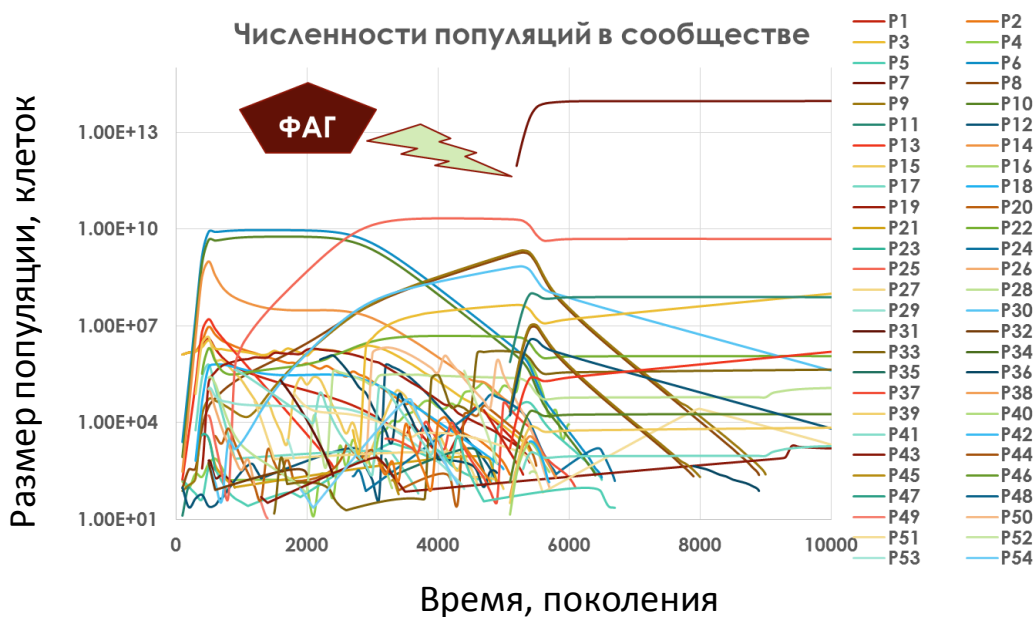


Рисунок 2.7.22. Типичная популяционная динамика при добавлении фага в среду на 5000 поколениях.

Отметим, что мы говорим о видообразовании с точки зрения вариаций набора генов, участвующих в метаболических сетях, а эволюция механизмов иммунитета не входит в рамки данного исследования. Это объясняет, почему мы наблюдаем снижение скорости видообразования в условиях фаговой

инвазии, в то время как в других исследованиях (Koskella, Brockhurst, 2014; Koskella, Parr, 2015), в которых акцент делается на эволюции стратегий защиты и противодействия, указывается на возможность увеличения скорости эволюции фагов и бактерий.

Этот эффект был оценён с помощью предложенного нами в (Klimenko et al., 2016) Индекса скорости видообразования (англ. Speciation rate index – SRI), который был рассчитывается по следующей формуле:

$$SRI = \frac{\| \{p\} \text{ - вновь возникшая жизнеспособная популяция} \|}{time_span} \quad (2.7.9)$$

где *time_span* – промежуток времени, на протяжении которого оценивается скорость видообразования; популяция считается жизнеспособной, если она просуществовала большее число поколений, чем заданный порог, который равнялся 500. Чем выше SRI, тем более интенсивные процессы видообразования происходят в данный промежуток времени. И наоборот, чем ниже SRI, тем меньше жизнеспособных популяций появляется в данное время.

Клетки сообщества не обладают хемотаксической способностью

Первоначально мы оценили SRI во всех пространственных ячейках среды обитания в отсутствие фаговой инвазии (рис. 2.7.23). Рисунок показывает, что скорость видообразования растёт по мере удаления от ячейки поступления неспецифического субстрата (вверху). Это можно объяснить тем, что в верхних ячейках существует несколько популяций, которые эффективно используют неспецифический субстрат. Эти популяции имеют доминирующую биомассу. В то же время, новые экологические ниши, связанные с высокими концентрациями специфических субстратов, возникают в средних и верхних ячейках благодаря их транспорту с протоком. Это открывает возможности для фиксации новых видов, специализирующихся на утилизации этих специфических субстратов. Иными словами, в дальних,

нижних ячейках более высокое биоразнообразие ассоциируется с низкой биомассой (рис. 2.7.24).

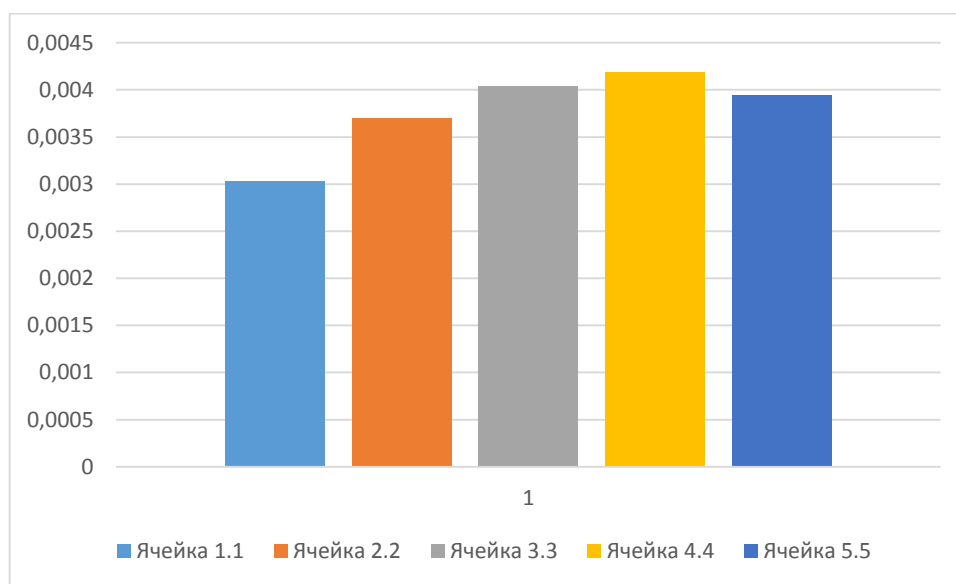


Рисунок 2.7.23. Значения SRI, посчитанные в течении 5000 поколений.

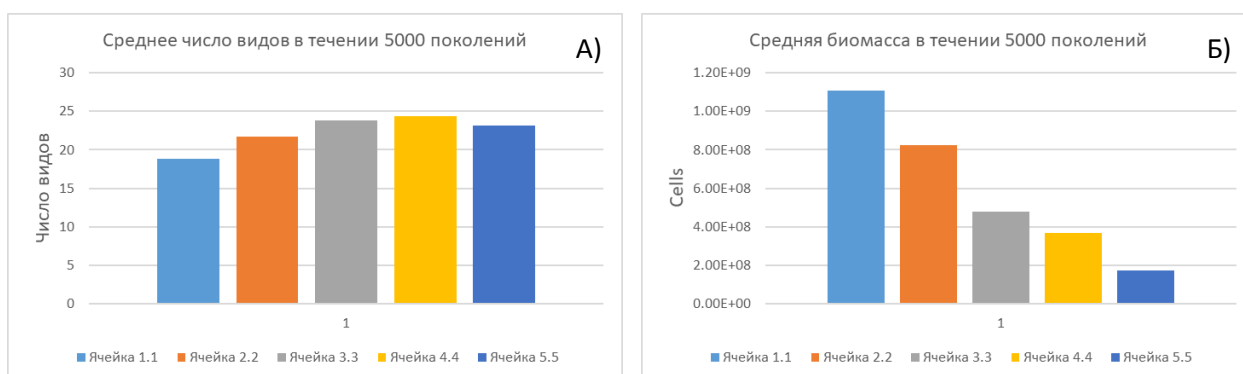


Рисунок 2.7.24. А) Среднее число видов в течении 5000 поколений; Б) Средняя биомасса в течении 5000 поколений.

После этого мы проанализировали, как фаговая инвазия влияет на скорость видообразования и видовое богатство сообщества. Полученные результаты показывают зависимость скорости видообразования от параметров инвазии (табл. 2.7.12).

Таблица 2.7.12. Средние значения SRI до и после фаговой инвазии (низкий уровень фагового всплеска) в различных местах обитания. Первые 1000 поколений не учитывались, так как существует погрешность, связанная с высоким начальным видообразованием.

	Ранняя инвазия (1-е поколение)		Средняя инвазия (5000-е поколение)		Поздняя инвазия (6600-е поколение)	
	SRI до	SRI после	SRI до	SRI после	SRI до	SRI после
ячейка (1,1)	-	0.000856	0.00685	0.00052	0.007143	0.000529
	SRI до	SRI после	SRI до	SRI после	SRI до	SRI после
ячейка (5,5)	-	0.000744	0.007275	0.0004	0.007946	0.000206
	SRI до	SRI после	SRI до	SRI после	SRI до	SRI после
ячейка (3,3)	-	0.000578	0.00825	0.0006	0.006964	0.000412
	SRI до	SRI после	SRI до	SRI после	SRI до	SRI после

Из таблицы видно, что инвазия фагов приводит к снижению скорости видообразования более чем на порядок. Легко заметить, что зависимость скорости видообразования от времени зависит от расстояния между ячейками инфицирования и поступления неспецифического субстрата:

1. В случае инвазии в ранние сроки (1-е поколение) наблюдается обратная зависимость – чем дальше место инвазии, тем меньше скорость видообразования после инвазии (рис. 2.7.25).

2. В случае инвазии в 5000-м поколении наименьшая скорость видообразования наблюдается, когда фаг проникает в центральную ячейку (3,3) среды обитания. Примечательно, что если фаг появляется в системе в самой нижней и дальней нижней ячейке (5,5), то средний показатель SRI становится выше по сравнению с самой верхней ячейкой (1,1).

3. В случае позднего заражения (6600-е поколение) результаты выглядят аналогично с предыдущим случаем, за исключением того, что средний SRI ниже, когда фаг вторгается в центральные и нижние узлы.

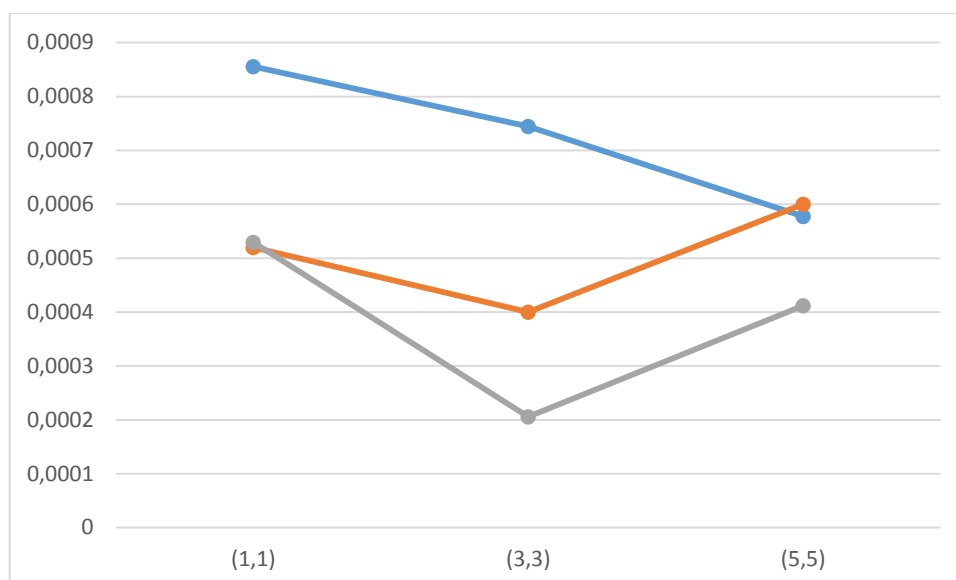


Рисунок 2.7.25. Зависимость SRI от времени инвазии и начальной локализации инвазии (по оси X: ячейки (1,1), (3,3) и (5,5)). Синий график: ранняя инвазия (1-е поколение); оранжевый график: средняя инвазия (5000-е поколение); серый: поздняя инвазия (6600-е поколение). Ось Y – средний SRI, рассчитанный за всё время моделирования эволюции.

Как было сказано выше, удаленные ячейки характеризуются более высокими концентрациями специфических субстратов. Таким образом, эти результаты согласуются с упоминаемым в литературе [41] предположением о том, что вирусный эффект, вероятно, больше в эвтрофных водах, чем в олиготрофных.

Из таблицы 2.7.13 видно, что в случае инвазии в средние сроки, средний SRI увеличивается от верхних (1,1) ячеек (богатых неспецифическим субстратом) к нижним (5,5) (бедным неспецифическим субстратом). Фаговая инвазия приводит к резкому сокращению видового богатства сообщества. Полученные результаты моделирования показывают, что SRI и видовое богатство выше в отсутствие фагов и при наличии большого количества экологических лицензий.

Таблица 2.7.13. Среднее значение SRI, рассчитанное за все время моделирования, относительно начального места инвазии фага.

	Ранняя инвазия (1-е поколение)	Средняя инвазия (5000-е поколение)	Поздняя инвазия (6600-е поколение)
ячейка (1,1)	0.0013	0.00561	0.0066
ячейка (3,3)	0.00117	0.00587	0.0072
ячейка (5,5)	0.00128	0.00633	0.0068

Клетки сообщества обладают хемотаксической способностью

В случаях, когда клетки обладают способностью к хемотаксису, пространственная стратификация сообщества по биомассе более выражена. Мы предположили, что это связано с наблюдаемым накоплением специфических субстратов в ячейках, расположенных на удалении от источника поступления неспецифического субстрата. Тем не менее, распределение видового богатства и SRI в целом выглядят аналогично соответствующим распределениям, полученным для предыдущего случая. Однако пространственные различия здесь более очевидны (рис. 2.7.26). С с другой стороны, общее число видов, а также SRI в случае, когда клетки обладают хемотаксической способностью, определено ниже по сравнению с предыдущим случаем (хемотаксис выключен). Это хорошо согласуется с биологическими данными, согласно которым миграция препятствует видообразованию, в то время как изоляция способствует ему (France et al., 1992; Gavrilets et al., 2000; Won et al., 2002).

На рисунке 2.7.27 показано, что инвазия фага в дальнюю нижнюю ячейку (5,5) приводит к росту скорости видообразования для случаев как раннего, так и позднего заражения. Это связано с тем, что фаговая инфекция не может закрепиться в среде обитания при инвазии в такие ячейки. Однако, если

инвазия происходила в верхних ячейках, видовое богатство было выше в нижних.

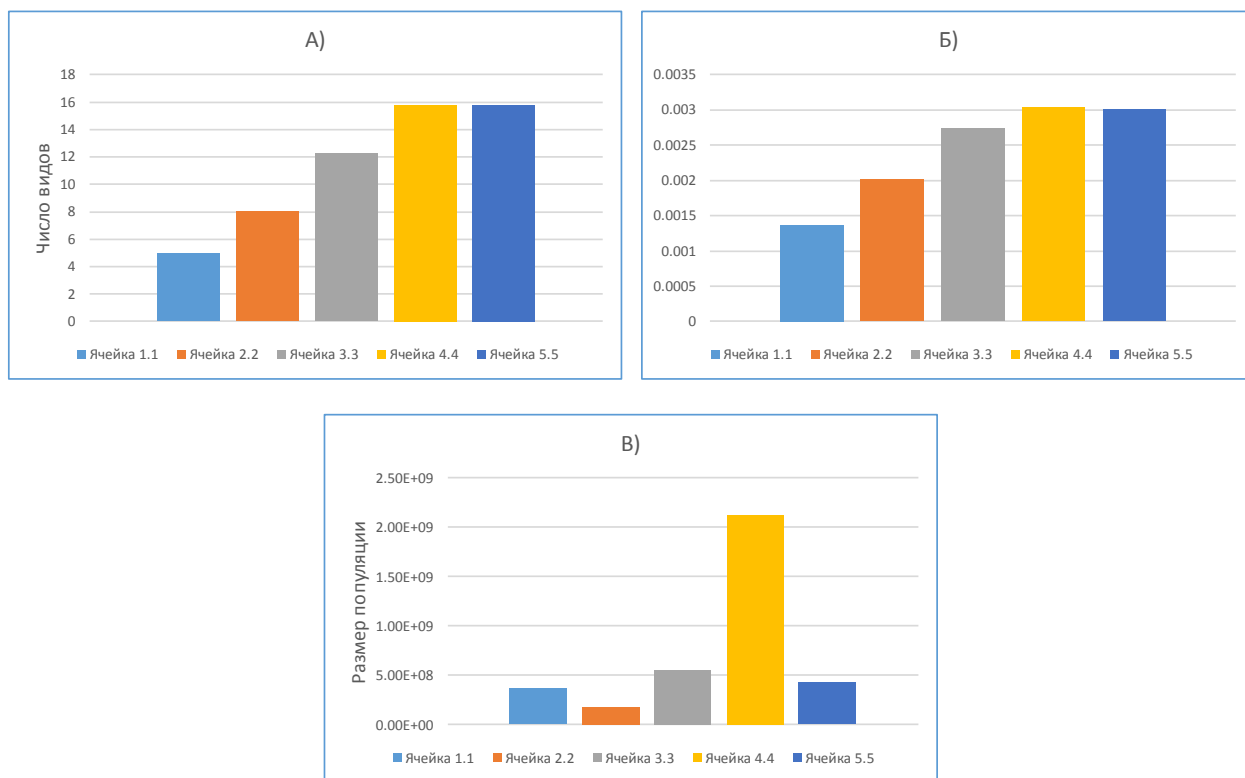


Рисунок 2.7.26. Распределения средних значений (А) видового богатства; (Б) SRI; (В) общей биомассы в пространственных ячейках. Значения рассчитаны для первых 5000 поколений.

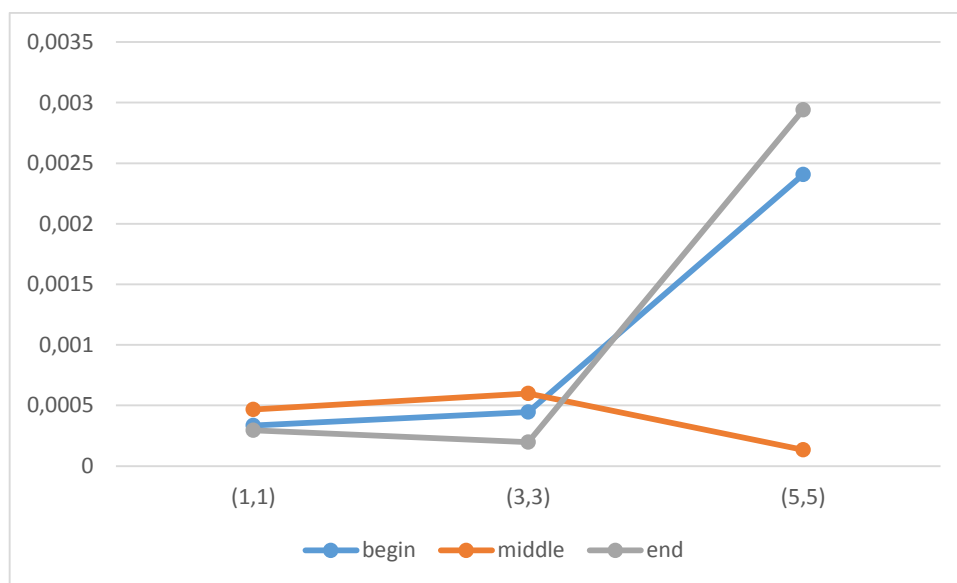


Рисунок 2.7.27. Зависимость SRI от времени инвазии и начальной локализации инвазии: хемотаксис включён (по оси X: ячейки (1,1), (3,3) и

(5,5)). Синий график: ранняя инвазия (1-е поколение); оранжевый график: средняя инвазия (5000-е поколение); серый: поздняя инвазия (6600-е поколение). Ось Y – средний SRI, рассчитанный за всё время моделирования эволюции.

Таким образом, мы показали, что бактериофаги могут выступать в качестве сдерживающих факторов эволюции микробных сообществ. Инфицирование фагами снижает скорость видообразования более чем на порядок, поскольку усиленный отбор блокирует возникновение новых жизнеспособных популяций/видов, которые могли бы занять потенциальные экологические ниши. В то же время фаги действуют как стабилизирующий фактор, приостанавливающий «излишнее» видообразование, что способствует поддержанию стабильного состояния системы (с точки зрения с точки зрения количества видов). Кроме этого, мы показали, что зависимость скорости видообразования от расположения места инвазии варьирует в зависимости от времени инвазии. Оказалось, что скорость видообразования ниже в тех случаях, когда инвазия происходит в уже сформированное сообщество (средние и поздние инвазии) в богатые видами ячейки окружающей среды (центральная ячейка (3,3)). Это особенно заметно в случае поздних инвазий.

В случае способности клеток к хемотаксическому движению, наблюдаемые зависимости отличаются. При ранних и поздних инвазиях скорость видообразования остается относительно высокой в тех случаях, когда фаг проникает в дальние нижние ячейки. Это явление требует дополнительных систематических исследований.

2.8. Методика моделирования микробных сообществ, основанная на агентно-ориентированной концепции

Программный комплекс ГЭК, описанный в предыдущих разделах данной главы (2.1-2.6), реализован в концепции супериндивида. Хотя использование данной парадигмы позволяет моделировать микробные сообщества с численностью, соответствующей природным сообществам (10^9 и выше) и,

более того, получать биологически значимые содержательные результаты, подобные методы моделирования имеют довольно существенные ограничения. В частности, такое моделирование может недостаточно полно описывать некоторые процессы, например, сигналинг, чувство кворума и другие процессы, возникающие между отдельными микробными клетками (Oldewurtel et al., 2015; Zöllner et al., 2017). Экспериментальные данные, полученные с помощью новых технологий в микроскопии, проточной цитометрии, микрофлюидике, спектроскопии и др. – свидетельствуют о том, что в ряде случаев микробные сообщества следует рассматривать как существенно дискретные системы, для которых не всегда представляется возможным пользоваться усредненными характеристиками для популяции, и наиболее подходящий метод моделирования для них – агентно-ориентированное моделирование (Hellweger et al., 2016). Поэтому нами было принято решение о разработке агентно-ориентированной методики моделирования микробных сообществ с разрешением до отдельной клетки.

В данном разделе приведены результаты разработки данной методики и её реализации в виде программного комплекса Microcosm (Microbial community simulator). Отметим, однако, что идеологически эта работа является продолжением работ по разработке ПК ГЭК, и многие ключевые концепции построения многоуровневых моделей иерархически организованных биологических систем, предложенные и разработанные нами в предыдущей части работы, были применены и для разработки данной методики. Поэтому в данной подглаве мы будем опускать некоторые детали там, где мы посчитали это допустимым, о чём дополнительно будем сообщать.

Программный комплекс Microcosm представляет собой агентно-ориентированную (на уровне отдельных клеток) систему для моделирования функционирования и эволюции микробного сообщества. Основные объекты этого ПК описаны ниже.

Объект *Cell* (Клетка) описывает микробную клетку, её метаболизм, жизненный цикл и экологическую активность.

В рамках данной модели предполагается, что жизненный цикл клетки состоит из следующих основных стадий: появление новой клетки -> поглощение субстратов и энергии из среды -> метаболизм -> перемещение клетки в пространстве -> деление клетки или её гибель (см. рис. 2.8.1 и пояснения к нему). Каждая из этих стадий (за исключением первой стадии) имеет свою длительность: T_{EAT} , T_{MET} , T_{CHEMO} , $T_{REPRODUCE}$.

Клетка представляется в виде кортежа $Cell = \langle M; G_M; G_T; G_R; T \rangle$, где

- M – набор внутренних метаболитов клетки;
- G_M – генная сеть общего метаболизма клетки;
- G_T – генная сеть транспортных процессов клетки (потребления и секреции метаболитов);
- G_R – генная сеть клеточного деления (репродукция клетки);
- G_C – генная сеть хемотаксиса (активного движения клетки);
- T – характерные времена клеточных процессов, $T = \langle T_{EAT}; T_{MET}; T_{CHEMO}; T_{REPRODUCE} \rangle$

Схематично функции клетки представлены на рис. 2.8.1. В природе за все эти функции отвечают свои генные сети (Колчанов et al., 2013). Т.е. модель выполнения той или иной функции клетки определяется соответствующей генной сетью.

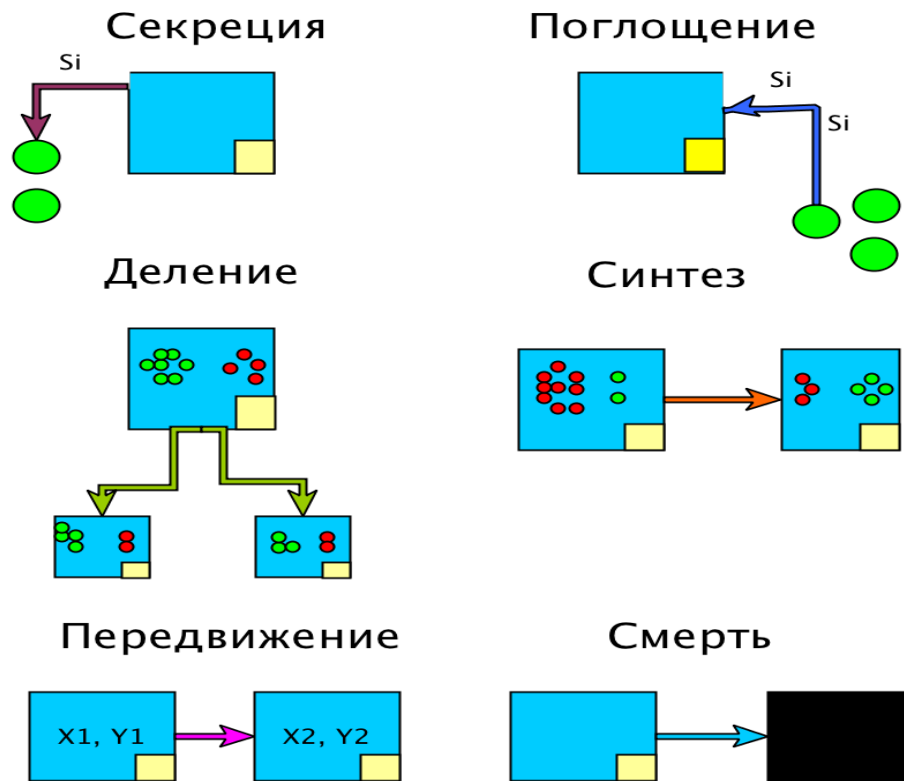


Рисунок 2.8.1. Функции клетки в ПК Microcosm. x_i, y_i – координаты клетки в пространстве; s_i – i -й нутриент. Детальные пояснения в тексте ниже.

2.8.1. Описание внутренних метаболитов клетки

Под метаболитами, как и прежде, будем подразумевать химические вещества как низко- так и высокомолекулярной природы. Набор M внутренних метаболитов представляется в виде множества

$$M = \{m_i\}, i = 1, \dots, N$$

где m_i , в свою очередь, представляет собой пару

$$m_i = \{name_i, concentration_i\}, i = 1, \dots, N$$

где $name_i$ и $concentration_i$, соответственно, название метаболита (строковое значение) и его внутриклеточная концентрация (число).

Не ограничивая общности, мы можем описывать энергию клетки через один или несколько метаболитов.

2.8.2. Описание генной сети общего метаболизма клетки

Генная сеть общего метаболизма клетки представляет собой тройку

$$G_M = \langle S_M, P_M, F_M \rangle, \text{ где}$$

S_M – структурная модель генной сети, записанная в одном из известных формализмов (подробнее см. литературный обзор и разделы 2.1, 2.6.5);

P_M – множество численных параметров генной сети;

F_M – процедура численного моделирования динамики генной сети, изменяющая значения внутриклеточных метаболитов за заданный момент времени T_M ($F_M: M_{T_0} \rightarrow M_{T_M}$).



Рисунок 2.8.2. Использование моделей генных сетей для описания метаболизма клетки в ПК Microcosm. Адаптация идеи, предложенной в ПК ГЭК, для ПК Microcosm.

Строго говоря, в рамках данной модели нас интересуют только концентрации внутриклеточных метаболитов в начальный момент T_0 и в финальный момент времени T_M . Под этими моментами времени мы подразумеваем, соответственно, время начала и окончания стадии метаболизма (тогда $T_M - T_0 = T_{MET}$). Таким образом, если, например, модель

генной сети S_M представлена в виде системы ОДУ, то в качестве F_M может служить любой подходящий метод численного интегрирования задачи Коши (рис. 2.8.2). Более детальное описание моделей генных сетей, записанных в разных формализмах, приведено в литературном обзоре (раздел 1.1.1).

2.8.3. Описание генной сети транспортных процессов клетки

Генная сеть транспортных процессов клетки представляет собой тройку

$$G_T = \langle S_T, P_T, F_T \rangle, \text{ где}$$

S_T – структурная модель транспорта веществ из среды обитания в клетку и из клетки в среду обитания, записанная в одном из известных формализмов (см. литературный обзор, а также разделы 2.1 и 2.2);

P_T – множество численных параметров генной сети;

F_T – процедура численного моделирования транспорта веществ из среды обитания в клетку и из клетки в среду обитания за заданный момент времени $T_T (F_T: \overline{M}_{T_0} \rightarrow \overline{M}_{T_T})$, где

$$\overline{M}_T = M_T \cup S_T$$

M_T и S_T – множества внутриклеточных и средовых (внеклеточных) метаболитов, соответственно в момент времени T .

2.8.4. Описание генной сети клеточного деления

Генная сеть клеточного деления (репродукции клетки) представляет собой двойку

$$G_R = \langle D_R, P_R \rangle, \text{ где}$$

D_R – решающее правило (англ. decision rule) клеточного деления, алгоритм, оценивающий текущие внутреннее (количество накопленных метаболитов, энергетический потенциал и т.п.) и внешнее (наличие/отсутствие соседей, химических сигналов от них и т.п.) состояния клетки, и на основе полученной оценки генерирующий две новых (дочерних) клетки на основе одной (материнской), также уничтожающий последнюю;

P_R – множество численных параметров решающего правила;

2.8.5. Описание генной сети хемотаксиса

Генная сеть хемотаксиса (активного перемещения клетки) представляет собой двойку

$$G_C = \langle D_C, P_C \rangle, \text{ где}$$

D_C – решающее правило хемотаксиса, алгоритм, оценивающий текущие внутреннее (количество накопленных метаболитов, энергетический потенциал и т.п.) и внешнее (наличие/отсутствие соседей, химических сигналов от них и т.п.) состояния клетки, и на основе полученной оценки изменяющий положение клетки в пространстве;

P_C – множество численных параметров решающего правила;

2.8.6. Моделирование жизненного цикла и деления клетки

Как уже было отмечено выше, в текущей модели мы полагаем, что жизненный цикл клетки состоит из следующих основных стадий: появление новой клетки, поглощение субстратов и энергии из среды, метаболизм, перемещение клетки в пространстве и, наконец, деление клетки или её гибель (рис. 2.8.1). Отметим, однако, что в реальной микробной клетке вышеперечисленные стадии могут происходить и в другой последовательности. Ранее в моделях с идеологией супериндивида (ГЭК, см. разделы 2.1-2.2) нами было показано (Лашин, 2010; Лашин et al., 2008), что эта последовательность не играет значимой роли при моделировании эволюции и функционировании микробного сообщества. Однако последние экспериментальные данные о развитии микробных популяций, полученные с помощью современных высокоразрешающих методов (например, микроскопия одиночных клеток, англ. single-cell microscopy), свидетельствуют о том, что появление сложноорганизованных паттернов в такой популяции может быть связано с рассинхронизацией определённых стадий клеточных процессов.

Таким образом, в рамках нашей модели мы решили предусмотреть два варианта проведения вычислительного эксперимента:

- Синхронный – все стадии жизненного цикла выполняются последовательно во всех клетках;
- Асинхронный – для каждой клетки стадии жизненного цикла выполняются независимо и случайно.

В следующем разделе приведены детали программной реализации ПК *Microcosm*.

2.8.7 Программный комплекс *Microcosm*

Опишем основные классы программного комплекса *Microcosm*:

Cell – класс, отвечающий за реализацию агента(клетки). В его наследниках должны быть реализованы основные функции клетки;

Environment – класс, отвечающий за реализацию одной ячейки среды. В его наследниках должны быть реализованы функции, определенные в описании модели среды;

EnvironmentContainer – контейнер, в которые помещаются одиночные ячейки среды в соответствии с их пространственным положением. Служит посредником между клетками и единичными ячейками среды;

GeneNetwork – класс, отвечающий за реализацию генных сетей внутриклеточного метаболизма. В данной работе использовалась единственная реализация данного класса – *TriggerGeneNetwork*, класс, реализующий сеть генетического триггера;

MoveStrategy, *DivisionStrategy*, *ForagingStrategy*, *DeathStrategy*, *SecretionStrategy* – классы, наследники которых должны реализовывать соответственно стратегии (модели) передвижения, деления, поедания, смерти и выделения субстратов;

Итоговую архитектуру программы можно упрощенно изобразить на следующей диаграмме (рис. 2.8.3).

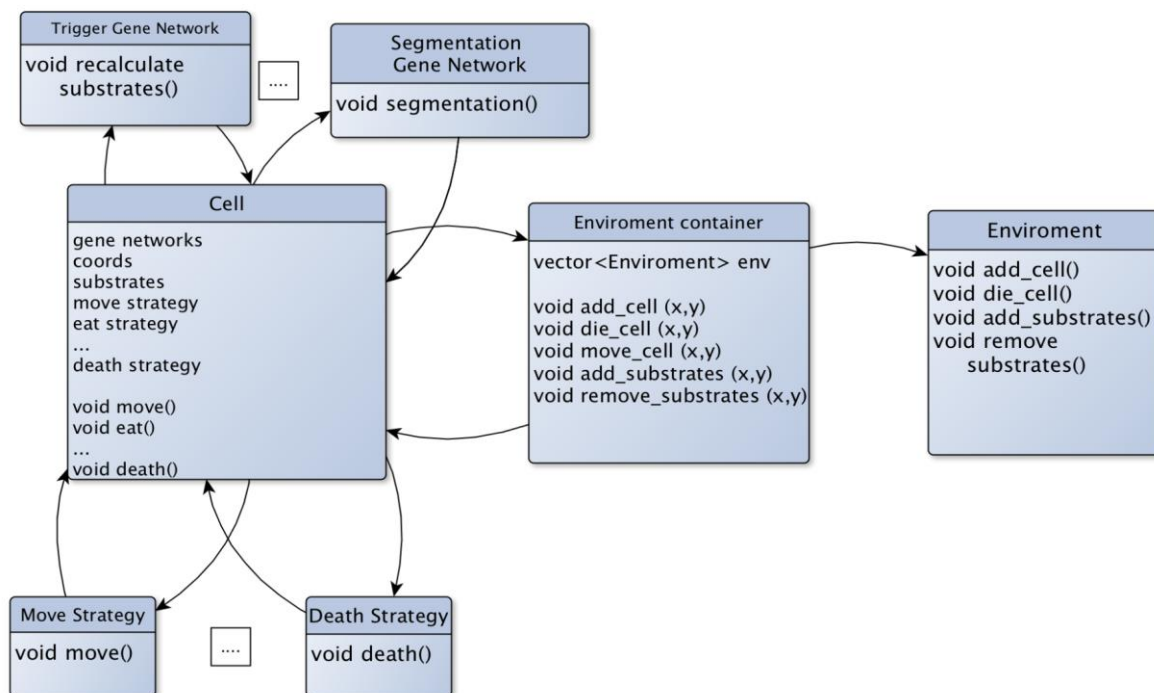


Рисунок 2.8.3. Архитектура ПК Microcosm. Стрелка А→В означает, что реализация класса А явно зависит от реализации класса В.

Для программной реализации генных сетей был выбран паттерн проектирования «стратегия» (Александреску, 2019). Данный паттерн позволяет очень гибко и «безболезненно» варьировать систему в процессе эволюции микробного сообщества, давая возможность комбинировать между собой различные подмодели, каждая из которых программно реализует ту или иную стратегию. Далее отождествим слова стратегия и подмодель.

Подмодели агента можно условно разбить на два типа: те, которые явно зависят от наблюдений, получаемых от среды и те, которые явно не зависят от наблюдений. В качестве примера подмодели первого типа можно привести простейшую модель поедания клеткой нутриентов из среды (рис. 2.8.4). В качестве стратегии, явно независимой от внешних наблюдений, можно

привести модель внутриклеточного синтеза субстратов (схема приведена в предыдущем разделе на рис. 2.8.4).

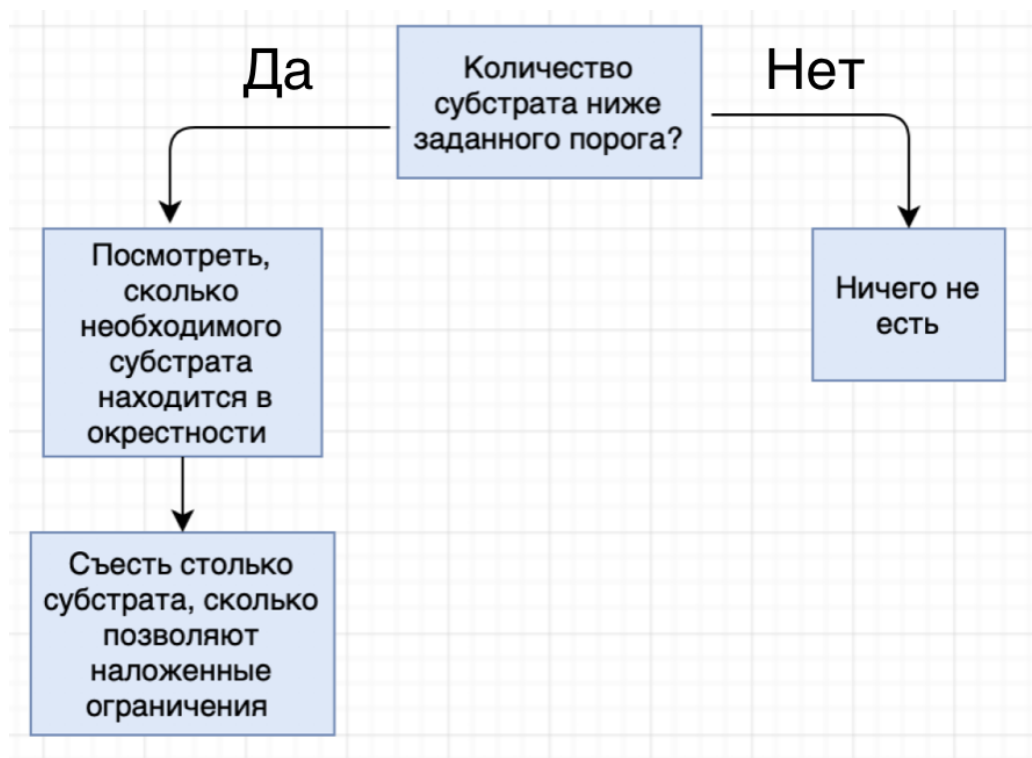


Рисунок 2.8.4. Пример алгоритма простейшей подмодели потребления нутриентов из среды.

Поток исполнения в ПК *Microcosm* в целом аналогичен таковому в ПК ГЭК (см. раздел 2.6.2). В ходе каждого «тика» (одной итерации по времени) программы сначала последовательно исполняются все функции, связанные с жизнедеятельностью клеток: метаболизм, секреция, потребление субстратов, деление, передвижение, смерть. Такая последовательность действий делается для каждого агента. В реализацию жизненного цикла заложено то, что некоторые функции исполняются не на каждом тике программы, а на каждом k -м тике (отметим, что значение k подбирается независимо для каждой функции). Это делается для того, чтобы учитывать характерные времена каждой из функций, соответствующие характерным временам реальных биологических процессов. После выполнения всех клеточных функций программа исполняет функции, связанные со средой.

Одна итерация модельного цикла в ПК Microcosm представлена диаграммой на рисунке 2.8.5.

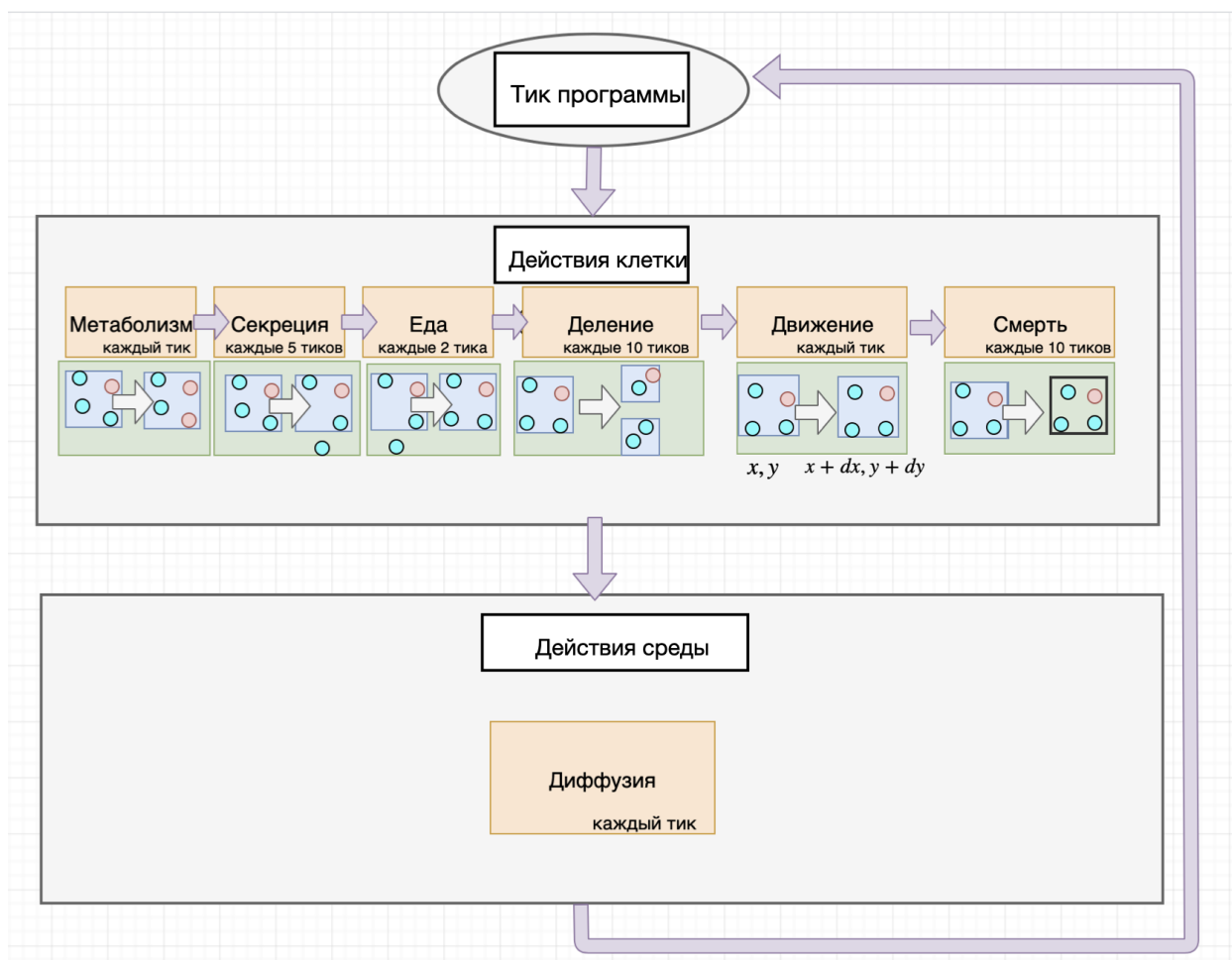


Рисунок 2.8.5. Схема потока вычислений в ПК Microcosm.

Программный комплекс Microcosm доступен по <https://gitlab.sysbio.cytogen.ru/kravchuk/microcosm>.

2.8.9. Верификация программного комплекса Microcosm

В ходе верификации ПК Microcosm были проведены вычислительные эксперименты по воспроизведению классических моделей динамики и генетики микробных популяций. Были проведены вычислительные эксперименты по сравнению эволюционной эффективности различных моделей перемещения клеток в среде обитания, по оценке влияния диффузионных параметров нутриентов на эволюцию сообществ и другие.

Были получены численные результаты, подтверждающие биологическую адекватность программного комплекса Microcosm.

2.9. Заключение к главе 2

В данной главе описаны результаты работ по разработке методов моделирования функционирования и эволюции сложноорганизованных микробных сообществ, их эволюции с учётом пространственной организации и гетерогенности сред обитания таких сообществ, а также их реализации в виде программных комплексов «Гаплоидный эволюционный конструктор» (ГЭК) и «Microbial community simulator» (Microcosm).

Программный комплекс ГЭК включает в себя: (1) консольную версию для запуска в пакетном режиме, в т.ч. и на высокопроизводительных кластерах с использованием MPI и OpenMP; (2) настольное кроссплатформенное графическое приложение, написанное с использованием платформы Qt и модуля QtConcurrent для проведения расчётов на настольных компьютерах с многоядерными процессами; (3) веб-приложение WebHEC (<http://webhec.sysbio.cytogen.ru>), предоставляющее возможность моделирования вычислительно тяжёлых задач с использованием кластеров Центра коллективного пользования «Биоинформатика» ИЦиГ СО РАН, Новосибирского Государственного Университета и Сибирского Суперкомпьютерного Центра, и включающее в себя библиотеку моделей микробных сообществ, написанных на специальном языке описания моделей ПК ГЭК (см. Приложение 1).

Вычислительное ядро ПК ГЭК позволяет моделировать пространственно распределённые сообщества в одно-, двух- и трёхмерном случае, соответственно. При этом рассматриваются такие процессы как диффузия и перенос под действием протока для метаболитов, а для клеток – пассивное (с помощью протока) и активное (с помощью хемотаксиса) перемещение. Язык описания моделей ГЭК (которые могут транслироваться ядром ГЭК) содержит средства для описания пространственной структуры сообществ: неоднородное

пространственное распределение, размерность, факторы перераспределения и другие команды, предназначенные для конструирования сложных модельных сценариев (см. Приложение 1).

Комплексное тестирование разработанных высокопроизводительных модулей показало следующие результаты: (1) эффективность распараллеливания до 180% с использованием небольшого числа процессов на MPI на сложных задачах; (2) ускорение до 40 раз с использованием большого числа процессов MPI; (3) ускорение до 3 раз с использованием QtConcurrent в графическом приложении на 6-ядерном процессоре. Были сформулированы следующие рекомендации по использованию высокопроизводительных модулей ПК ГЭК. На задачах с расчётом сложных моделей стоит использовать высокопроизводительный кластер с распределенной памятью при наличии не более 50 процессов. Если есть возможность использования 50 процессов на SMP машине, то это будет намного эффективнее по сравнению с высокопроизводительным кластером с распределенной памятью.

Выполнена интеграция ПК ГЭК с существующими средствами моделирования генных сетей и базами данных (Biomodels, CellML, SabioRK, MAMMOTH) с помощью стандарта Systems Biology Markup Language (SBML) и сопутствующих программных средств. Кроме того, выполнена интеграция базы данных наименований химических веществ ChEBI. Для всех модулей выполнена верификация и показано соответствие их функциональности заявленных задачам. Разработано большое количество содержательных моделей микробных сообществ, для которых показана биологическая адекватность как самих моделей, так и разработанных средств моделирования. В результате создана библиотека моделей, интегрированная в веб-приложение WebNES.

Программный комплекс Microcosm, реализованный в агентно-ориентированной парадигме (с разрешением до отдельной клетки), является дальнейшим развитием ПК ГЭК и предназначен для исследования микробных сообществ относительно небольшого размера, в которых возможно появление

сложноорганизованных паттернов, связанных с рассинхронизацией клеточных процессов и/или особых сигнальных путей, участвующих в формировании т.н. чувства кворума.

С помощью разработанных ПК получены содержательные биологические результаты, опубликованные в большом количестве научных статей. В частности, были исследованы эволюционные ландшафты приспособленности бактерий (Лашин et al., 2020). Показано, что успех в преодолении «долин приспособленности» определяется как различной приспособленностью разных аллельных комбинаций, так и частотой мутаций. Для каждого ландшафта приспособленности существует своя оптимальная частота мутаций. Более того, в зависимости от характера ландшафта приспособленности оптимальными для достижения пика оказываются либо градуальный (постепенная дивергенция между родственными таксонами путём равномерного накопления небольших изменений – «природа не делает скачков» (Иорданский, 2001; Лашин et al., 2012b)), либо сальтационный (быстрые и значительные преобразования предковых форм (Иорданский, 2001; Лашин et al., 2012b)) эволюционные режимы.

Были построены и исследованы модели конкуренции «оседлой» и «кочевой» популяции в среде обитания с периодическим во времени и пространстве субстратом, а также модель формирования популяций подвижных микроорганизмов с различными адаптационными стратегиями в периодически изменяющейся среде. Показано, что даже при 5% увеличении доли активно мигрирующей субпопуляции (0.0105 у кочевника против 0.01 у оседлого) наблюдается преобладание численности кочевника и постепенное вымирание оседлого. Однако, этот процесс идёт значительно медленнее в том случае, если оседлый эффективнее утилизирует субстрат, чем кочевник. Преимущество кочевой популяции выше в условиях ограниченности субстрата, в то время как преимущество эффективной утилизации субстрата, которым обладают клетки оседлой популяции, более явно проявляет себя в условиях обилия субстрата.

В условиях высокоизменчивой в смысле доступности субстрата среды обитания и однородного начального распределения клеток действует дизруптивный отбор по признаку энергетической стоимости миграции.

Проведено моделирование влияния фаговой инфекции на эволюцию микробных сообществ при различных экологических условиях. В результате моделирования показано, что фаговая инфекция влияет на эволюцию микробного сообщества, замедляя видообразование и стабилизируя систему в целом. Это влияние отличается по своей эффективности в зависимости от пространственно-экологических факторов, а также от состояния сообщества на момент инвазии фага.

3. Моделирование популяций диплоидных организмов

Базируясь на опыте разработки методов и программ для моделирования функционирования и эволюции микробных сообществ, описанных в предыдущей главе, мы провели серию работ по разработке инструментария для теоретического исследования популяций более сложных живых организмов с диплоидной организацией генома и половым размножением. Представителями таких организмов являются животные и люди. В этой связи в качестве основных объектов методики рассматриваются: среда, особь, а также популяция, как совокупность особей.

3.1. Методика моделирования популяций сложноорганизованных диплоидных организмов

Разрабатываемая в данной работе методика – «Диплоидный эволюционный конструктор» (ДЭК) – ориентирована на агентное (agent-based) моделирование эволюционных, популяционно-генетических и популяционно-экологических процессов в *популяциях* сложноорганизованных *организмов* (*особей*). Основным параметром такого сложноорганизованного организма, влияющим на его выживаемость и способность к репродукции, является *внутреннее состояние* (или «здоровье») организма. Этот параметр обуславливается отдельными *фенотипическими признаками*, которые, в свою очередь, зависят как от наследственной неизменяемой компоненты (*генотип*), так и от изменяемой *средовой компоненты*. Основным объектом эволюции в биологии является *популяция*, под которой понимается элементарная группировка организмов определённого вида, обладающая всеми необходимыми свойствами для поддержания своей численности длительное время в постоянно изменяющейся среде. К таким свойствам можно отнести (Байков et al., 2006): способность к свободному *скрещиванию* (панмиксии), благодаря чему популяция формирует свой собственный *генофонд*, отличный от других популяций, и способна эволюционировать; существование

минимальной численности, достаточной для поддержания жизнеспособности – устойчивого оборота поколений на минимально возможной территории и характеризующегося временем жизни одного поколения, минимальным пространством, необходимым для потока поколений и уровнем плотности особей, необходимым для устойчивого размножения (Загульнова et al., 1993; Смирнова et al., 1993); функциональное единство особей в пределах ареала, т.е. возможность замещения любой особи другими (Беклемишев, 1970).

Исходя из вышеизложенного, были выделены основные объекты методики моделирования (в терминах агентно-ориентированного моделирования – агенты): *особь* и *окружающая среда (экосистема)*. Ещё один важнейший объект методики (не являющийся агентом) – *популяция* особей. Взаимоотношения этих ключевых объектов между собой схематично представлены на рисунке 3.1.1: окружающая среда может содержать несколько популяций, каждая из которых может содержать несколько особей (характерные размеры популяции – 100-10000 особей).

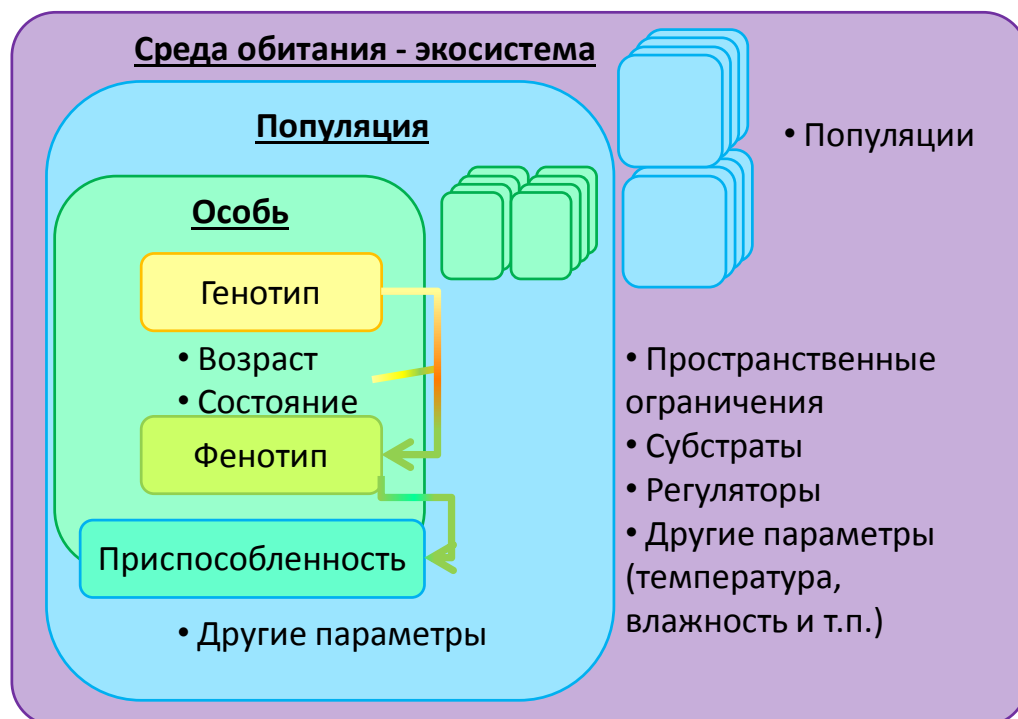


Рисунок 3.1.1. Основные объекты методики ДЭЖ и их свойства.

Взаимодействие между объектами модели в методике осуществляется по принципу *многослойного моделирования* (multilayered model), предложенному нами ранее для описания эволюционно-популяционных процессов в бактериальных сообществах (см. литературный обзор). В многослойной модели каждый «слой» представляет собой подмодель, описывающую отдельный иерархический уровень биологической организации. Например, геномный слой описывает только организацию генома, не касаясь при этом ни описания отдельных генов, их функций и т.д., ни описания влияния генома и его частей на функционирование организма. Для каждого слоя определяется набор спецификаций, которому должна удовлетворять каждая модель данного слоя. Выделение отдельных слоёв в модели отвечает основными принципам объектно-ориентированного программирования (абстракция, инкапсуляция, наследование и полиморфизм), в результате чего появляется возможность создавать множественные реализации (подмодели – как детали конструктора) для всех объектов модели, не меняя при этом её структуры. Отметим, что применение «многослойной» концепции позволяет комбинировать различные методы математического моделирования в одной общей модели. Например, поступление питательных веществ в организм из среды можно описать с помощью сетей Петри, а их метаболизм, с учётом генотипа организма, его возраста и общего состояния (здоровья) уже описать с помощью обыкновенных дифференциальных уравнений (численно решить задачу Коши).

Другой важной составляющей многослойного моделирования является концепция *стратегий* – специальных моделей, описывающих законы изменения слоёв, а также взаимодействия между отдельными слоями. С точки зрения программной реализации, слои содержат данные, а стратегии манипулируют этими данными (Александреску, 2019).

Перечислим основные классы слоёв в нашей методике.

Слои уровня отдельной особи:

1. Генетический слой. Содержит информацию об отдельных «генах». Ген в модели может быть представлен в виде числа (целого или с плавающей точкой), элемента перечислимого множества (`enum` в C++) или строки (название гена, нуклеотидная последовательность).
2. Геномный слой. Содержит информацию о полном геноме организма. Отдельные гены расположены в *хромосомах*, для описания которых в модели введены вспомогательные объекты.
3. Метаболический слой. Содержит информацию о метаболитах и других функциональных факторах организма (концентрации отдельных субстратов и несубстратных регуляторов).
4. Фенотипический слой. Содержит информацию о фенотипических признаках организма. Признаки могут быть представлены в виде чисел (целых или с плавающей точкой), элементов перечислимого множества (`enum` в C++), строк.
5. Слой внутреннего состояния организма и его выживаемости. Содержит информацию о возрасте и других характеристиках организма (в качестве одной из характеристик может использоваться информация других слоёв модели – метаболического, фенотипического и т.д.).

Слои уровня популяции и экосистемы:

6. Популяционный слой. Содержит информацию о множестве особей, составляющих популяцию, с учётом поколений; минимальной и эффективной численности популяции, ареале обитания (пространственном распределении особей) и т.д.
7. Экологический слой. Содержит информацию об окружающей среде (ареале – среде обитания популяций): субстратах и

несубстратных регуляторах, географических (топологических) особенностях ареала, его границах с другими ареалами и, наконец, распределении популяций по ареалу.

Основные классы стратегий, описывающих изменение отдельных слоёв, а также взаимодействие между разными слоями, приведены ниже:

1. Мутационная стратегия. Описывает мутацию отдельного гена. В зависимости от представления гена в модели (число, элемент перечислимого множества, строка) мутация может изменять количественную или качественную характеристику гена, а, в случае строкового представления, моделировать известные типы генетических мутаций: замены нуклеотидов, инсерции, делеции и транспозиции.
2. Рекомбинационная стратегия. Описывает образование нового гаплоидного генома из двух родительских в процессе мейоза (при моделировании процесса полового размножения). Из двух гомологичных наборов генов (диплоидного генома), хранящихся в хромосомах, собирается один набор генов (гаплоидный геном), который затем будет использован при формировании диплоидного генома новой особи. Комбинаторика, возникающая в процессе такой сборки, соответствует таким биологическим процессам как кроссинговер (обмен участками гомологичных хромосом), конверсия, перестройки/транслокация хромосомы, дупликация отдельных генов и т.д.
3. Стратегия «генотип→фенотип». В широком смысле, описывает отображение множества элементов генома (отдельных генов) на множество фенотипических признаков с учётом общего состояния организма. Важнейший подкласс данной стратегии – стратегия метаболизма, описывающая изменение концентраций отдельных веществ (субстратов) в течение определённого времени под действием

процессов синтеза, деградации и т.д. Эффективность этих метаболических процессов детерминирована генетически. Подчеркнём, что для описания стратегий метаболизма можно использовать любые математические подходы, применяемые для описания динамики функционирования генных сетей (Колчанов et al., 2008; Фадеев, Лихошвай, 2003). В частности, это дифференциальные и дифференциально-алгебраические уравнения, стохастическое моделирование, булевы сети и сети Петри, пороговые модели, конечные автоматы и другие подходы (Афонников, Миронова, 2014; Колчанов et al., 2008; Ратушный et al., 2005).

4. Стратегия «фенотип→приспособленность» или «организм–среда». Описывает взаимодействие отдельной особи со средой обитания, в том числе процессы субстратного и информационного обмена. Определяет приспособленность особи к конкретным условиям среды с учётом фенотипических признаков особи.
5. Стратегия размножения и отбора. Описывает такие популяционные процессы, как выбор брачных партнёров, конкуренцию, половое размножение. Определяющую роль в этих процессах играет приспособленность особей, а также их фенотипические и генотипические особенности.
6. Миграционная стратегия. Описывает распределение особей внутри ареала, миграцию особей внутри одного ареала и между разными ареалами. Важную роль в этих процессах также играет приспособленность особей.

Отметим, что описанная в данном разделе структура комплексной модели эволюционных, популяционно-генетических и популяционно-экологических процессов в популяциях особей не является статичной. При накоплении новых данных и знаний об этих процессах возможно расширение модели за счёт добавления новых типов стратегий. Это позволяет рассматривать нашу

методику как некий конструктор моделей, где в качестве составных блоков-подмоделей используются модели слоёв и стратегий. Исходя из этой особенности методики, программный комплекс, описанный ниже, тоже получил название ДЭК.

Как было сказано выше, основные объекты ДЭК – среда и индивидуум (агент, особь).

Среда содержит следующие основные объекты:

- Особи
- Субстраты (в т.ч. и субстратные регуляторы)
- Несубстратные регуляторы

При этом субстраты – являются исчерпаемыми, их наличие среде зависит от функционирования особей, тогда как несубстратные регуляторы – неисчерпаемы (в том смысле, что их наличие не зависит от деятельности особей, а зависит от других внешних факторов). Такая классификация соответствует принятым в экологии понятиям: «ресурсам и параметрам».

Особь характеризуется следующими параметрами – слоями описания, каждый из которых представляет собой подмодель (возможно, иерархически содержащую другие подмодели):

- Генотип
- Возраст
- Внутреннее состояние (концентрации различных веществ и т.д.)
- Фенотип – производная трёх предыдущих параметров
- Приспособленность – производная фенотипа и среды

Для описания организации генома особи используется отдельный тип подмоделей: хромосомы, в которых содержатся конкретные гены, что

позволяет моделировать группы сцепления (маркированные). Использование подмоделей хромосом позволяет моделировать следующие характеристики/процессы:

- Диплоидность
- Кроссинговер, конверсии
- Перестройки/транслокации
- Дупликации
- Половой процесс

Признаки, кодируемые генами, могут рассматриваться как точечные (моногенные), полигенные и/или как компенсируемые признаки (для реализации скрытого резерва, как изменчивости, так и устойчивости к мутациям и/или инбридингу). Связь фенотипа с генотипом осуществляется по схеме, предложенной ещё А.А. Ляпуновым (Ляпунов, 1972), развивавшейся впоследствии во многих работах (см., например, (Knibbe et al., 2008, 2007)).

По генотипам с помощью специальных функционалов строятся интегральные признаки, совокупность которых формирует фенотип, который сравнивается с фенотипом «идеальной приспособленности», который, в свою очередь, строится с помощью функционалов, зависящих от параметров окружающей среды. При этом возможно моделирование как полимерии, так и плейотропии.

В моделях ДЭЖ рассматриваются 3 группы генов:

1. «Коровые» гены – в основном влияющие на фенотип и приспособленность;
2. Гены регуляции – влияют на коровые гены;
3. Гены стресса – в нормальных условиях работают как гены регуляции, дублируя гены регуляции; в пессимальных условиях обеспечивают

кратковременную, но мощную экспрессию комплекса регуляторных и коровых генов, изменяя вид кривой выживаемости (вместо постепенного ухудшения состояния организма в условиях недостатка ресурса(ов) пролонгированное неизменное состояние с последующим резким падением жизнеспособности или гибелью).

В итоге для отображения «значений» или «параметров» генов на их «функцию» используется три класса фильтр-функционалов, каждый из которых также представляет собой слой моделирования. Обратная связь обеспечивается влиянием изменённого организма, как на его состояние, так и на окружающую среду (что потом может опять повлиять на регуляторные гены и гены стресса).

3.2. Программный комплекс «Диплоидный эволюционный конструктор»

В результате проектирования программного комплекса ДЭК были выделены следующие основные модули системы:

- Вычислительное ядро ДЭК. Реализует все описанные выше объекты и процессы модели.
- Консольный модуль управления процессом моделирования, включая транслятор специального языка моделирования ДЭК.
- Графический модуль управления процессом моделирования (графический интерфейс пользователя).
- Модуль(и) поддержки высокопроизводительных вычислений в ДЭК. На текущий момент в стадии проектирования.

Каркас ДЭК составляют следующие классы:

Gene – предназначен для описания модельного гена. Содержит перечисление GeneType, определяющее тип представления гена (число, элемент перечислимого множества, последовательность).

```
class Gene {
```

```

public:
    enum GeneType { Continuous, Discrete };
    typedef unsigned int idType;

protected:
    GeneType geneType;          // Дискретный или непрерывный
    idType ID;                  // Идентификатор
    std::string name;           // Название признака
    float geneValueCont;        // Значение признака, если он непрерывный
    int geneValueDiscr;         // Значение признака, если он дискретный

public:
    // методы
};

```

Chromosome – предназначен для хранения набора генов в виде модельной хромосомы (группы сцепления генов). Обеспечивает операции вставки и удаления генов. Более сложные операции (моделирующие рекомбинацию и т.п. биологические процессы) делегированы отдельному классу **рекомбинационной стратегии** ChromosomeRearrangementStrategy.

```

class Chromosome {
    friend class ChromosomeRearrangementStrategy; // Стратегия рекомбинации
protected:
    std::string name;
    std::vector<Gene> genes;
public:
    Chromosome(std::string _name) : name(_name) {};
    Chromosome(std::string _name, Gene gene);
    Chromosome(std::string _name, const std::vector<Gene>& _genes);
    void insertGene(Gene gene, int pos = 0);
    void insertGeneToEnd(Gene gene, int pos = 0);
    // остальные методы
};

```

Классы, описывающие геном особи, представлены двумя классами: HaploidGenotype – для описания гаплоидного генома, и Genotype – для описания диплоидного генома. Первый класс выполняет сервисную функцию контейнера хромосом:

```

class HaploidGenotype {
    friend class ChromosomeRearrangementStrategy;
protected:
    std::vector<Chromosome> chromosomes;
public:
    const Chromosome& getChromosome(unsigned int i) const;
    HaploidGenotype(const std::vector<Chromosome>& _chrom);
};

```

Второй класс обеспечивает операции доступа к гаплоидным геномам родителей, а также создание рекомбинантного гаплоидного генома (делегировав при этом эту функциональность классу рекомбинационной стратегии).

```
using namespace std;
class Genotype {
    friend class ChromosomeRearrangementStrategy;
protected:
    HaploidGenotype fatherGenome;
    HaploidGenotype motherGenome;
public:
    Genotype(const vector<Chromosome>& _fGenome, const vector<Chromosome>&
_mGenome);
    Genotype(const HaploidGenotype& _fGenome, const HaploidGenotype&
_mGenome);
    Genotype(const Genotype& _fDiplGenome, const Genotype& _mDiplGenome);
    HaploidGenotype
    recombinantHaploidGenotype(ChromosomeRearrangementStrategy*) const;
    const HaploidGenotype& getFatherGenome() const;
    const HaploidGenotype& getMotherGenome() const;
};
```

Trait – предназначен для описания модельного фенотипического признака. Аналогично классу Gene, содержит перечисление TraitType, определяющее представление признака.

```
class Trait {
public:
    enum TraitType { Continious, Discrete };
    typedef unsigned int idType;

protected:
    TraitType traitType; // Дискретный или непрерывный
    idType ID; // Идентификатор
    std::string name; // Название признака
    float traitValueCont; // Значение признака, если он непрерывный
    int traitValueDiscr; // Значение признака, если он дискретный

public:
    // методы
};
```

Phenotype – предназначен для описания всего набора фенотипических признаков. Обеспечивает операции вставки признака и доступа к признаку. Операции вычисления признаков по геному и приспособленности по фенотипу делегированы отдельным классам: GenotypeToPhenotypeStrategy и PhenotypeToFitnessStrategy, соответственно.

```

class Phenotype {
    friend class GenotypeToPhenotypeStrategy;
    friend class PhenotypeToFitnessStrategy;
protected:
    std::map<std::string, Trait> traitsMap;
public:
    Phenotype(){};
    Phenotype(const std::map<std::string, Trait>& _tm) : traitsMap(_tm) {}
    Phenotype(Trait _trait);
    Trait getTraitByName(std::string name) const;
    void addTrait(Trait _trait);
};

```

InnerSubstratesPool – предназначен для описания внутренних субстратов организма. Изменения концентрации субстратов, связанные с метаболизмом, делегируются классам GenotypeToPhenotypeStrategy и MetabolicStrategy.

```

using namespace std;
class InnerSubstratesPool {
protected:
    vector<float> subConcentrations; // Концентрации субстратов
    map< string, int> subNamesMap; // Словарь названий субстратов
    vector<float> regConcentrations; // Концентрации несубстратных
регуляторов
    map< string, int> regNamesMap; // Словарь названий несубстратных
регуляторов
public:
    InnerSubstratesPool();
    InnerSubstratesPool(const vector<float>& s, const map<string, int>& n);
    InnerSubstratesPool(const vector<float>& s, const map<string, int>& n,
        const vector<float>& _regs, const map< string, int>& _regNames);

    virtual void utilizeSubstrate(int num, float count);
    virtual void utilizeRegulator(int num, float count);
    virtual void utilizeSubstrate(string name, float count);
    virtual void utilizeRegulator(string name, float count);
    virtual void addSubstrate(int num, float count);
    virtual void addRegulator(int num, float count);
    virtual void addSubstrate(string name, float count);
    virtual void addRegulator(string name, float count);
};

```

Region – предназначен для описания отдельного ареала (среды обитания) популяции.

```

class Region {
protected:
    map<SpeciesID, long int> capacities // ёмкость среды для видов
    vector<long int> currentSubstrates; // текущие объёмы субстратов в среде
    vector<long int> restorationSubstrates; // объёмы восстановления
субстратов
    vector<long int> maxSubstrates; // максимальные объёмы субстратов
    std::vector<double> regulators; // регуляторы
    // temperature, humidity и другие параметры

    Position3D<double> position3D; // Топологические характеристики ареала

```



```

public:
    Region(Position3D<double> pos, long int capac);
    Region(Position3D<double> pos, long int capac,
           const vector<long int>&, const vector<long int>&,
           const vector<long int>&, const vector<double>&);
    void restoreSubstrate();
    long int getCapacity() const { return this->capacity;}
};

```

Environment – предназначен для описания совокупности ареалов, составляющих всю доступную среду обитания в текущей модели. Операции миграции особей внутри ареала и между ареалами делегированы классу MigrationStrategy.

```

class Environment {
    friend class MigrationStrategy;
protected:
    std::vector<Region> regions;
    map<SpeciesID, Population*> populations;
public:
    ////
};

```

Классы, описывающие популяцию организмов, представлены несколькими классами: абстрактным классом Population, классом AsexualPopulation (популяция бесполовых организмов) и классом BisexualPopulation (популяция двуполовых организмов). Операции, связанные с отбором и размножением особей, делегированы классу PopulationBreedingStrategy. Классы AsexualPopulation и BisexualPopulation реализуют чисто виртуальные методы и хранят информацию об особях, в первом случае в одном контейнере, во втором в двух (мужские и женские особи).

```

class Population {
    friend class PopulationBreedingStrategy;
protected:
    PopulationBreedingStrategy* breedingStrategy;
    Region* region;
public:
    Population(PopulationBreedingStrategy* _strategy) :
breedingStrategy(_strategy) {}
    virtual ~Population() {};
    virtual void setBreedingStrategy(PopulationBreedingStrategy* _strategy);
    virtual void setRegion(Region* _region) { region = _region;}
    virtual void calculateFitnessAll() = 0;
    virtual void breedAll() = 0;
    virtual void selectionAll() = 0;
};

```

```

virtual void mutationAll() = 0;
virtual std::string getStatistics() = 0;
virtual void putGeneticStatisticsToStream(std::ostream& out) = 0;
virtual void putGeneticSimpleStatisticsToStream(std::ostream& out) = 0;
// другие статистические сервисные методы
};

```

Классы стратегий, описанные выше и перечисленные в списке ниже, являются интерфейсными классами. Они не имеют нестатических полей (статические поля используются в некоторых случаях для хранения настроек отдельных алгоритмов) и обладают одним методом или несколькими перегруженными версиями одного метода. В текущей версии ДЭЖ имеются следующие классы стратегий:

- ChromosomeRearrangementStrategy
- GenotypeToPhenotypeStrategy
 - (и её подкласс) MetabolicStrategy
- PhenotypeToFitnessStrategy
- PopulationBreedingStrategy
- MigrationStrategy

Пример класса стратегии рекомбинации и простейшая реализация метода `buildRecombinantGenotype` приведены ниже:

```

class SimpleRecombStrategy : public ChromosomeRearrangementStrategy {
public:
virtual HaploidGenotype buildRecombinantGenotype(const Genotype*) {
    std::vector<Chromosome> chromosomes;
    const std::vector<Chromosome>* fGenome = &parentGenotype->
fatherGenome.chromosomes;
    const std::vector<Chromosome>* mGenome = &parentGenotype->
motherGenome.chromosomes;

    for (unsigned int i = 0; i < fGenome->size(); i++){
        int flag = rand() % 2;
        const std::vector<Gene>* first;
        const std::vector<Gene>* second;
        if(flag == 0){
            first = &fGenome->at(i).genes;
            second= &mGenome->at(i).genes;
        }
        else{
            first = &mGenome->at(i).genes;
            second= &fGenome->at(i).genes;
        }
        Chromosome childChromosome(fGenome->at(i).name, *first);
        // Для каждой хромосомы выбираем свою "точку перегиба"
        int bound = rand() % first->size();
        for(unsigned int j = bound; j < first->size(); j++){
            childChromosome.genes.at(j) = second->at(j);
        }
    }
}

```

```

        chromosomes.push_back(childChromosome);
    }
    return HaploidGenotype(chromosomes);
};

```

Простейшая стратегия размножения и отбора в популяции приведена ниже:

```

void SimplePopulationBreedingStrategy::breed(Population* _pop){
    float deathRate = Settings::NaturalDeathRate; // Коэффициент смертности

    // Выбираем слабых и удаляем их
    AsexualPopulation* pop = dynamic_cast<AsexualPopulation*>(_pop);
    if(pop == NULL){
        std::cerr<<"Wrong dynamic cast to AsexualPopulation\n";
        return;
    }
    std::sort(pop->individs.begin(), pop->individs.end(), compareOnFitness);
    int inds = pop->individs.size();

    int toDie = int (inds * deathRate);
    std::vector<Individual*>::iterator start = pop->individs.begin();
    std::vector<Individual*>::iterator end = pop->individs.begin()+toDie;
    for(std::vector<Individual*>::iterator it = start; it != end; it++){
        delete *it;
    }
    pop->individs.erase(start, end);

    // Рожаем новых особей
    float birthRate = Settings::BirthRate;
    int toBorn = int (inds * birthRate);
    inds = pop->individs.size();
    srand((unsigned int)time(NULL));
    for(int i = 0; i < toBorn; i++) {
        int male = rand() % inds;
        int female = rand() % inds;
        female = (female != male ? female : (female + 1) % inds);

        Individual* ind = new Individual(*(pop->individs.at(male)), *(pop->
individs.at(female)));
        // Здесь по необходимости добавляем настройки среды и т.д.
        pop->individs.push_back(ind);
    }
}

```

Подобное архитектурное решения для классов стратегий позволяет создавать множественные реализации для всех этапов моделирования (см. подробнее Приложение 2. Система плагинов ДЭК).

3.2.1. Модули поддержки высокопроизводительных вычислений ПК ДЭК

В рамках работы над ПК ДЭК были разработаны, программно реализованы и протестированы модули поддержки высокопроизводительных вычислений (HPC, high performance computing) с

использованием таких технологий как MPI, OpenMP и Nvidia Cuda. Использование всех трёх технологий обусловлено с одной стороны большим спектром разнообразия решаемых задач, с другой – необходимостью иметь возможность использовать программный пакет как на высокопроизводительных кластерах, так и на стационарных компьютерах и ноутбуках. При расчёте популяционных моделей с большим генетическим разнообразием (сотни тысяч и миллионы аллельных комбинаций в отдельной популяции) программы без поддержки высокопроизводительных вычислений требуют десятков и сотен часов для расчёта хотя бы нескольких сотен поколений жизни популяции. Отметим при этом, что в ряде задач (например, при компьютерной реконструкции истории популяций древних людей на основе палеогенетических и/или археологических данных) необходимо моделировать тысячи и десятки тысяч поколений.

С помощью ДЭК была построена тестовая модель популяционно-генетических и миграционных процессов в популяциях древних людей (Gunbin, Lashin, 2013) (детальное описание в подглаве 3.3). На этой модели было проведено профилирование ДЭК, которое показало, что большая часть времени (80-90%) уходит на процедуру выбора брачных пар в текущей стратегии размножения (процедурах смертности и рождаемости).

OpenMP версия ДЭК

Была проведена оптимизация данной функции ДЭК с использованием технологии OpenMP (в версии ДЭК для настольных компьютеров и ноутбуков). Оптимизация состояла из двух этапов:

- 1) При моделировании процесса смертности исходный список популяции особей был разбит на несколько списков меньшей длины, смертность рассчитывалась в каждом из них и, в дальнейшем, производилась их склейка.

Экспериментальным путём был получен примерный порог, когда склейка все большего числа списков (разрезов) начинает занимать время, достаточное для остановки прироста производительности (см. рис. 3.2.1 – зависимость времени моделирования от числа разрезов). К 20-30 разрезам было достигнуто состояние, в котором выигрыш времени был минимален. Далее было взято три последних варианта и для каждого проведено тестирование процесса смертности с использованием OpenMP (т.е. параллельная обработка этих разрезов). Выигрыш оказался минимальным.

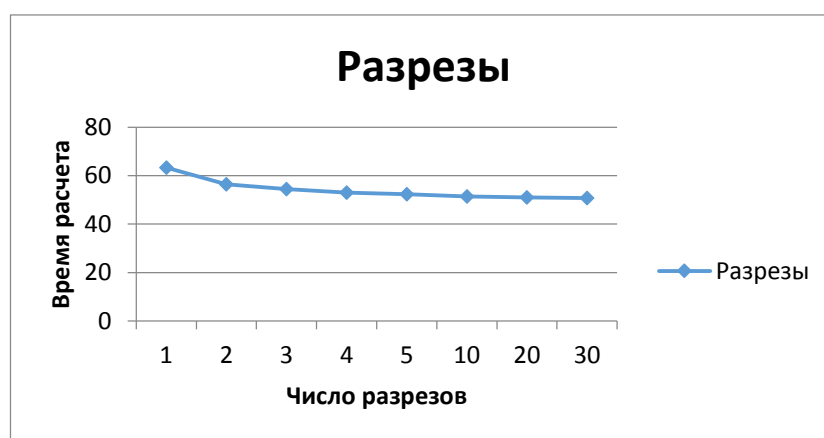


Рисунок 3.2.1. Время расчёта тестовой модели ДЭК на первых 1000 итерациях (в секундах) в зависимости от числа разрезов списка особей.

2) Технология OpenMP была использована для моделирования процесса рождаемости. Вместо последовательного добавления новых особей в исходный список особи добавлялись в пустой список, индивидуальный у каждого потока. По окончании моделирования стадии рождаемости эти списки были присоединены к исходному. Зависимость времени выполнения от числа потоков можно наблюдать на рисунке 3.2.2.

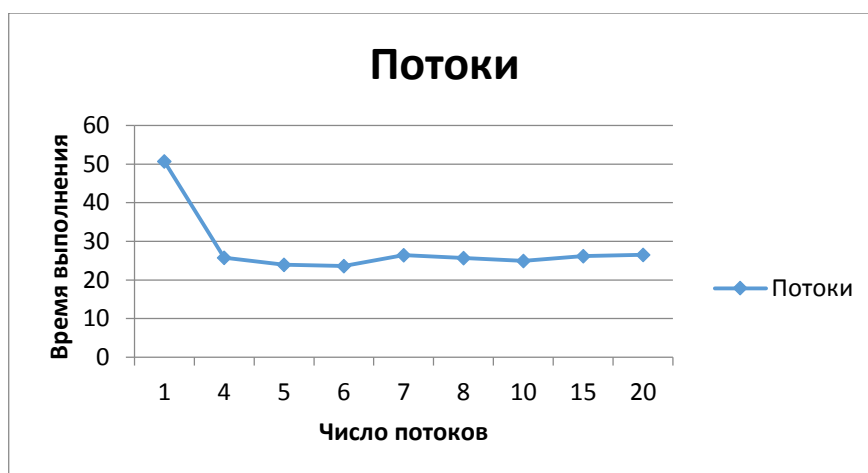


Рисунок 3.2.2. Время расчёта тестовой модели ДЭК на первых 1000 итерациях (в секундах) в зависимости от числа потоков в OpenMP версии, при распараллеливании процесса рождаемости.

Были проведены замеры скорости выполнения оптимизированной OpenMP версии ДЭК на 10000 поколениях тестовой модели. В таблице 3.2.1 приведены результаты расчета без оптимизации и с оптимизацией (расчёты проведены на ноутбуке с 6-ядерным процессором AMD Phenom).

Таблица 3.2.1. Результаты расчётов 10000 поколений в тестовой модели ДЭК

Число потоков	Без оптимизаций	1	2	4	6	8
Время счета, сек.	5294	4471	2437	1536	1237	1410

В OpenMP можно задавать различные стратегии для распределения итераций в цикле между параллельными потоками с помощью модификатора `schedule(type[, size])` для директивы компилятору `#pragma omp for`, находящейся в параллельном регионе. Type означает тип стратегии, size является необязательным параметром и задает размер блока итераций, передающегося потоку. Если параметр size не задан, размер блока будет приблизительно равен общему количеству итераций в цикле, разделенному на

количество потоков, обрабатывающих этот цикл. В OpenMP присутствуют следующие стратегии:

- **Static.** Цикл делится на блоки размером `size` или стандартного размера, примерно равного результату деления количества итераций на количество потоков. Затем каждый поток обрабатывает свой блок итераций.
- **Dynamic.** Организуется внутренняя очередь, выдающая потокам блоки итераций размера `size` или стандартного размера, равного одной итерации. По окончании работы над блоком, поток получает новый блок. Потоки будут получать новые блоки, пока итерации не закончатся.
- **Guided.** Стратегия, похожая на **Dynamic**. Отличие состоит в том, что размер блока, выдаваемого потоку, сначала большой и затем уменьшается при следующей выдаче блока, что может помочь сбалансировать нагрузку на потоки. Размер блока уменьшается вплоть до размера `size`, или до стандартного размера, равного одной итерации.
- **Auto.** При данной стратегии компилятор сам выбирает стратегию распределения итераций, параметр `size` не задается.
- **Runtime.** Стратегия распределения итераций задается пользователем во время выполнения программы с помощью переменной окружения `OMP_SCHEDULE`. Выбранной стратегией может быть **Static**, **Dynamic** или **Guided**, параметр `size` не задается.

Были протестированы стратегии **Static**, **Dynamic**, **Guided** и **Auto**. Наилучший результат показала стратегия **Static**, близкий к нему – **Auto**. Стратегии **Dynamic** и **Guided** показали схожие результаты, худшие, чем остальные стратегии.

Возможной причиной худших показателей у динамических стратегий является тот факт, что в каждой итерации циклов, подвергнутых распараллеливанию, находится относительно небольшое количество вычислений, при этом, количество итераций со временем (при прохождении

итераций основного цикла) увеличивается, из-за чего тратится лишнее время на добавление потоков в очередь для получения новых блоков, являющееся значительным. Стратегия `guided` показала худший результат, так как количество итераций постоянно увеличивается, а размер блоков уменьшается, из-за чего тратится дополнительное время.

Стратегия `Auto` показала результат, близкий к результату стратегии `Static`. Возможной причиной этого является то, что компилятор сам мог выбрать стратегию `Static` со стандартным размером блоков.

Позднее было принято решение отказаться от использования `Intel C++ Compiler` с `OpenMP` версии 3.0 с целью продолжить разработку на `Microsoft Visual Studio`, следовательно, необходимо было переписать программу, используя возможности `OpenMP` версии 2.0.

Проблемы в коде на `OpenMP 2.0` возникали при обработке итератора, использовавшегося в цикле, поэтому было решено имитировать стратегию статического распределения итераций между параллельными потоками, так как в предыдущей версии программы стратегия `Static` показала наилучший результат.

Мужские и женские особи в ДЭК хранятся отдельно в списках `std::list`. В новой версии программы каждый из этих списков разрезается на отдельные примерно равные по размеру списки. Количество новых списков равно количеству `n` параллельных потоков (по `n` списков для мужских и женских особей). Затем создаются два вектора `std::vector` для списков мужских и женских особей, в которые помещаются эти списки.

После этого в параллельном регионе начинается параллельная обработка цикла `for` с помощью директивы `#pragma omp parallel for` и модификатора `num_threads(n)`, где `n` – количество потоков. При этом, так как количество списков равно количеству потоков, реализуется статическая стратегия распределения итераций, так как каждый поток

работает со своим списком и потоки не конфликтуют друг с другом. Номер каждого потока и соответствующего ему списка можно узнать во время выполнения блока с помощью функции `omp_get_thread_num()`.

Данная версия программы показывает лучшие результаты, чем предыдущая и позволяет продолжать работу над программным комплексом с помощью уже имеющегося программного обеспечения Microsoft Visual Studio.

Тестирование программы производилось на разных машинах, таких как компьютер с 2-ядерным процессором Intel Core i3 2.93 GHz с включенным Hyper-Threading, 8-ядерным компьютером с процессором AMD FX-8350 4GHz с включенным Hyper-Threading, а также на кластере Сибирского Суперкомпьютерного Центра СО РАН. Ускорение программы рассчитывалось как отношение времени выполнения последовательной версии программы ко времени выполнения параллельной версии.

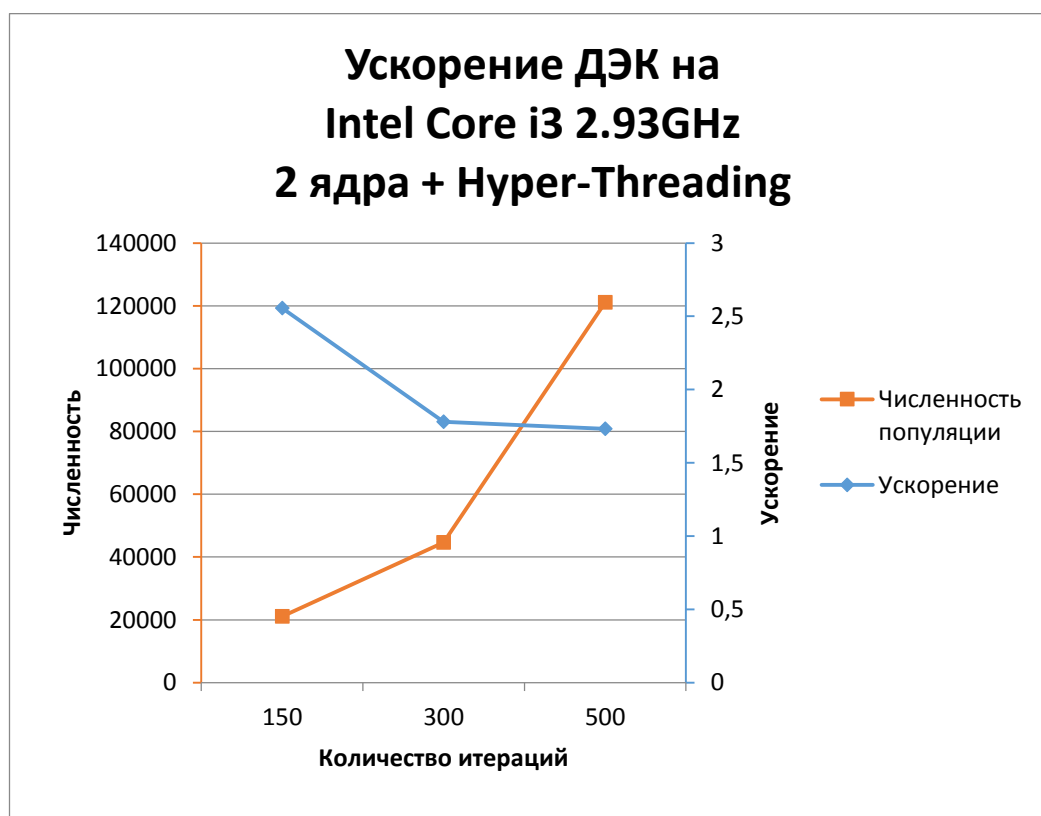


Рисунок 3.2.3. Уровень ускорения расчётов модели в зависимости от численности популяции (Intel Core i3 2.93GHz).

Первым проводилось тестирование на компьютере с процессором Intel Core i3 (Рисунок 3.2.3). Начальные условия для всех тестов были одинаковые, с различием только в количестве задействованных потоков. Для 2-ядерного компьютера с Hyper-Threading количество потоков было равно 4. Значения ускорения были получены в зависимости от итераций, так как с каждой итерацией количество особей увеличивалось экспоненциально (из-за превышения коэффициентом рождаемости коэффициента смертности), поэтому было необходимо рассмотреть, как будет зависеть ускорение от увеличения количества особей (численность популяций которых на момент окончания процесса моделирования достигала порядка 120000).

Как видно на рисунке 3.2.3, ускорение достигло уровня 2.5 раза, затем при увеличении численности популяций снизилось до уровня в 1.75 раз и оставалось на том же уровне. Уменьшение быстродействия по сравнению с максимально достигнутым ускорением может быть связано с низкой производительностью процессора, из-за которой при увеличении численности особей процессор больше не успевал с такой же скоростью обрабатывать вычисления, производимые в параллельных потоках.

Затем проводилось тестирование на более мощном компьютере с процессором AMD FX-8350 (Рисунок 3.2.4). Начальные условия были такие же, количество потоков для 8-ядерного компьютера с Hyper-Threading было равно 16.

На рисунке 3.2.4 можно видеть отличие от первого случая: ускорение увеличивается при увеличении численности популяций, то есть, при увеличении объема задачи. Замечено, что некоторые задачи, в которых объем параллельной части моделирования увеличивается при увеличении объема самой задачи, являются более масштабируемыми и показывают больший прирост производительности. То есть, такие задачи «раскрывают» свою производительность при увеличении своей сложности на достаточно производительном компьютере.

ПК ДЭК является примером подобного типа задач: на рисунке 3.2.4 видно, что программа достигает четырехкратного ускорения на минимальном количестве итераций, и быстродействие улучшается с увеличением количества особей в популяции. Ускорение достигает 13 раз для полного процесса моделирования.

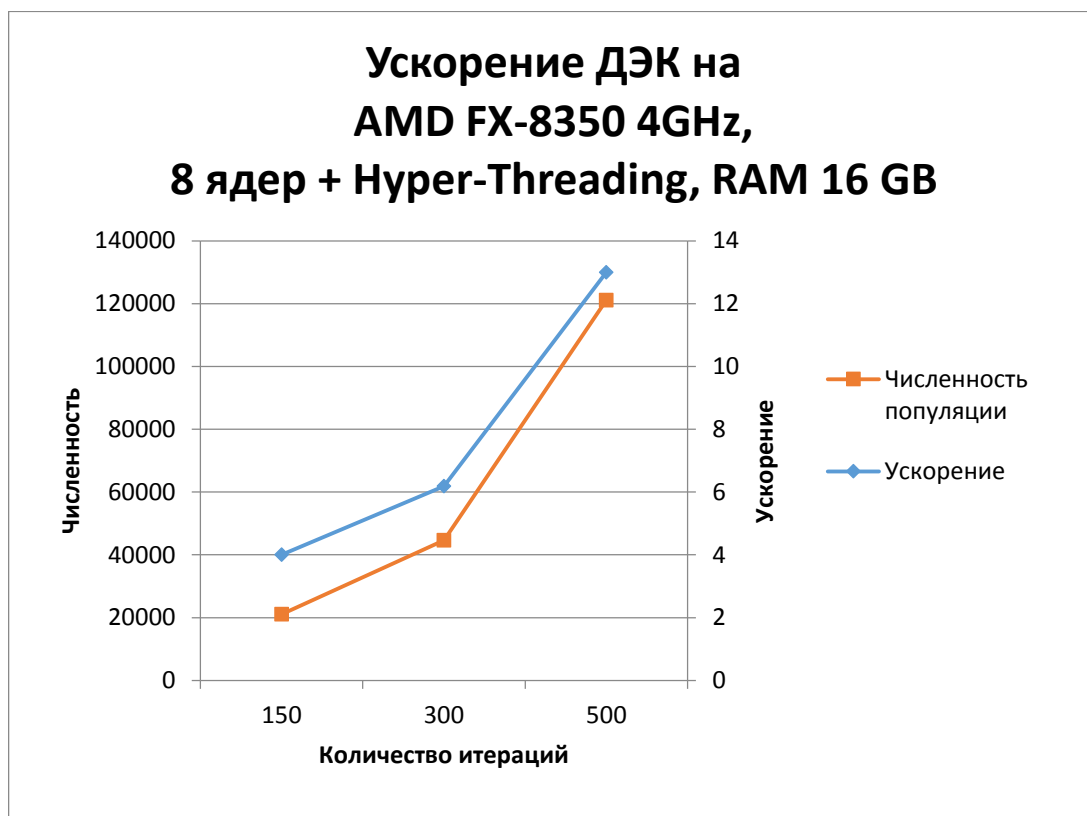


Рисунок 3.2.4. Уровень ускорения расчётов модели в зависимости от численности популяции (AMD FX-8350 4GHz).

Последним проводилось тестирование на кластере ССКЦ СО РАН (Рисунок 3.2.5), для моделирования были выделены 15 4-ядерных процессоров Intel Xeon X5560. Начальные условия были такие же, число потоков для 15 4-ядерных процессоров было равно 60.

На рисунке 3.2.4 можно видеть похожую на предыдущий случай картину. Это обусловлено тем, что для моделирования используется достаточно мощная система, которая позволяет задаче показать свою производительность при увеличении сложности. При этом, видно, что из-за большего количества параллельных потоков программа показывает большее ускорение, которое начинается с 15 раз для минимального количества итераций и достигает 68 раз

при максимальном количестве итераций во время моделирования при предыдущих условиях. Затем ускорение перестает увеличиваться, находясь для последующих итераций в пределах 65 раз, на рисунке прекращение увеличения производительности можно увидеть при увеличении максимального количества итераций в процессе моделирования, численность популяций при этом достигает порядка 200000.

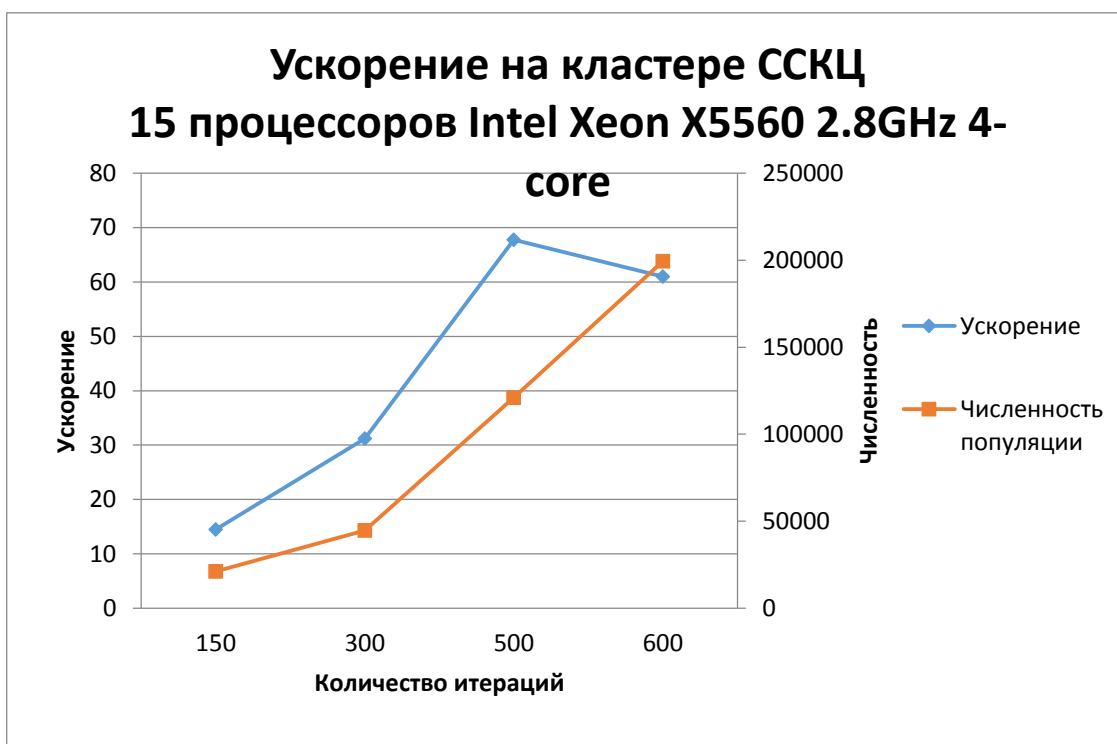


Рисунок 3.2.5. Уровень ускорения расчётов модели в зависимости от численности популяции (Intel Xeon X5560 2.8GHz 4-core).

Cuda версия ДЭК

Была реализована Cuda-версии ДЭК, также направленной на ускорение функции размножения. Отметим, что одной из особенностей технологии Cuda является ограничение использования объектно-ориентированного подхода программирования в GPU-части. Вследствие этого, широкое использование основного архитектурного паттерна ПК ДЭК, подразумевающего независимую разработку слоёв (классов, описывающих данные) и стратегий (классов, описывающих правила изменения слоёв), и основанную на *динамическом полиморфизме*, в текущей версии технологии

Cuda весьма затруднено. Практический вывод, сделанный нами в результате разработки Cuda-модулей, заключается в том, что для эффективного использования данной технологии необходима совместная и зависимая реализация Cuda-версий классов всех стратегий и слоёв, задействованных в конкретной модели ПК ДЭК. Фактически, это означает необходимость перехода к *статическому полиморфизму* и изменение архитектуры программного комплекса. Для упоминавшейся выше модели популяций древних людей в Cuda-версии были реализованы следующие функции:

```
1) void breedCuda(int generations_on_cuda, int generations_full, double &counttime);
```

где

`generations_on_cuda` – число поколений, выполняющихся подряд на одном раунде запуска генератора случайных чисел;

`generations_full` – число поколений за 1 цикл передачи данных на видеокарту. Используется 100 по умолчанию, т.к. каждые 100 поколений на ЦП происходит миграция (в данной конкретной модели);

`counttime` – переменная для профилирования времени выполнения процессов.

Данная функция осуществляет подготовку данных для расчета на видеокарте – преобразование из недоступных для Cuda структур в доступные, выделение и копирование памяти, расчет того, какую часть данных получит каждый поток.

```
2) __global__ void setup_kernel(curandState *state, unsigned long seed)
```

где

`*state` – массив состояний генератора случайных чисел для каждого потока;

`seed` – инициализирующее значение датчика случайных чисел;

Данная функция осуществляет инициализацию генератора случайных чисел для каждого потока. Функция выполняется ядром графического процессора.

```
3) __global__ void calculate(char *alph, char *m_DNA, char *f_DNA, int *numPeoples, double deathRate, double birthRate, int generations, curandState *state, int* devDies, int* devBorns, int* devMales, int* devFemales)
```

Данная функция выполняет моделирование процессов рождаемости, смертности и мутаций для всех популяций. Функция выполняется на графическом процессоре. Параметрами служат всевозможные данные о популяция для каждого потока (в функции происходит «развёртка» данных из инкапсулированного вида в «открытые» данные).

Процесс моделирования в Cuda-версии осуществляется следующим образом. В основном цикле ПК ДЭК вызывается функция `breedCuda` для всех популяций, затем происходит миграция и процесс повторяется. Внутри данной функции происходит инициализация всех необходимых данных и осуществляется цикл выполнения функции `calculate`. В предельном случае планировался однократный вызов данной функции, в которой одновременно рассчитывалось бы все 100 итераций до миграции. Но, ввиду большой вычислительной сложности задачи, расчёт на видеокарте NVIDIA GTX 570 прекращался из-за перегрузки данными с ростом числа особей в каждой популяции. Для исправления данной проблемы функция вызывается не 1 раз, а `generations_full/generations_on_cuda` раз. После всех вычислений происходит копирование данных на ЦП и обратное присваивание полученных результатов. Тестирование последовательной Cuda-версии дало следующие результаты (табл. 3.2.2):

Таблица 3.2.2. Сравнение времени выполнения расчётов модели с использованием последовательной Cuda-версии ДЭК, исходной и OpenMP версий

Тип версии	Время выполнения, сек.
Исходная версия, без распараллеливания	558
Параллельная версия с использованием OpenMP	181
Cuda-версия, 1 итерация за вызов <code>calculate</code>	742
Cuda-версия, 2 итерации за вызов <code>calculate</code>	726
Cuda-версия, 5 итераций за вызов <code>calculate</code>	718
Cuda-версия, 10 итераций за вызов <code>calculate</code>	715

Как можно видеть, последовательная версия на графическом процессоре оказалась даже медленнее исходной непараллельной версии. Различие между вызовами `calculate` 100 раз или же 10 раз по 10 поколений за расчет объясняются тем, что между каждым вызовом функции приходится заново вызывать функцию `setup_kernel`, иначе будут переданы те же самые случайные числа.

Было проведено профилирование, которое показало, что основные затраты времени уходят на работу с памятью и копированием данных. С другой стороны, дополнительно был проведен замер времени с помощью обработчика событий `cuda`, который показал, что 700 секунд уходит именно на процедуру `calculate`.

Были предприняты попытки написания нескольких многопоточной Cuda-версии версий, которые укрепили нас в мнении, что портирование на Cuda программ, подобных ПК ДЭК, описывающих сложные модели иерархических (в нашем случае биологических) систем, должно сопровождаться кардинальной модификацией архитектуры программы в

сторону отхода от классической объектно-ориентированной схемы динамического полиморфизма в сторону как минимум полиморфизма статического. Отметим при этом, что технологии MPI и OpenMP не требуют такой кардинальной перестройки архитектуры и могут рассматриваться как более подходящее решение для разработки подобных программных комплексов, по крайней мере, с точки зрения временных затрат на разработку.

3.2.2. Графический интерфейс пользователя ПК ДЭК

Был разработан и программно реализован графический интерфейс пользователя (ГИ) ПК ДЭК. Для создания ГИ использовались платформа Qt и библиотека QCustomPlot (<http://www.qcustomplot.com/>). На данный момент доступны два варианта использования ГИ ПК ДЭК:

1. Интерактивный режим. Позволяет всесторонне исследовать отдельную модель, изменять свойства агентов (в том числе, и во время моделирования), детально исследовать все характеристики модели, а также наблюдать анимированную динамику модели (пример на рис. 3.2.6).
2. Пакетный режим. Позволяет рассчитывать серии моделей с одинаковыми стартовыми условиями, или на заданном множестве стартовых условий без необходимости отдельного запуска расчётов для каждой модели. Позволяет набирать большую статистику расчётов, а затем анализировать различные статистические закономерности в рассчитанных моделях.

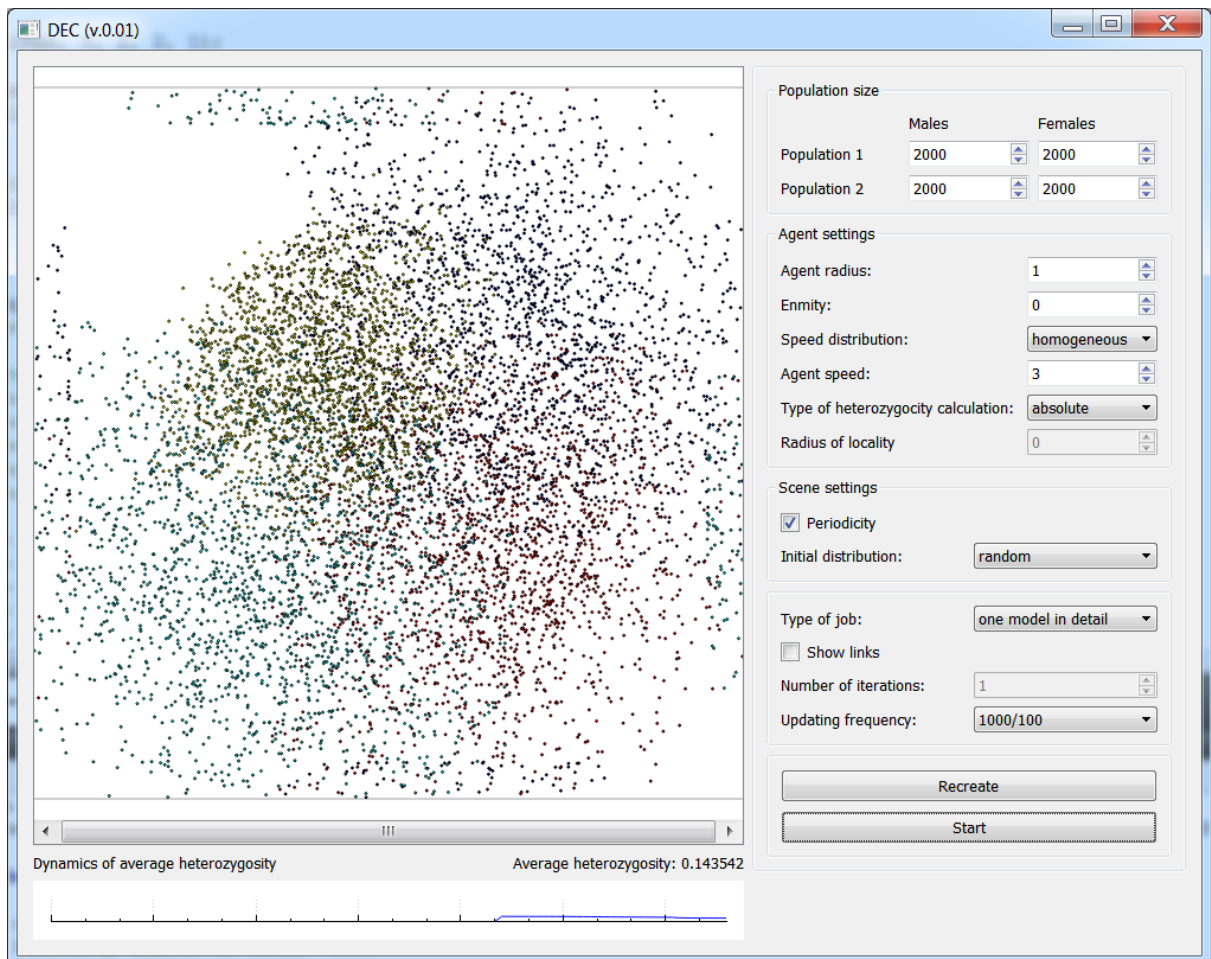
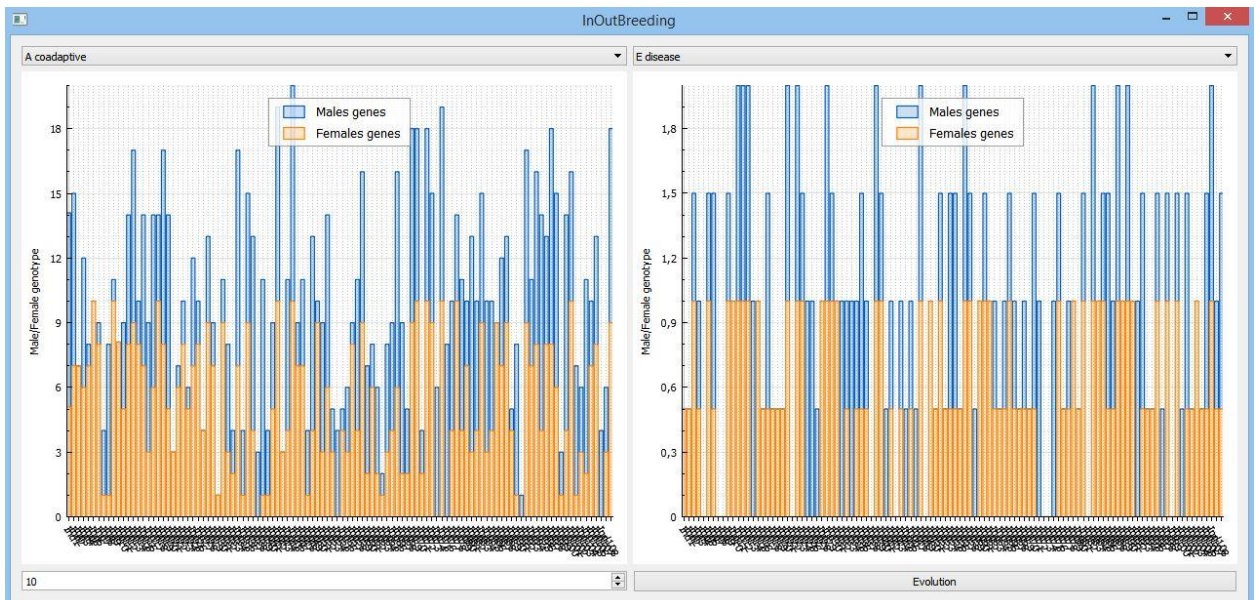


Рисунок 3.2.6. Графический интерфейс ПК ДЭК. Модель смешивания двух популяций в результате миграций.



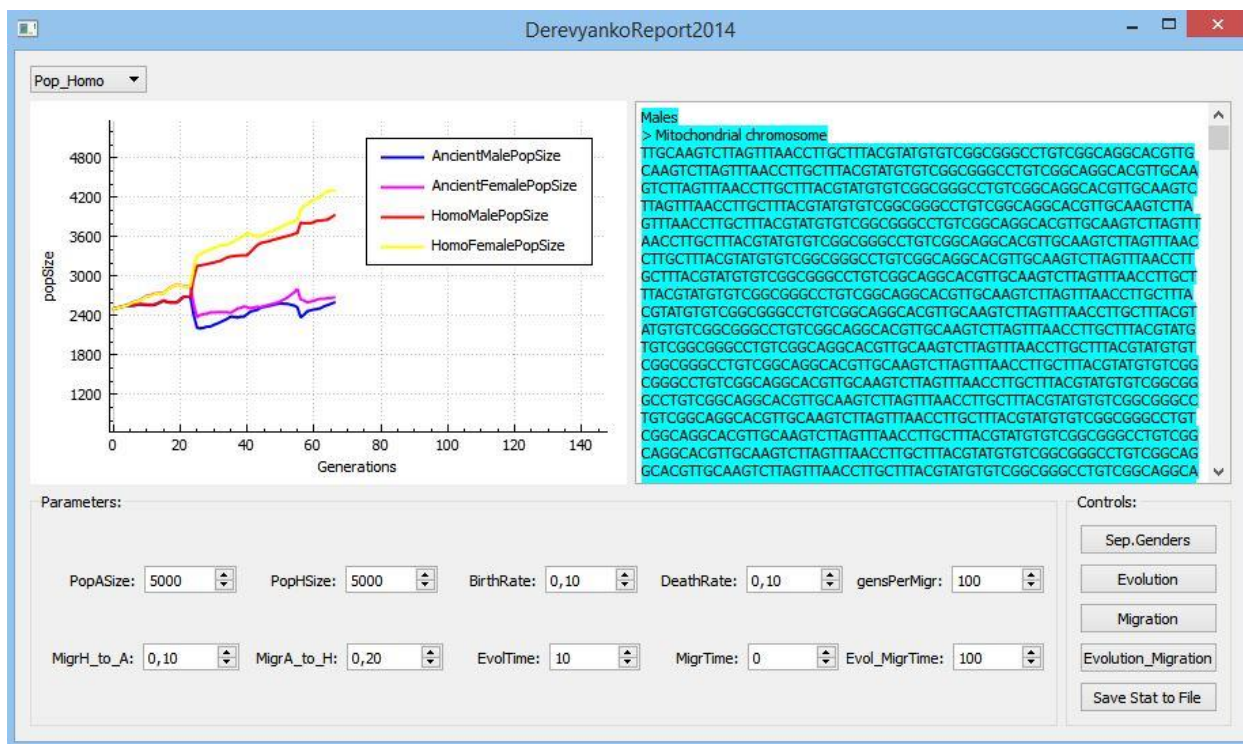


Рисунок 3.2.7. Примеры графического интерфейса ПК ДЭК. Сверху – модель инбридинга-аутбридинга; снизу – модель смешивания двух популяций в результате миграций.

На рисунке 3.2.7 приведенный примеры графического интерфейса ПК ДЭК на моделях, описанных в следующей подглаве – модели инбридинга-аутбридинга (сверху, подробнее см. раздел 3.3.2) и модели смешивания двух популяций в результате миграций (снизу, подробнее см. раздел 3.3.3).

3.3. Содержательные биологические результаты, полученные с использованием ПК ДЭК

С помощью ДЭК были построены и исследованы содержательные биологические модели, включающие в себя как классические модели генетики популяций, разработанные с целью верификации программного комплекса, так и ранее не исследованные модели популяций древних людей и модели инбридинга и аутбридинга в популяциях человека.

3.3.1. Верификация ПК ДЭК на классических задачах генетики популяций

Верификация вычислительного ядра ДЭК проводилась на классических задачах генетики популяций. Одной из таких задач является моделирование фундаментального принципа естественного отбора Фишера, традиционно называемого в генетике популяций теоремой Фишера, формулируемого в следующем виде: «В бесконечной панмиктической популяции, в которой происходит постоянный отбор генотипов по аллелям одного гена, скорость возрастания средней приспособленности популяции пропорциональна генетической дисперсии приспособленности этой популяции» (Ли, 1978; Фишер, 2011). Моделировалась популяция диплоидных модельных организмов. Диплоидный геном особи состоял из двух генов (один экземпляр от отца, другой от матери). Единственный фенотипический признак в данной модели, тождественный приспособленности, определялся численными значениями генов (имевших в модели смысл ферментативной активности белков, экспрессируемых с данных генов) по принципу доминантности – признак определялся как $\max(A1, A2)$, где $A1$ и $A2$ – значения аллелей (вариантов генов), полученных от отца и матери, соответственно. От приспособленности особи прямо пропорционально зависела вероятность оставить потомство. При этом вероятность не оставить потомства у наиболее приспособленной особи, равно как и вероятность оставить потомство у наименее приспособленной особи, не были нулевыми, т.е. имел место так называемый «мягкий отбор». В ходе эволюции могли происходить мутации, повышающие или снижающие приспособленность. Тем не менее, основной рост средней приспособленности популяции в основном обеспечивался не мутационным, а половым процессом (ср. кривые средней и максимальной приспособленности на рис. 3.3.1). Как видно из рисунка 3.3.1, средняя приспособленность популяции монотонно возрастает, и на длительном промежутке времени (200 поколений) линейно аппроксимируется. Рисунок 3.3.2 показывает, что с течением времени и ростом средней приспособленности особей генетическая дисперсия популяции

(эквивалентная в данной модели приспособленности) снижается. Таким образом, рассмотренная нами модель, созданная с помощью ПК ДЭК, в целом удовлетворяет теореме Фишера, однако, с учётом конечности моделируемой популяции, не полностью.

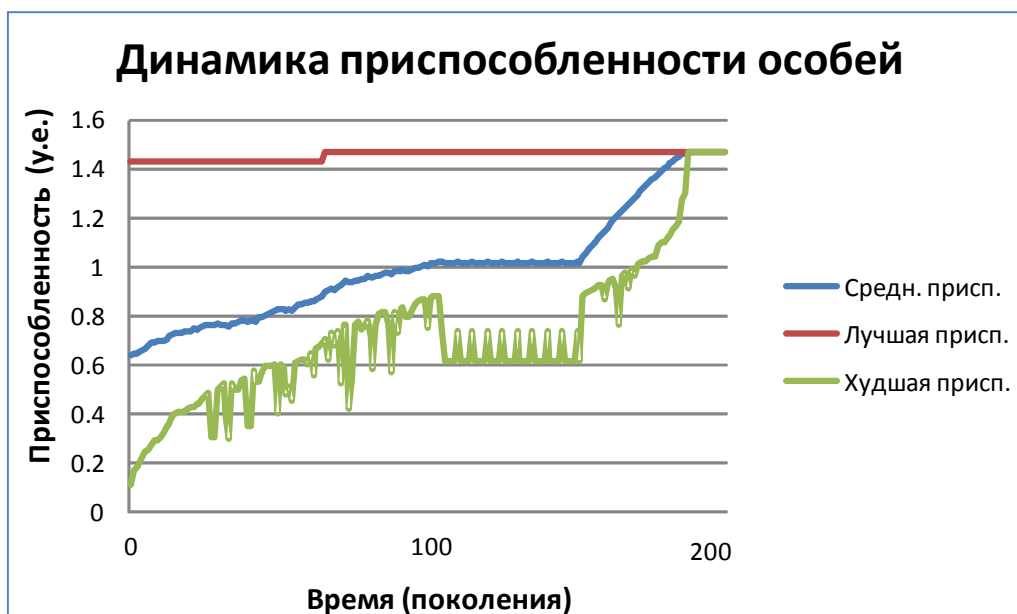


Рисунок 3.3.1. Динамика приспособленности особей в популяции: средняя приспособленность, максимальная (лучшая) приспособленность, минимальная (худшая) приспособленность особей в популяции.

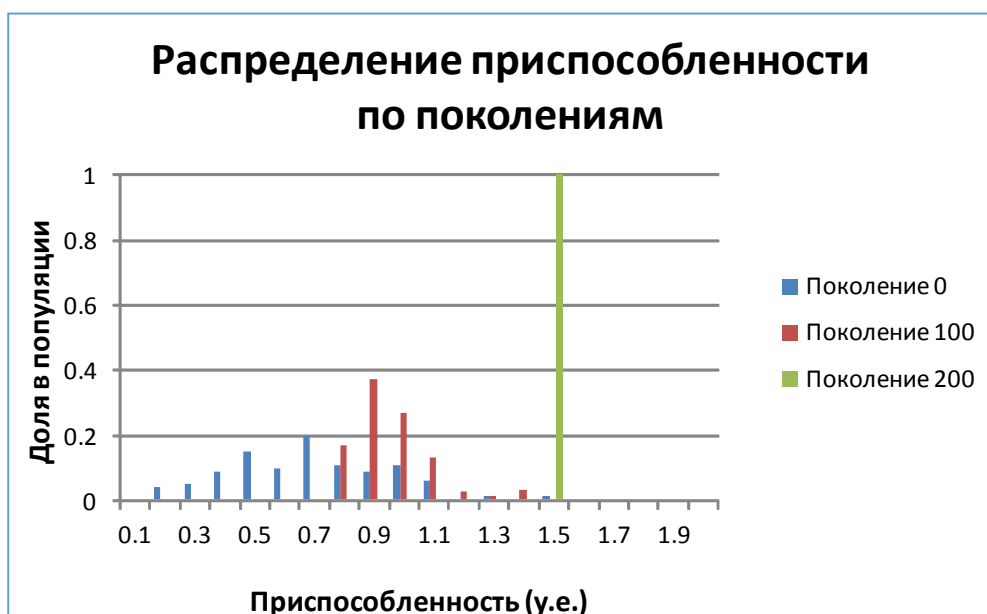


Рисунок 3.3.2. Распределение приспособленности особей популяции по поколениям (0, 100 и 200 поколения).

В целом, верификация ПК ДЭК показала биологическую адекватность как выбранной методики, так и её программной реализации.

3.3.2. Исследование инбридинга и аутбридинга в популяциях диплоидных организмов

С помощью ДЭК была разработана и исследована компьютерная модель процессов инбридинга и аутбридинга в популяции. Рассматривалась популяция диплоидных модельных организмов. Гаплоидный геном особи состоял из одной хромосомы, которая, в свою очередь, содержала шесть генов: А, В, С, D, E, F. Фенотип особи определялся двумя сложными признаками – фертильностью и устойчивостью к болезни. Первый признак определялся простыми количественными признаками А и В, кодируемыми, соответственно, генами А и В. При определении этих простых признаков рассматривались разные схемы доминирования: полное доминирование, кодоминирование, а также материнский эффект. Фертильность же определялась с учётом кумулятивного действия признаков А и В (мультипликативный эффект), при этом ландшафт значений для фертильности имел два пика (повышенной приспособленности, см. рис. 3.3.3). Признак устойчивости к болезни определялся простыми количественными признаками С, D, E, F, кодируемых соответствующими генами. При этом также рассматривались разные схемы доминирования (см. выше), и, кроме того, аддитивные и мультипликативные эффекты.

Моделировалось функционирование популяции на протяжении 100 поколений. Генотипы популяций генерировались стохастическим образом. Был рассмотрен эффект миграции одной популяции (Р2) в ареал обитания другой популяции (Р1). Особи популяций Р1 и Р2 находились на значительном генетическом расстоянии, в частности, у них наблюдались различающиеся коадаптации генов А и В (соответствующие разным пикам приспособленности на рис. 3.3.3).

В результате скрещиваний особи Р1 и Р2 образовывались гибриды, в которых, благодаря рекомбинации, могло происходить нарушение

коадаптивного комплекса аллелей генов А и В, что, в свою очередь, приводило к снижению фертильности (и, как следствие, приспособленности) этих гибридов. Подобный отрицательный эффект аутбридинга наблюдался экспериментально в различных человеческих популяциях. Долговременный отбор (в течение 100 поколений) в результирующей популяции P1+P2, как правило, приводил к «вымыванию» аллелей генов А и В, характерных для популяции P1. При этом скорость вымывания этих аллелей, а также генетическое разнообразие в результирующей популяции по генам А и В зависело как от типа доминирования для этих генов, так и от степени инбридинга по генам С, D, E и F (незначительная степень инбридинга на рис. 3.3.3а, 3.3.3б, 3.3.3в; значительная степень инбридинга на рис. 3.3.3г, 3.3.3д, 3.3.3е).

Наиболее сильно изменение степени инбридинга влияет на генетическое разнообразие результирующей популяции в случае материнского эффекта для генов А и В. В отсутствие инбридинга результирующая популяция не содержит аллелей популяции P2 (по генам А, В), однако генетическое разнообразие по этим генам в ней по-прежнему сохраняется (рис. 3.3.3в). В случае давления инбридинга генетическое разнообразие по этим генам меньше, несмотря на то, что, вообще говоря, после элиминации аллелей популяции P2 (по генам А, В), отбор в популяции шёл в основном по генам С, D, E и F. В случае полного доминирования и кодоминирования степень инбридинга влияет на генетическое разнообразие результирующей популяции значительно меньше: оно невелико и в отсутствие давления инбридинга, и в его присутствии (ср. рис. 3.3.3а и 3.3.3г, также 3.3.3б и 3.3.3д).

Таким образом, показано, что эволюция коадаптивных генов может идти под воздействием как инбридинга, так и аутбридинга. Давление инбридинга по одной группе генов (в модели гены С, D, E и F) может влиять на эволюцию и генетическое разнообразие других генов, в том числе коадаптивных (гены А

и В). Также на эволюцию коадаптивных генов влияет тип доминирования при их взаимодействии (в диплоидном геноме).

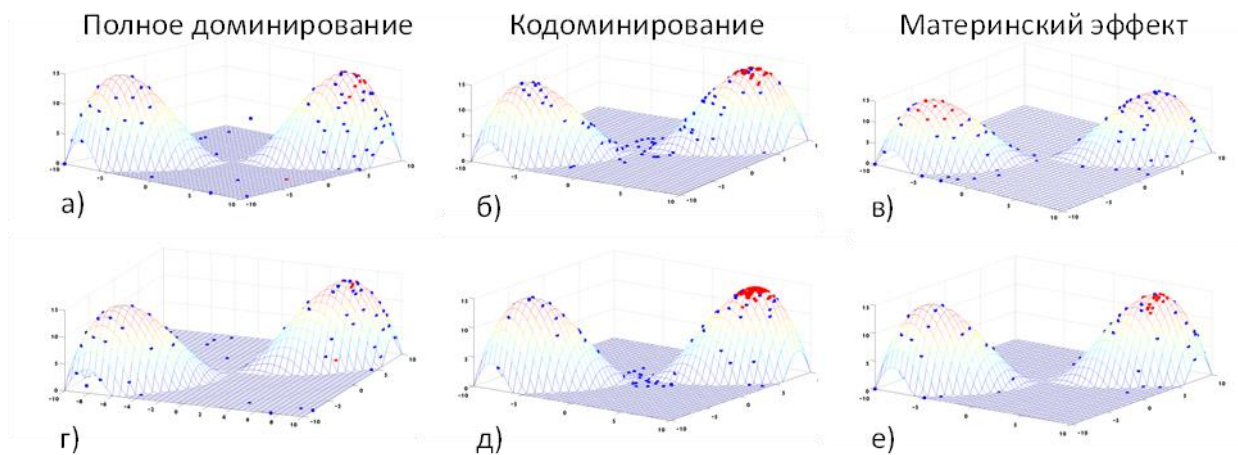


Рисунок 3.3.3. Генетическое разнообразие популяции на первом поколении (синие точки) и на сотом поколении (красные точки) по генам А и В (оси x и y, соответственно) при полном доминировании, кодоминировании и материнском эффекте. Ось z – значение сложного признака фертильности. Поверхность – ландшафт приспособленности. Верхний ряд (а, б, в) – в популяции слабое давление инбридинга; нижний ряд (г, д, е) – сильное давление инбридинга.

3.3.3. Моделирование популяционно-генетических и миграционных процессов в популяциях древних людей на территории Евразии

Одним из основных факторов, сильно влияющих на оценку скорости фиксации в митохондриальной ДНК (мтДНК), является генетическое смешение при гибридизации. Отцовская мтДНК в редких случаях может проникать в яйцеклетку во время оплодотворения и существенно изменять материнскую мтДНК путем рекомбинации. Такой тип эволюционных событий можно легко проследить вплоть до самых ранних стадий эволюции человека, поскольку эти события очень редки, в отличие от материнского наследования мтДНК. Для того чтобы найти наиболее вероятные пути генетического смешения, мы исследовали наследование мтДНК человека, основываясь на всех имеющихся в настоящее время полных

митохондриальных геномах. Были извлечены 16810 полных последовательностей мтДНК современных людей из PhyloTree.org (van Oven, Kayser, 2009) и 5 полных мтДНК древних неандертальцев и Денисовского человека из базы данных Генбанк. Затем с помощью программы IMa2 (Heu, 2010) была восстановлена картина наиболее вероятных миграций древних человеческих популяций. Анализ миграций показал интенсивную гибридизацию ($P < 0.001$) между современными и древними людьми, что согласуется с ранее опубликованными данными об эволюции ядерных генов (Reich et al., 2011).

Интересно, что поток митохондриальных генов от древних к современным людям происходил тремя путями: от древних людей к носителям гаплотипов L0 и L5 и к потомкам гаплотипа L2. Обнаружен также обратный поток современных аллелей генов мтДНК к древним людям: от потомков гаплотипов L2 и L0. Таким образом, было показано, что почти все митохондриальные гаплотипы современного человека обладают генными аллелями древних людей. Таким образом, эволюция *Homo sapiens* носит скорее ретикулярный, чем филетический характер, и, следовательно, возникает вопрос об изменении таксономического статуса *Homo neanderthalensis* и *Homo denisova* на *Homo sapiens neanderthalensis* и *Homo sapiens denisova*, соответственно.

На основе полученных оценок популяционно-генетических параметров с помощью ПК ДЭЖ была построена серия компьютерных моделей эволюции предков современных людей на территории Евразии. Проанализировано несколько возможных эволюционных сценариев, представленных на рис. 3.3.4.

Общая схема сценария выглядела следующим образом:

1. Рассматривались популяции двуполой особей, геном которых моделировался в виде последовательности ДНК длиной 360 нуклеотидов.

Наследование генома происходило по материнской линии. Эта последовательность являлась моделью варибельного участка мтДНК.

2. Генерировались начальные популяции f , G (здесь и далее – по рис. 1). Исходно в одной популяции все особи обладали одинаковым генотипом (для популяции f использовался фрагмент Homo sapiens isolate Nairobi-020 control region, gb|AY632914.1; для популяции G в этом фрагменте были произведены 4 замены). Исходный размер популяций варьировался, в зависимости от сценария, от 1000 до 10000 особей.

3. В течение 7000 поколений (исходя из оценки 25 лет на одно поколение) моделировались популяционные и миграционные процессы.

a. Коэффициенты рождаемости (b) и смертности (d) варьировались, в зависимости от сценария, таким образом, чтобы коэффициент прироста популяции составлял от $1+10^{-4}$ до 1.1.

b. Мутации возникали в геномах с частотой 10^{-6} на поколение; моделировались только точковые замены нуклеотидов.

c. Рассматривалось два варианта миграций особей между популяциями: а) миграции происходят относительно часто (1 раз в 100 поколений), но относительно малочисленными группами (1% популяции); б) миграции происходят относительно редко (1 раз в 1000 поколений), но большими группами (10% популяции)

4. Из популяции f выделялась субпопуляция e (50% особей), и в дальнейшем моделировалась эволюция только этой популяции и её потомков. В течение следующих 10 поколений моделировались только популяционные процессы (без миграций) в популяциях e и G .

5. Моделировалась дивергенция популяции e на популяции F и d (10% и 90%, соответственно). В течение следующих 50 поколений моделировались только популяционные процессы в популяциях d , F , G .

6. Моделировалась дивергенция популяции d на популяции E и c (10% и 90%, соответственно). В течение следующих 100 поколений моделировались только популяционные процессы в популяциях c, E, F, G.

7. Моделировалась дивергенция популяции c на популяции D и b (10% и 90%, соответственно). В течение следующих 100 поколений моделировались только популяционные процессы в популяциях b, D, E, F, G.

8. Моделировалась дивергенция популяции b на популяции C и a (10% и 90%, соответственно). В течение следующих 200 поколений моделировались только популяционные процессы в популяциях a, C, D, E, F, G.

9. Моделировалась дивергенция популяции a на популяции A и B (50% и 50%, соответственно). В течение следующих 2540 поколений моделировались как популяционные, так и миграционные процессы в популяциях A, B, C, D, E, F. Рассматривались взаимные миграции особей между популяциями A и B, а также односторонние миграции их популяции A в популяции C, D, E и F. Частоты миграций и их интенсивность (количество мигрантов) варьировались в зависимости от эволюционного сценария.

10. В шагах 4-9 при дивергенции популяций в минорной субпопуляции демографические параметры сохранялись прежними, а в мажорной субпопуляции изменялись таким образом, что скорость её прироста увеличивалась на 10%

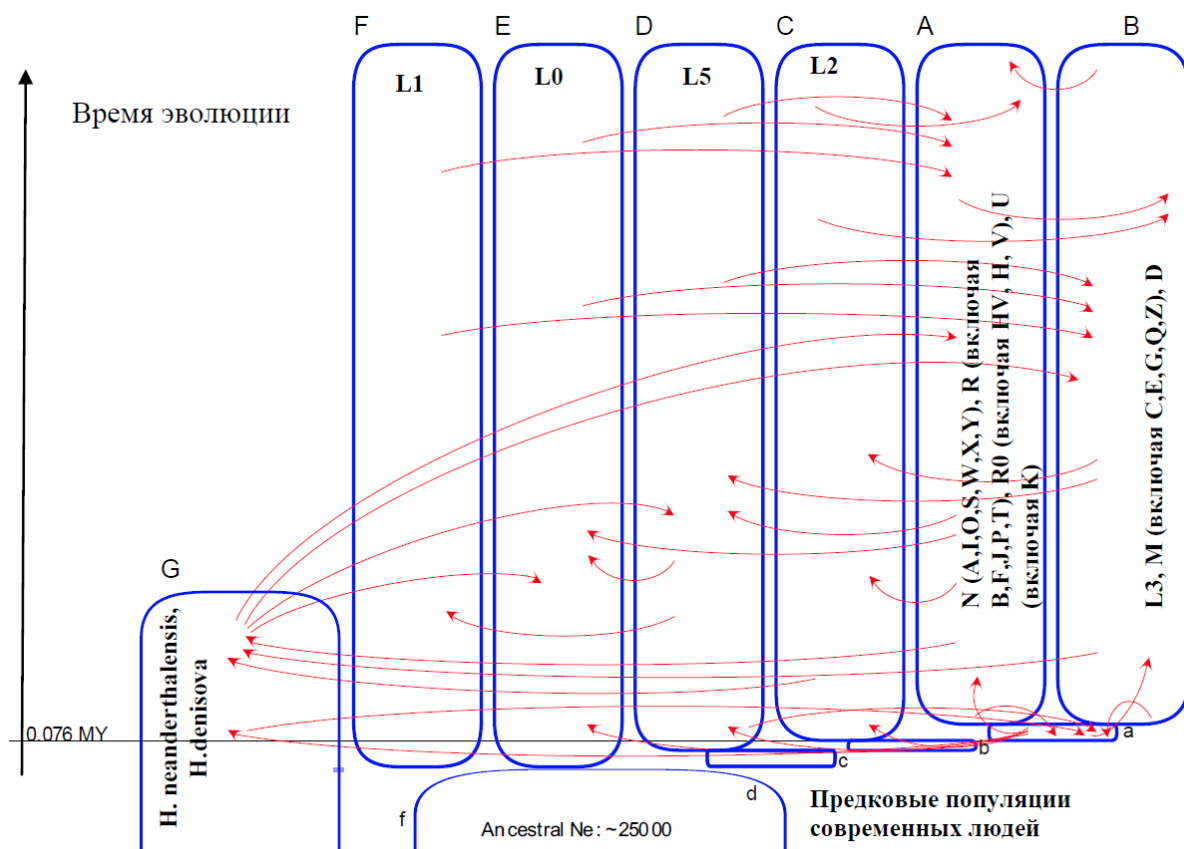


Рисунок 3.3.4. Схема миграций в популяциях человека, реконструированная на основе последовательностей мтДНК (По: (Gunbin et al., 2012)).

По итогам моделирования в каждой из субпопуляций мы анализировали последовательности фрагментов мтДНК женских особей. Одним из основных выводов из нашего анализа может служить оценка влияния миграций небольших групп особей (1% популяции) на изменение генотипа мажорной популяции: такая доля особей-мигрантов практически не изменяет частот аллелей в мажорной популяции особей. Модельный сценарий приведён в Приложении 4.

3.3.4. Модель эволюции генной сети с отрицательными обратными связями при разных типах отбора

Модель эволюции генной сети с регуляторным контуром, работающим по принципу отрицательной обратной связи (ОС), предложенная Н.А. Колчановым и И.Н. Шиндяловым (Колчанов, Шиндялов, 1991), была исследована с помощью ПК ДЭК.

Численность общей популяции (состоящей из особей без ОС – P1, и особей с ОС – P2) была постоянной и равнялась 1000. Размножение особей осуществлялось в соответствии с их приспособленностью, которая зависела от концентрации белка X, в соответствии с соотношением

$$W(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-X_0}{\sigma_X}\right)^2} \quad (\text{ур. 3.3.1})$$

Концентрация белка X определялась следующим уравнением:

$$\frac{dX}{dt} = k_s - k_d X \quad (\text{ур. 3.3.2})$$

Где параметры определяются следующим образом:

$$k_d = \begin{cases} \frac{C_3}{E + C_4}, & \text{если } E \geq 0 \\ \frac{C_3}{C_4}, & \text{если } E < 0 \end{cases}$$

В системе без обратной связи $k_s = C_1$

В системе с обратной связью

$$k_s = \frac{C_5}{1 + (X/C_2)}$$

При моделировании эволюционного процесса параметры отдельных особей определялись наследованием (от «матери» и «отца») с учётом следующих факторов:

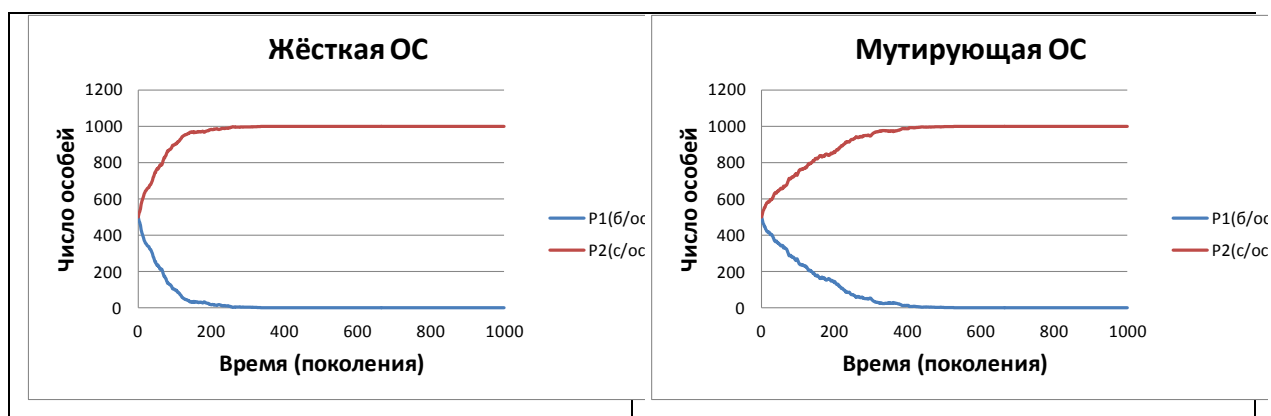
- тип доминирования:
 - кодоминирование: $A = (A1+A2)/2$;
 - полное доминирование: $A = \max(A1, A2)$;
 - материнский эффект: $A = A1$;
 - стохастическое возмущение параметров (изменение параметра на некоторую случайную величину σ из некоторого интервала).

Кроме того, в модель могли вноситься дополнительные возмущения:

- отклонение начальных данных от эталона на некоторую случайную величину σ из некоторого интервала
- варьирование «времени измерения признака» (в данном случае – времени расчёта T_{\max})

Были получены следующие результаты: в условиях стабилизирующего отбора (неизменные условия окружающей среды и неизменное значение X_0 в уравнении 3.3.1) преимущество получала субпопуляция с ОС. При этом, при возникновении движущего отбора (X_0 начинает изменяться на величину dX_0 каждую итерацию) преимущество постепенно переходило к субпопуляции без ОС (рис. 3.3.5).

В целом полученные результаты согласуются с результатами из (Колчанов, Шиндялов, 1991). Однако, в отличие от опубликованных результатов, в нашей системе не наблюдается полного вытеснения одной популяции другой. По-видимому, это связано с отличием в стратегии отбора. В нашей модели приспособленность влияет на выживаемость особей: определённая доля наименее приспособленных погибает, а их место в популяции (при её постоянном значении) занимают потомки выживших. При этом вероятность размножиться одинакова для всех выживших (вне зависимости от приспособленности).



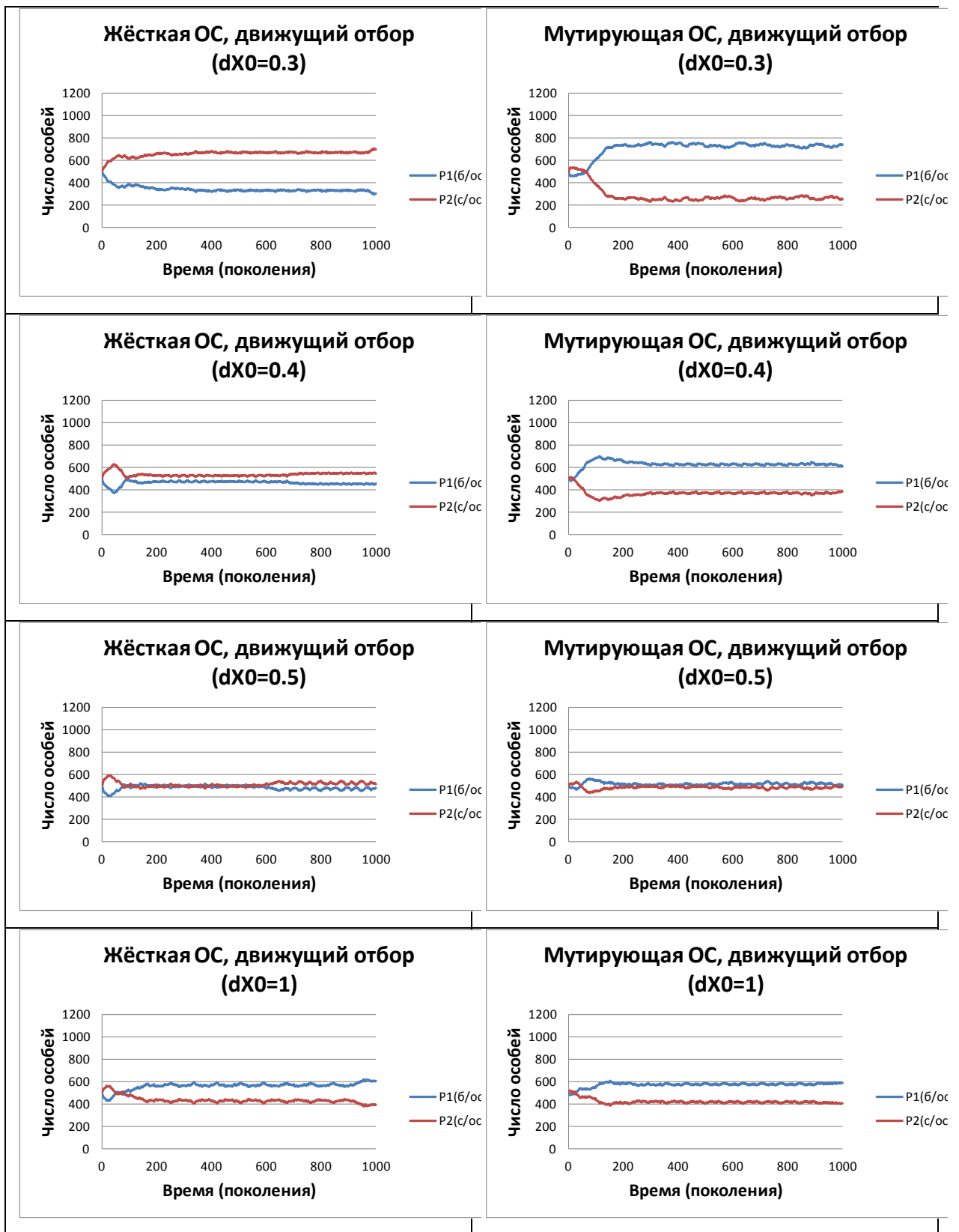


Рисунок 3.3.5. Популяционная динамика субпопуляций с обратной связью и без обратной связи в разных условиях отбора.

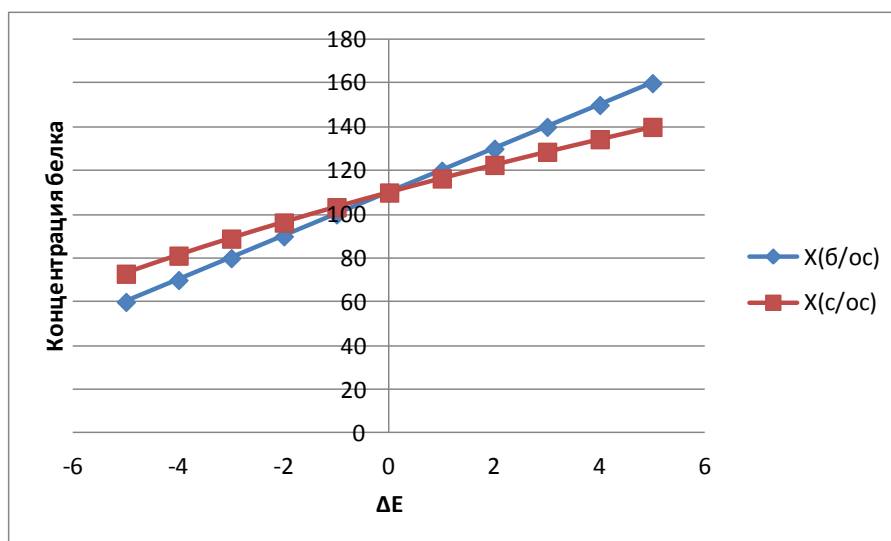


Рисунок 3.3.6. Зависимость концентрации белка (X) от его стабильности (E) в системах с учётом и без учёта обратной связи.

Другим интересным результатом является то, что было показано, что в популяции с обратной связью существенно дольше сохраняется биоразнообразие, что при меняющихся условиях среды может играть существенную роль, обеспечивая пул скрытой изменчивости, которая при смене условий будет вскрываться.

3.4. Заключение к главе 3

В данной главе описаны результаты работ по разработке методов моделирования функционирования и эволюции сложноорганизованных диплоидных организмов с половым размножением, а также их реализация в виде программного комплекса «Диплоидный эволюционный конструктор» (ДЭК).

С помощью ПК ДЭК моделируются эволюционные, популяционно-генетические и популяционно-экологические процессов в популяциях сложноорганизованных организмов (особей). В методике используется принцип многослойного моделирования, предложенный нами ранее для моделирования популяционной динамики простейших одноклеточных организмов и их эволюции, описанный в Главе 2 данной диссертации.

Были выделены основные объекты методики (слои модели) и процессы взаимодействия между объектами (стратегии модели).

Формально описаны слои: генетический, геномный, метаболический, фенотипический, популяционный и экологический слои, а также слой внутреннего состояния организма и его выживаемости; стратегии: мутационная, рекомбинационная, «генотип↔фенотип», «фенотип↔приспособленность» («организм↔среда»), размножения и отбора, и, наконец, миграционная.

Разработаны спецификации использования программного комплекса ДЭК. Программно реализованы: вычислительное ядро, консольный модуль управления процессом моделирования, включая транслятор специального языка моделирования ДЭК; графический интерфейс пользователя, система поддержки подключаемых модулей (плагинов) для классов стратегий, модули поддержки высокопроизводительных вычислений в ДЭК.

Было показано, что для эффективного использования технологии CUDA для параллелизации ДЭК на видеокарте необходима кардинальная переработка вычислительного ядра программного комплекса, поэтому в настоящий момент использование технологии CUDA для увеличения быстродействия является нецелесообразным.

Версия ПК ДЭК с использованием OpenMP была протестирована на различных компьютерах и показала значительное ускорение как на обычной машине (до 13 раз), так и на кластере Сибирского Суперкомпьютерного Центра СО РАН (до 60-70 раз), что позволит значительно сократить время, затрачиваемое на проведение расчетов в программе ДЭК.

Большой прирост производительности программного комплекса является важным результатом, так как это открывает перед биологами новые возможности по эволюционному моделированию с использованием агентно-ориентированных моделей, в том числе, по массовому моделированию, так как численность популяций особей в ДЭК может достигать десятков и сотен тысяч особей.

Верификация ПК ДЭК проведена на классических задачах генетики популяций, в частности, смоделирована теорема Фишера. Верификация показала применимость ПК ДЭК для решения задач моделирования эволюционно-популяционных процессов. Кроме того, с помощью ПК были исследованы различные сценарии популяционно-генетических и миграционных процессов в популяциях древних людей. Наконец, была исследована эволюция генных сетей с отрицательными обратными связями под действием разных типов отбора. В частности, установлено, что наличие регуляторных отрицательных обратных связей в генных сетях, отвечающих за приспособленность особей к условиям окружающей среды, способствует сохранению биоразнообразия популяции и формированию пула скрытой изменчивости – одного из важных факторов эволюции в изменяющихся условиях внешней среды.

4. Моделирование эволюции популяций с учётом социальных факторов

Важной отличительной особенностью популяций животных и, особенно, человека, по сравнению с популяциями микробов и других простых организмов, является возрастающая роль социальных факторов в возникновении, функционировании и эволюции этих сложных систем. Последние несколько лет ознаменовались ростом интереса к системным подходам в области социальной эпидемиологии (El-Sayed et al., 2012). Основными инструментами исследований в этой области являются анализ социальных сетей и агентно-ориентированное моделирование (АО модели или АОМ). Анализ социальных сетей позволяет ответить на вопросы о связи между топологией сети и скоростью распространения (и угасания) в ней информации различного рода, тогда как АО модели позволяют во многих случаях провести прямое имитационное моделирование популяции с возможностью явного задания правил генерации и обработки информации агентами. Хотя сетевой анализ подходит для исследования механизмов «социального заражения» (Buchanan, 2003), для его проведения требуется наличие большого объёма экспериментальных данных (что на практике встречается достаточно редко), что может ограничивать применение данных методов на практике. В этой ситуации использование АО моделей для оценки факторов, влияющих на распространение идей, оказывается наиболее естественным решением, особенно при рассмотрении факторов, относящихся к разным уровням биологической/социальной организации сообщества или его отдельных индивидуумов (El-Sayed et al., 2012).

Возникновение принципиально новых социальных явлений (таких как Интернет), а также возможности использования больших баз данных послужили толчком к применению численного анализа в социологии, в том числе с использованием математических моделей, заимствованных из физики. Общим вопросом таких моделей является возникновение порядка в

изначально неупорядоченной системе. Под порядком понимается консенсус, согласие, однородность (например, принятие общего решения, формирование иерархии, общего языка или культуры), под беспорядком – фрагментация или несогласие; ключевыми моментами формирования общего мнения, культуры или языка являются взаимодействие агентов и их стремление становится похожими друг на друга.

Перечислим некоторые известные модели социальной динамики, использующие идеи физики и агентное моделирование: модель избирателя (Clifford, Sudbury, 1973; Holley, Liggett, 1975), приобретающая известность как пример неравновесного стохастического процесса, аналитически разрешимого для любой размерности; модель с правилом большинства (Galam, 2002), реализующая принцип социальной инерции, согласно которому люди неохотно принимают изменения, если их не поддерживает явное большинство; модели социального влияния (Latané, 1981; Lewenstein et al., 1992; Nowak et al., 1990), рассматривающие зависимость силы взаимодействия от расстояния (как пространственного, так и расстояния в абстрактном пространстве мнений); модель Sznajd (Sznajd-Weron, Sznajd, 2000), изначально предназначенная для описания распространения мнений, однако повторяющая динамику модели избирателя; модель Deffuant (Deffuant et al., 2000) и модель Hegselmann-Krause (Hegselmann, Krause, 2002), имеющие дело с непрерывными мнениями (т. е. мнениями, имеющими диапазон возможных вариантов, описываемых в модели вещественными числами); модель Axelrod (Axelrod, 1997), базирующаяся на двух базовых механизмах культурной ассимиляции: стремлении агентов становится более похожими друг на друга и тенденции схожих агентов чаще взаимодействовать друг с другом.

В методологии АОМ система моделируется в виде набора автономных объектов, каждый из которых принимает решение самостоятельно. Данные объекты называются агентами. Каждый агент независимо от других оценивает ситуацию и затем изменяет своё состояние («принимает решение») на основе

некоторого набора правил. Агентные модели можно рассчитывать необходимое число раз, что позволяет значительно шире оценивать статистические характеристики моделируемых процессов, по сравнению с классическими математическими моделями. Даже в моделях, где поведение агентов задаётся с помощью простейших правил, возможно появление сложной динамики поведения системы. Естественно, более сложные модели, в которых агенты могут эволюционировать (приобретая новые свойства или избавляясь от старых свойств, а также изменяя своё поведение), зачастую порождают ещё более сложные паттерны поведения.

4.1. Методы моделирования биологических популяций с учётом социальных факторов

В данном разделе рассматривается простая модель, введённая Bonabeau (Bonabeau, 2002), в которой феномены возникают за счёт двух стратегий поведения у агентов; на её основе реализована социально-биологическая модель, в которой агенты обладают большим числом характеристик, что позволяет варьировать социальные и биологические подмодели в рамках одного программного средства.

4.1.1. Модель «Агрессор — Защитник»

Модель «Агрессор — Защитник» предложена в работе (Bonabeau, 2002). Она представляет собой агентно-ориентированную модель динамики сообщества, члены которого (агенты, игроки) движутся согласно определённым правилам. В частности, выделяются стратегии движения «трусов» и «защитников». В первом случае агент А (трус) стремится «спрятаться» от агента В (агрессора) за агента С (защитника). Во втором случае агент А сам выступает в роли «защитника» или «миротворца», стремясь занять положение между агентами В и С. Данная модель является классическим примером модели, показывающей эмергентную природу социальных процессов: (1) простые правила индивидуального поведения могут привести к неожиданным результатам на уровне системы/сообщества;

(2) небольшие изменения правил могут повлечь за собой значительные изменения поведения системы.

Нами была воспроизведена и существенно расширена модель «Агрессор-Защитник» Vonabeau (Vonabeau, 2002). Были предложены две альтернативные стратегии движения – как для «трусов», так и для «защитников». Был введён параметр «враждебности» – минимального расстояния, на которое сближаются игроки (при нулевой враждебности они могут совпадать по координатам, при враждебности, равной $2r$, где r – радиус круга, символизирующего игрока, агенты будут обходить друг друга, не пересекаясь). Было введено понятие скорости агента. Кроме того, реализована возможность задания любыми параметрами модели случайной величиной из заданного распределения (в настоящее время используется бета-распределение).

4.1.2. Расширенная модель «Агрессор - Защитник»

Модель Vonabeau была расширена за счёт усложнения поведения отдельных агентов, а также введения параметров, относящихся к популяции в целом. Во-первых, помимо двух режимов (трус/защитник) для каждого агента в доработанной модели возможен выбор стратегии движения: по нормали или по медиане/антимедиане.

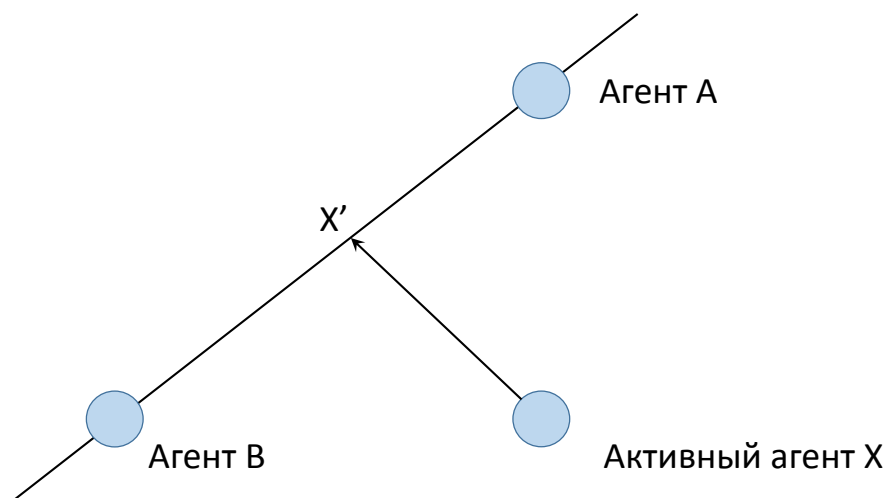


Рисунок 4.1.1. Возможная схема расположения тройки агентов.

Пусть A и B — выбранная пара агентов для игрока X . Под **движением по нормали** понимается следующий алгоритм: осуществляется проверка, соответствует ли движение по нормали к прямой AB выбранному режиму (попадает ли желаемая координата между A и B в режиме защитника, «спрячется» ли игрок за игрока B в режиме труса). Если не соответствует, игрок пытается спрятаться за агента B на минимальном допустимом расстоянии от него (в режиме труса) или занять позицию ровно посередине между A и B (в режиме защитника). Если игрок X на рис. 4.1.1. является трусом с приоритетным движением по нормали, то точка X' на прямой AB не удовлетворяет желаним труса спрятаться от агрессора, и игрок направится «за спину» агента B . Если же предположить, что X – защитник, то точка X' удовлетворяет его требованиям.

Под **движением по медиане** понимается следующее: игрок стремится занять позицию ровно посередине между A и B (режим защитника) или позицию, симметричную агрессору A относительно защитника B (режим труса). Последний тип движения назовём движением по **антимедиане** (см. рис. 4.1.2).

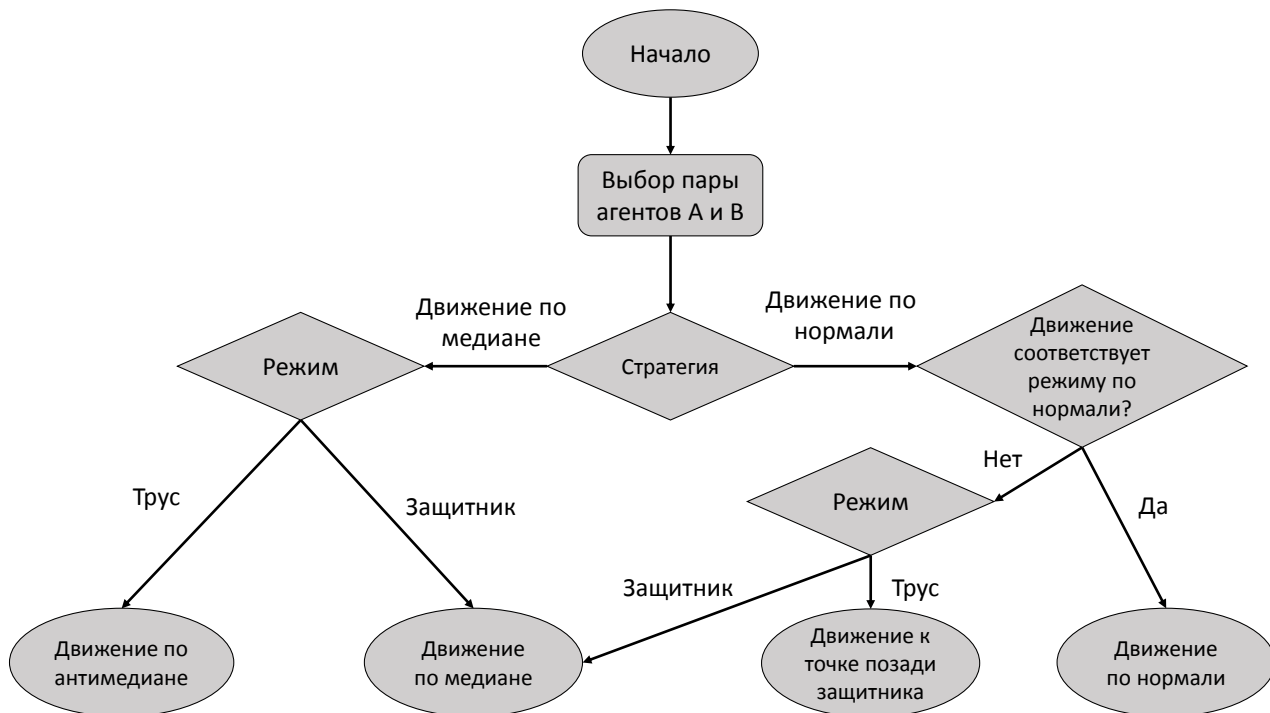


Рисунок 4.1.2. Блок-схема принятия решения о направлении движения, осуществляемая агентом на каждой итерации.

Пусть $X(t)$ – радиус-вектор активного агента в момент времени t , $A(t)$, $B(t)$ – радиус-векторы агрессора и защитника/защищаемого соответственно. Тогда в зависимости от пары режим-стратегия движение агента описывается формулами (4.1.1) — (4.1.9).

1. PM (protector-along-median): агент стремится занять место ровно посередине между агрессором и защищаемым агентом, двигаясь тем самым по медиане в треугольнике ABX :

$$\overrightarrow{PM}(t) = \vec{X}(t) + \min(\text{speed}, |\vec{A}(t)/2 + \vec{B}(t)/2|) * \text{normalize}(\vec{A}(t)/2 + \vec{B}(t)/2); \quad (4.1.1)$$

$$\vec{X}(t+1) = \overrightarrow{PM}(t); \quad (4.1.2)$$

2. PN (protector-along-normal): агент проверяет, соответствует ли движение к прямой AB по кратчайшему пути (т. е. по нормали) режиму защитника; если нет — движется по медиане.

$$\overrightarrow{PN}(t) = \vec{X}(t) + \min(\text{speed}, \text{normal}(\vec{B}(t) - \vec{A}(t))) * \text{normalize}(\text{normal}(\vec{B}(t) - \vec{A}(t))) \quad (4.1.3)$$

$$\vec{X}(t+1) = \begin{cases} \overrightarrow{PN}(t) \\ \text{если } |\vec{B}(t) - \vec{A}(t)| = \\ \max(|\vec{B}(t) - \vec{A}(t)|, \\ |\overrightarrow{PN}(t) - \vec{A}(t)|, \\ |\overrightarrow{PN}(t+1) - \vec{A}(t)|); \\ \overrightarrow{PM}(t) \\ \text{иначе} \end{cases} \quad (4.1.4)$$

3. CM (coward-along-median): агент стремится занять позицию, симметричную агрессору относительно позиции защитника (движется по антимедиане).

$$\overrightarrow{CM}(t) = \vec{X}(t) + \min(\text{speed}, |\vec{2B}(t) - \vec{A}(t)|) * \text{normalize}(\vec{2B}(t) - \vec{A}(t)); \quad (4.1.5)$$

$$\vec{X}(t+1) = \overrightarrow{CM}(t); \quad (4.1.6)$$

4. CN (along-normal coward): агент проверяет, соответствует ли движение к прямой АВ по кратчайшему пути (т. е. по нормали) режиму труса; если нет — движется по антимедиане.

$$\overrightarrow{CN}(t) = \vec{X}(t) + \min\{\text{speed}, |\text{normal}(\vec{B}(t) - \vec{A}(t))|\} * \text{normalize}(\text{normal}(\vec{B}(t) - \vec{A}(t))); \quad (4.1.7)$$

$$\overrightarrow{hide}(t) = \vec{B}(t) + \text{enmity} * \text{normalize}(\vec{B}(t) - \vec{A}(t)); \quad (4.1.8)$$

$$\vec{X}(t+1) = \begin{cases} ANP(t), \\ \text{если } |\overrightarrow{CN}(t) - \vec{A}(t)| = \\ \max\{|\overrightarrow{CN}(t) - \vec{A}(t)|, |\overrightarrow{CN}(t) - \vec{B}(t)|, \\ |\vec{B}(t) - \vec{A}(t)|\}; \\ \vec{X}(t) + \min\{\text{speed}, |\overrightarrow{hide}(t)|\} * \text{normalize}(\overrightarrow{hide}(t)) \\ \text{иначе;} \end{cases} \quad (4.1.9)$$

где *speed* – скорость агента, *normalize* - оператор нормализации вектора, *normal* — оператор взятия нормали (с учётом направления вектора), *enmity* – «враждебность».

Следующие новые параметры — скорость и враждебность — уже были использованы в формулах (4.1.1) — (4.1.9). Под **враждебностью** понимается минимальное допустимое расстояние между двумя агентами (при нулевой враждебности они могут совпадать по координатам, при враждебности, равной $2r$, где r – радиус окрестности, занятой игроком, агенты будут обходить друг друга, не пересекаясь); под **скоростью** понимается расстояние, на которое может сдвинуться агент за единицу времени.

Также для модели был введён выходной параметр – **неудовлетворённость** как мера относительной отдалённости от желаемых координат. Неудовлетворённость агента под номером i рассчитывается по формуле (4.1.10):

$$discontentment_i = distance_i / maxDistance ; \quad (4.1.10)$$

где $distance$ – расстояние до желаемых координат, $maxDistance$ – максимальное расстояние среди агентов заданной окрестности.

Неудовлетворённость агента имеет два способа подсчёта: максимальное значение $maxDistance$ считается по всем агентам игры (такой тип расчёта неудовлетворённости будем называть **абсолютным**) или максимальное значение $maxDistance$ считается внутри окрестности заданного радиуса (такой тип расчёта неудовлетворённости будем называть **относительным**). В таком случае радиус окрестности является дополнительным параметром модели.

Средняя неудовлетворённость рассчитывается как среднее по неудовлетворённостям всех агентов:

$$discontentment_{av} = (\sum_1^n discontentment_i) / n. \quad (4.1.11)$$

Наконец, модель допускает настройку периодичности сцены. В случае периодических граничных условий полагаем, что квадрат сцены окружён четырьмя аналогичными сценами с идентичным расположением, состоянием и поведением агентов на них (с той же конфигурацией популяции); если агент покидает основную сцену, допустим, через левый её край, аналогичный агент приходит на сцену через правый, тем самым конфигурация популяции сохраняется и имитируется бесконечность системы. В противном случае границы предполагаются непроницаемыми для агентов.

4.1.3. Основные режимы динамики модели «агрессор — защитник»

Добавление большого числа параметров модели сделало её гораздо более сложной и гибкой. Рассмотрим поведение системы при варьировании основных параметров: стратегии движения, враждебности, соотношения агентов-трусов и защитников. Заметим, что при движении по медиане и нулевой враждебности расширенная и исходная модели совпадают, движение агентов модели Vonabeau и модели «Агрессор — Защитник» идентично.

При движении по медиане/антимедиане популяция в целом характеризуется большей неудовлетворённостью на начальной стадии по сравнению с движением по нормали, однако проявляет большую приспособляемость и выходит на стационар с меньшим значением неудовлетворённости. При движении по нормали значение выходного параметра мало изменяется как внутри групп, так и в целом для популяции. На рис. 4.1.3 изображено сравнение поведения системы при движении по нормали и по медиане для популяции с равным соотношением трусов и защитников.

Введение параметра враждебности не внесло особой изменчивости для популяций, где все агенты движутся по нормали (ср. рис. 4.1.4 – 4.1.5).

Однако для популяций, где агенты движутся по медиане, возникает особенность. При малом числе трусов в популяции (не более 5%) неудовлетворённость при нулевой враждебности достигает порядка 0.95 (рис. 4.1.6), в то время как при ненулевой враждебности устойчивый достигается устойчивый стационар со значением около 0.56 (рис. 4.1.7).

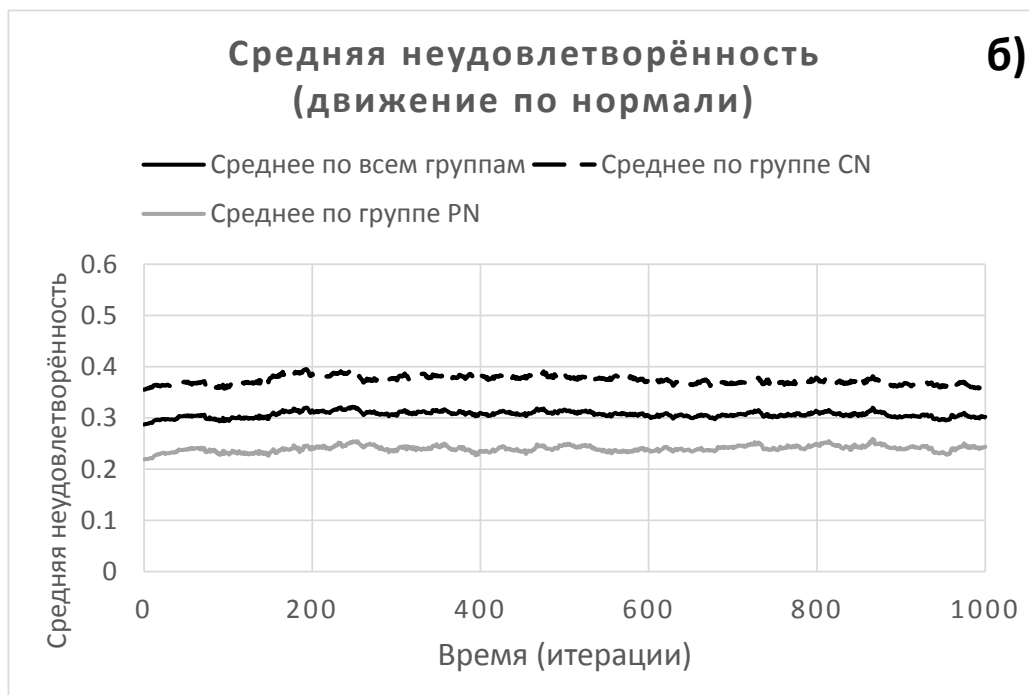
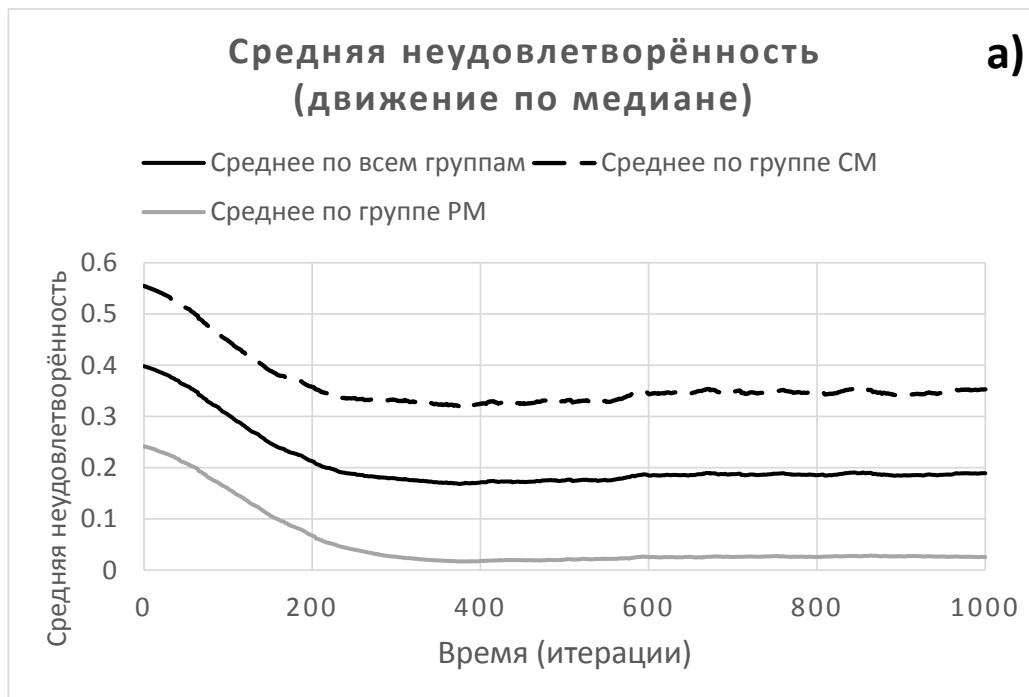


Рисунок 4.1.3. Динамика модели для движения по медиане (а) и по нормали (б). Популяция размера 100, соотношение трусов и защитников 1:1, radius = 3, enmity = 0, speed = 1. Усреднение на 100 моделей. По оси X – время (число итераций), по оси Y - неудовлетворённость (средняя если для сообщества).

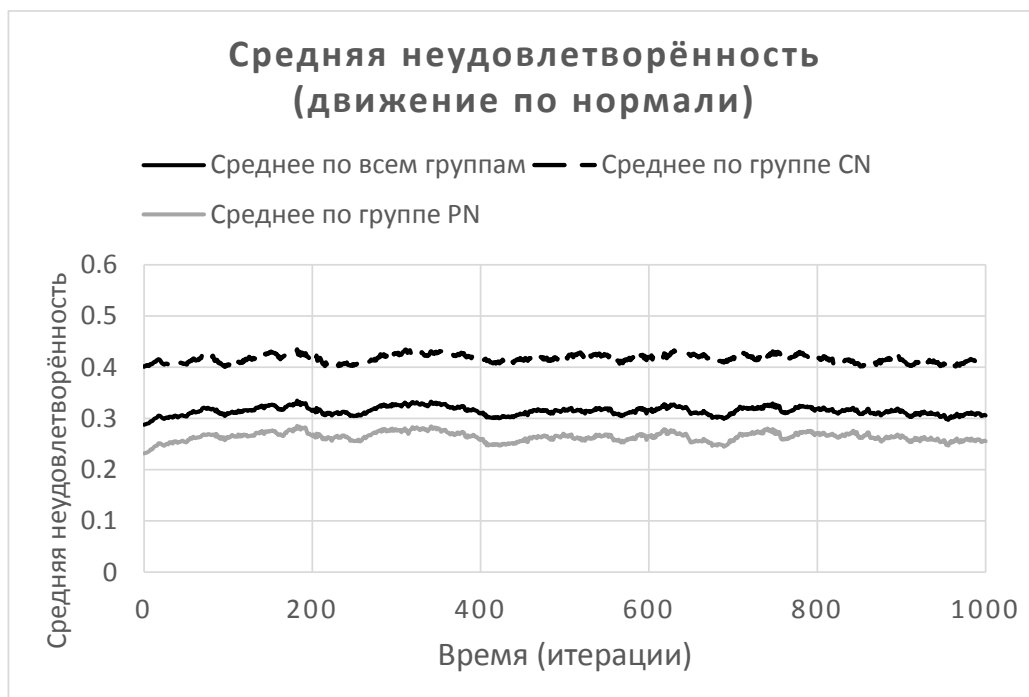


Рисунок 4.1.4. Динамика модели для движения по нормали для $enmity = 0$. Популяция размера 100, соотношении трусов и защитников 1:2, $radius = 3$, $speed = 1$. Усреднение на 100 моделей. По оси X – время (число итераций), по оси Y - неудовлетворённость (средняя если для сообщества).

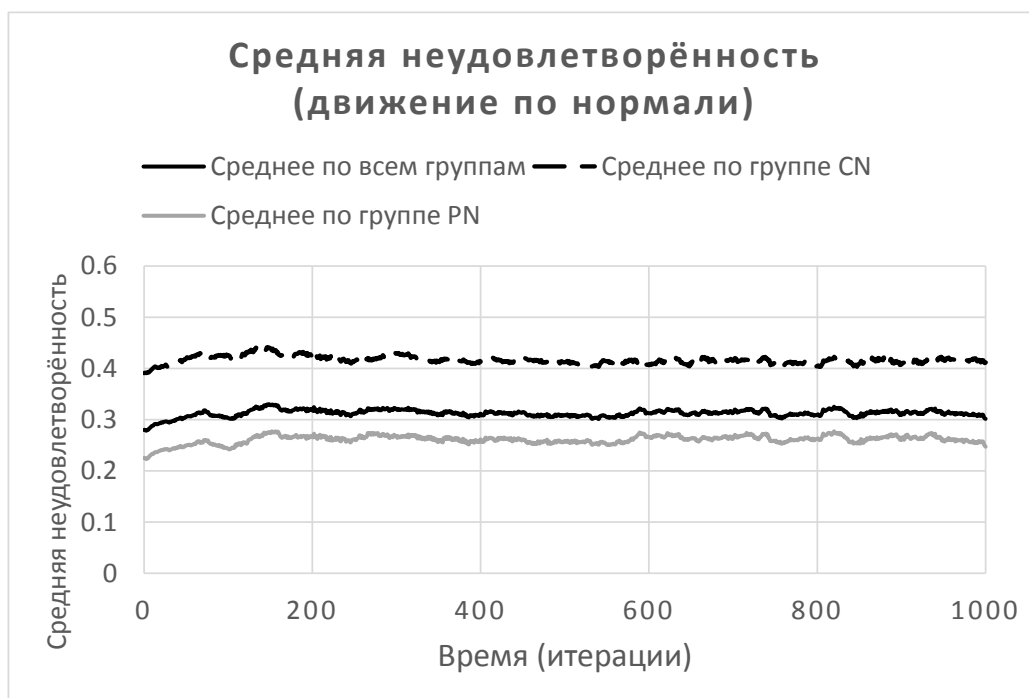


Рисунок 4.1.5. Динамика модели для движения по нормали для $enmity = 3$. Популяция размера 100, соотношение трусов и защитников 1:2, $radius = 3$, $speed = 1$. Усреднение на 100 моделей. По оси X – время (число итераций), по оси Y - неудовлетворённость (средняя если для сообщества).

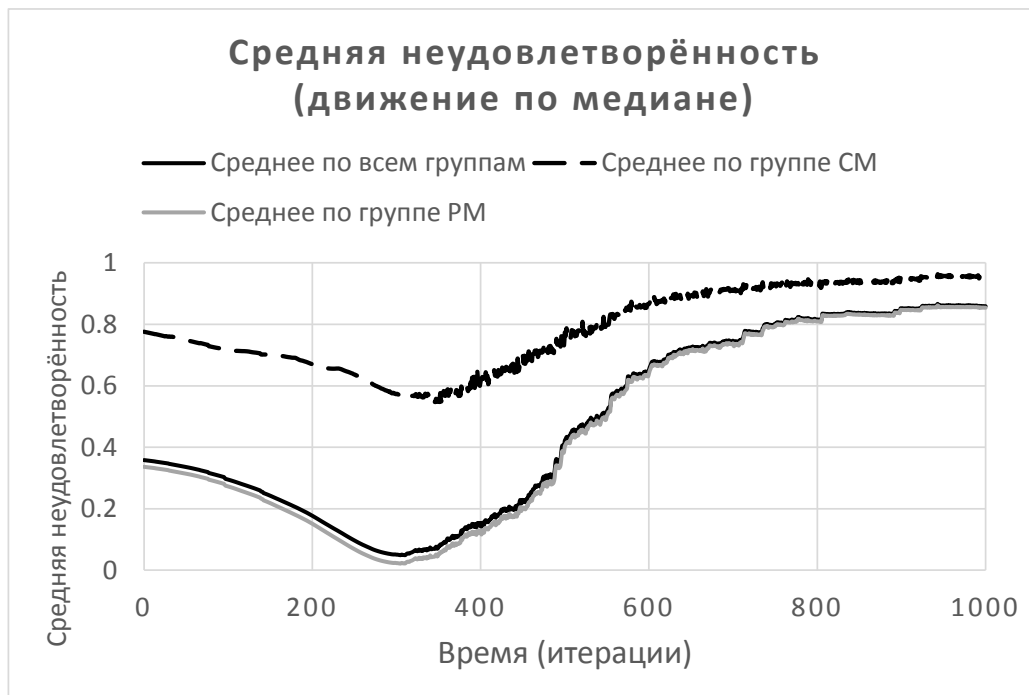


Рисунок 4.1.6. Динамика модели для движения по медиане для $enmity = 0$. Размер популяции 100, соотношение трусов и защитников 5:95, $radius = 3$, $speed = 1$. Усреднение на 100 моделей. По оси X – время (число итераций), по оси Y - неудовлетворённость (средняя если для сообщества).

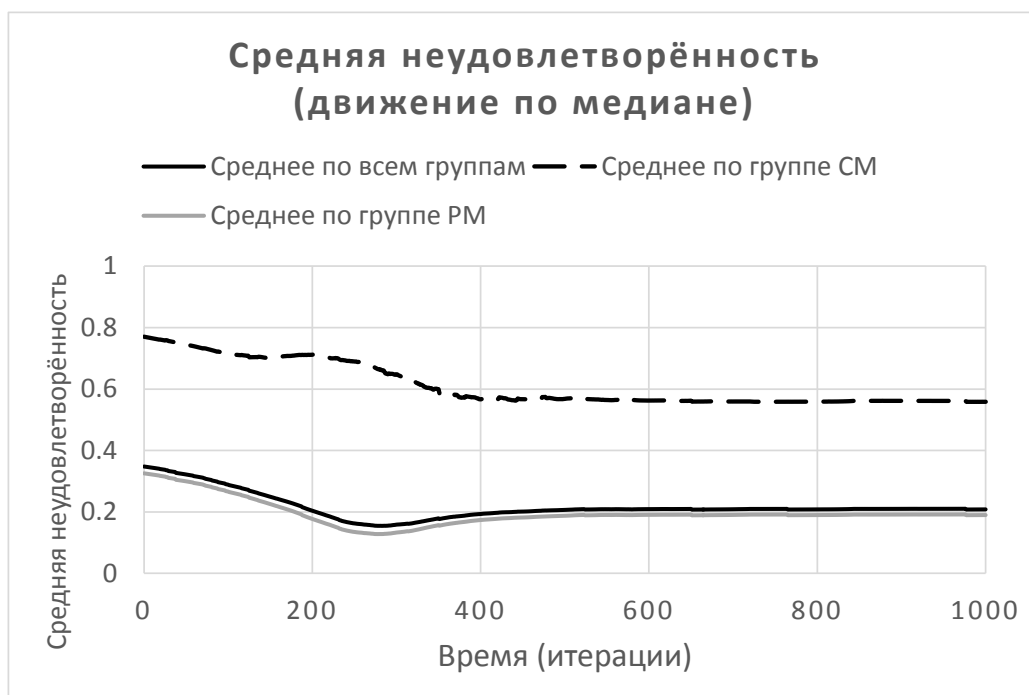


Рисунок 4.1.7. Динамика модели для движения по медиане для $enmity = 3$. Размер популяции 100, соотношение трусов и защитников 5:95, $radius = 3$, $speed = 1$. Усреднение на 100 моделей. По оси X – время (число итераций), по оси Y - неудовлетворённость (средняя если для сообщества).

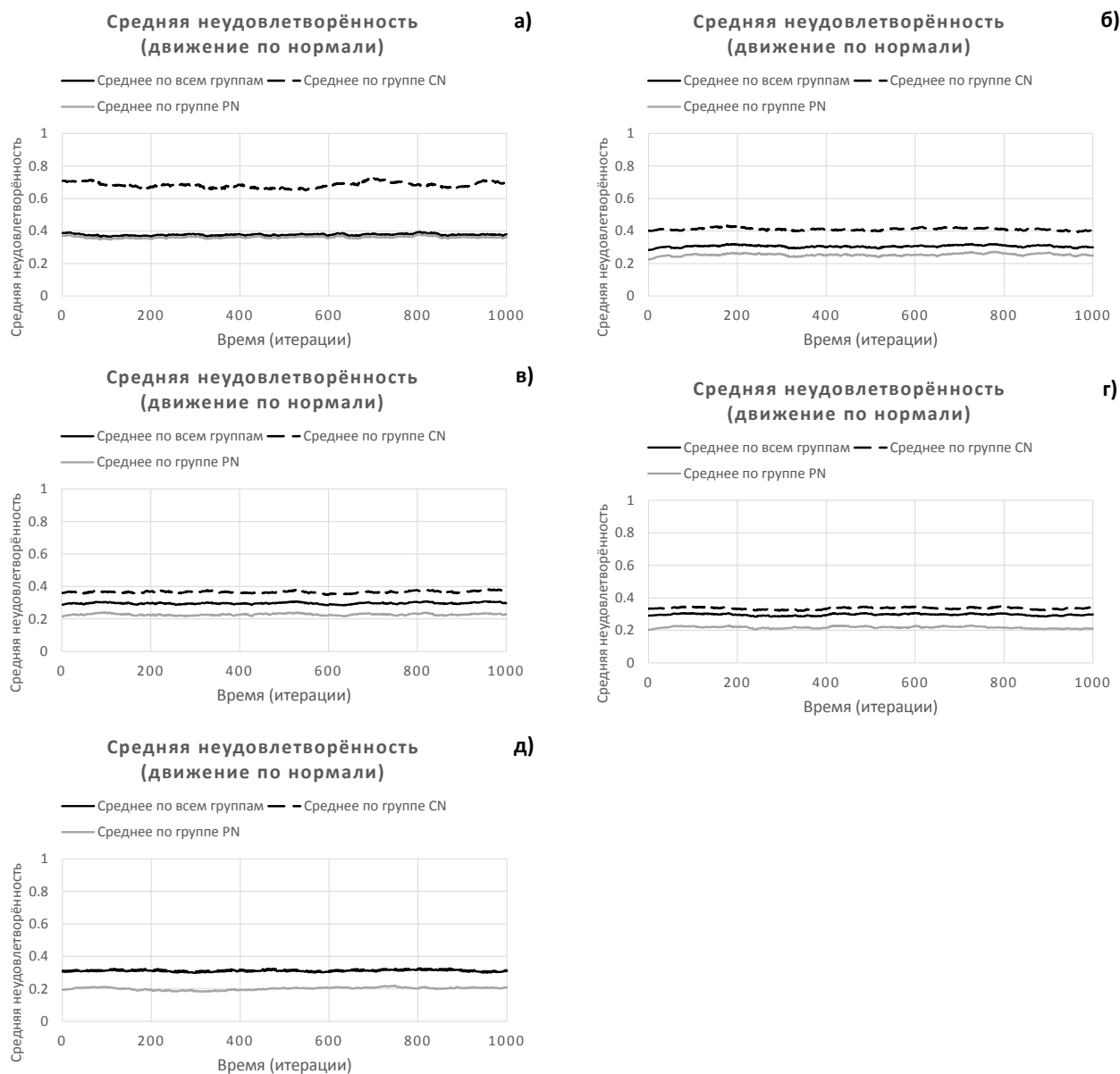


Рисунок 4.1.8. Динамика модели для движения по нормали при различном соотношении тросов и защитников: а) 5:95, б) 1:2, в) 1:1, г) 2:1, д) 95:5. Размер популяции 100, radius = 3, enmity = 6, speed = 1. Усреднение на 100 моделей. По оси X – время (число итераций), по оси Y - неудовлетворённость (средняя если для сообщества).

Такие численные результаты объясняются следующим образом: при нулевой враждебности агенты популяции стягиваются в единую точку, тем самым каждый агент достигает своих желаемых координат. Но, так как расстояние до желаемых координат всех агентов оказываются равным нулю и

среднему значению внутри окрестности (которое также равно нулю), и так как неудовлетворённость стремится иметь значения выше среднего, то неудовлетворённость каждого агента оказывается максимальной, а значит максимальна и средняя неудовлетворённость по всей популяции. В случае ненулевой враждебности защитники стягиваются в плотный узел, система приходит в неподвижность, немногочисленные трусы занимают желаемую позицию вне узла; однако достичь желаемых координат для всех агентов невозможно в силу враждебности.

Наконец, заметим, что неудовлетворённость для группы агентов-трусов всегда оказывается значительно выше, чем для агентов-защитников, независимо от выбора прочих параметров. В то же время увеличение доли трусов в популяции ведёт к уменьшению средней неудовлетворённости (рис. 4.1.8).

Таким образом, по расширению модели Vonabeau, рассмотренному в данном разделе, можно сделать ряд следующих наблюдений, новых относительно исходной модели:

- динамика модели зависит от выбора движения по медиане /по нормали следующим образом: в первом случае популяция в целом характеризуется большей неудовлетворённостью на начальной стадии по сравнению с движением по нормали, однако проявляет большую приспособляемость и выходит на стационар с меньшим значением неудовлетворённости; во втором случае значение выходного параметра мало изменяется как внутри групп, так и в целом для популяции;
- введение параметра враждебности породило особенность для популяции с движением по медиане и малым числом трусов в ней (~5%): величина неудовлетворённости выходит на стационар со значением в 1;

- неудовлетворённость трусов всегда выше неудовлетворённости защитников, однако увеличение доли трусов в популяции ведёт к уменьшению данной величины.

Модель может расширяться и далее, посредством добавления в неё механизмов передачи идей и мнений, что позволит моделировать распространения идей внутри популяции (в то же время моделировать распространение болезней, слухов, любой информации); реализацией неоднородного распределения параметров внутри популяции (например, скорости, враждебности — в настоящее время это общий параметр для всех агентов популяции), рассмотрением различных топологий социальной сети. Но уже сейчас модель «Агрессор-Защитник» способна воспроизвести возникновение большого числа социальных и биологических явлений.

4.2. Модель распространения генетической глухоты в популяциях людей с разными социально-демографическими структурами

Тяжелая потеря слуха является актуальной медико-социальной проблемой, поскольку в той или иной степени затрагивает более чем 10% населения мира, приводит к инвалидности и существенно снижает качество жизни глухих людей. Потеря слуха может быть обусловлена как средовыми, так и генетическими причинами. В среднем, 1 из 1000 новорожденных рождается глухим, и у 50-60% из них глухота имеет генетическую природу. Наследственная потеря слуха характеризуется уникально высокой гетерогенностью генетического контроля (много генов – один фенотип). Распространенность различных генетических форм глухоты широко варьирует в разных регионах мира и может определяться, как и для многих других генетических заболеваний, известными факторами, формирующими генетическую структуру популяций (этнический состав, изоляция, миграции, эффекты основателя и «бутылочного горлышка», доля близкородственных браков, селективное преимущество гетерозигот). Наибольшую частоту (до 40-50% в некоторых европейских странах и США) имеет рецессивно наследуемая

глухота, обусловленная мутациями гена GJB2 (Cx26). Предполагается, что, в отличие от других генетических заболеваний, высокая в настоящее время частота этой формы наследственной глухоты может определяться долговременной традицией заключения ассортативных браков между глухими людьми (выбор по патологическому фенотипу) в сочетании с ростом их биологической приспособленности (genetic fitness) и социальной адаптации, и стартом для этих событий, вероятно, явился «прорыв» глубокой социальной изоляции глухих людей, произошедший 300-400 лет назад, когда в Европе стали создаваться школы для глухих с обучением жестовому языку как единому средству коммуникации (Nance et al., 2000). Результаты компьютерного моделирования показали, что такие социальные процессы могли бы привести к удвоению частоты глухоты, обусловленной мутациями гена GJB2, за 200-летний период (Nance, Kearsey, 2004). Существенное изменение брачной структуры глухих людей выявлено при ретроспективном (XIX-XX вв.) анализе уникальных архивных данных. Данные о достоверном росте частоты мутаций гена GJB2 за период в три поколения получены в экспериментальном исследовании (Arnos et al., 2008).

Ретроспективные исследования социально-демографической структуры сообществ глухих людей крайне ограничены практическим отсутствием соответствующих архивных данных, но изучение социальных, демографических и медико-генетических характеристик современных сообществ глухих людей имеет важное значение как для прогнозирования распространенности наследственных форм потери слуха, так и понимания роли социальных факторов в эволюционных процессах, происходящих в популяциях человека. К сожалению, раньше к этому явлению не привлекалось достаточно внимания (особенно в Российской Федерации) и информация по этому заболеванию почти никем не собиралась, и сейчас накоплена статистика за не очень продолжительный промежуток времени. Тем не менее, изучение социально-демографических и генетических характеристик сообществ глухих

людей имеет важное значение в современном мире, как для прогнозирования дальнейшего распространения глухоты, так и для изучения зависимости между социальными факторами и распространенности генетической глухоты.

При исследовании различных заболеваний в популяциях человека возникает проблема с проведением эксперимента для проверки выдвинутых заранее гипотез. Особенно это касается болезней с выраженными генетической и социальной компонентами в механизмах распространения этих заболеваний. В таких случаях математическое и компьютерное моделирование являются чуть ли не единственным методом научного исследования данных задач, не вызывающим большого количество этических противоречий.

В данном разделе мы описываем компьютерную модель распространения генетической глухоты в популяциях человека, учитывающую как генетические, так и социальные факторы. Модель предназначена для выявления того, как на распространение генетической глухоты влияет наличие жестокого языка и социальные предпочтения при выборе партнера для образования брака. В работе использовались данные по распространенности генетической глухоты в республике Якутия и некоторых регионах Сибири, любезно предоставленные к.б.н. Барашковым Н.А., Романовым Г.П., к.б.н. Бады-Хоо М.С., а также курировавшей данную работу с биологической стороны к.б.н. Посух О.Л.

Модель была построена в терминах описанной в Главе 3 модели, лежащей в основе программной платформы «Диплоидный Эволюционный Конструктор» (ДЭК). Для детального исследования данной модели нами был разработан специализированный программный комплекс PGD-S (Population Genetics of Deafness – Simulator). PGD-S является расширением ПК ДЭК и позволяет проводить численные исследования модели в двух режимах: (1) Режим исследования отдельной модели; (2) Пакетный режим. В первом режиме проводится детальное исследование одного запуска модели (раунда

моделирования) с возможностью визуализации всех характеристик модели, а также с возможностью изменять параметры непосредственно в ходе расчётов, между отдельными шагами моделирования. Под пакетным заданием понимается запуск заданного числа моделей при фиксированных входных данных и возможностью дальнейшей обработки как данных по отдельным моделям, так и статистики по всем моделям.

Базовые объекты модели – агент (индивидуум, человек) и среда обитания (рис. 4.2.1).

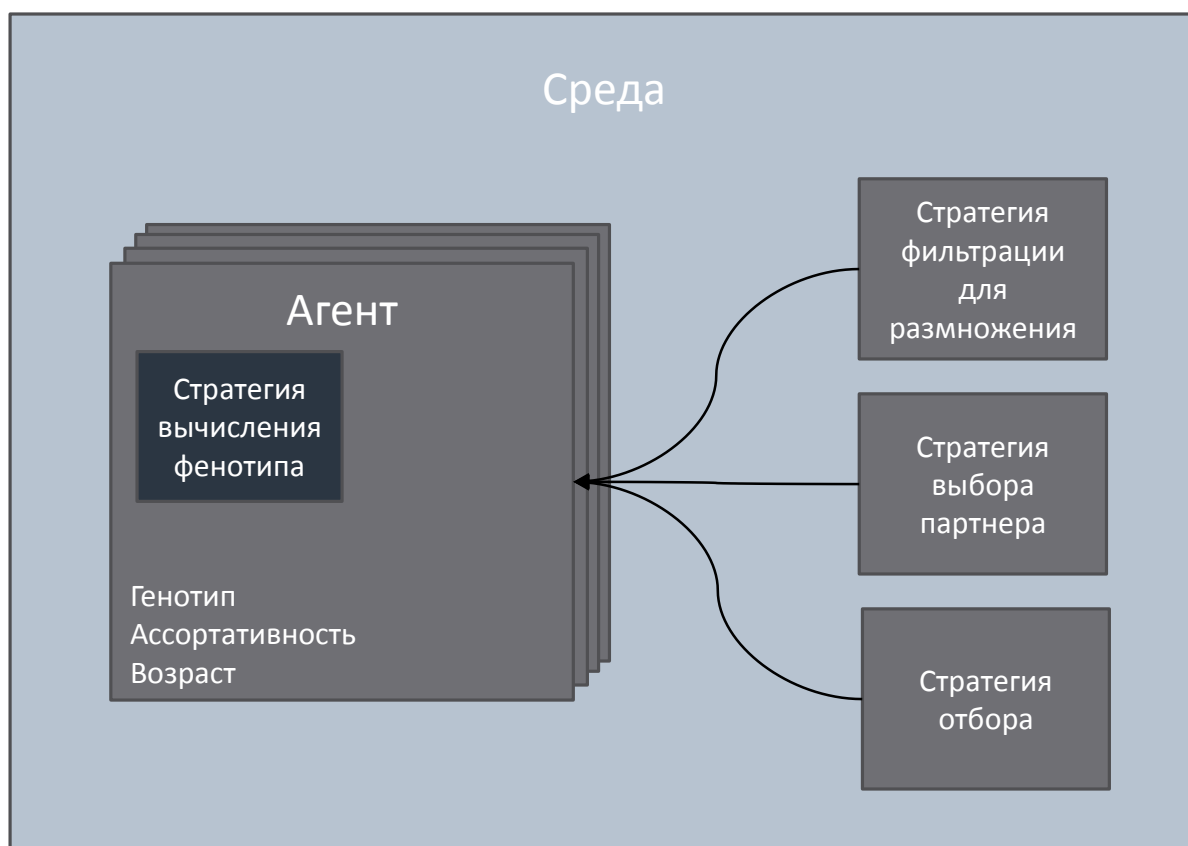


Рисунок 4.2.1. Принципиальная схема агент-ориентированной модели возможных трендов распространенности наследуемых форм болезней (в данном случае – потери слуха).

Модель в PGD-S описывает популяцию «агентов» (особей), каждый из которых обладает следующими индивидуальными характеристиками: пол, возраст, генотип, фенотип (слышащий, глухой), социальная позиция (мера предпочтительности вступления в ассортативный брак). Генотип

представляется как пара векторов $\{m, f\}$ (диплоидный геном), где $m = (m_1, m_2, \dots, m_k)$ – набор генов (аллельных вариантов), унаследованный от матери, $f = (f_1, f_2, \dots, f_k)$ – набор генов, унаследованных от отца. В настоящее время исследуется модель со следующими ограничениями: $k = 1$, f_i, m_i принимают значения 0 («здоровый» аллель) или 1 (мутантный аллель). Доминантность/рецессивность мутантного аллеля задаётся пользователем.

Итерационный шаг модели составляет один год. На каждом шаге моделируются следующие события:

1. Формирование пар (вступление в брак) – стохастическое событие для выбранных случайным образом агентов, ещё не вступивших в брак, достигших минимально допустимого возраста, и не достигших максимально допустимого возраста. Вероятность вступления в брак, минимальный и максимальный допустимые возраста задаются в виде параметров модели. Дополнительным параметром, определяющим возможность создания пары, является «социальная позиция» агента, представляющая собой число в диапазоне от 0 до 1. Этот параметр описывает вероятность образования ассортативного брака по фенотипическому признаку наличия или отсутствия глухоты: 0 – при формировании пары этот признак не учитывается, 1 – агент согласен исключительно на ассортативный брак. *Отметим, что вступление в брак в нашей модели соответствует в т.ч. и т.н. гражданскому браку.*
2. Генерация новых агентов – для каждой пары стохастическим образом принимается решение о «рождении» у неё нового агента. Вероятность этого события обуславливается как общедемографическими (коэффициент рождаемости в популяции), так и индивидуальными (количество «детей», уже имеющих у пары) параметрами. Новый агент получает гены от «отца» и от «матери».
3. Элиминация агентов – часть агентов удаляется из популяции стохастическим образом. Вероятность элиминации обуславливается

коэффициентом смертности в популяции. Отметим, что в модели можно задать индивидуальные коэффициенты смертности для разных возрастных групп, а внутри этих групп – отдельно для слышащих и глухих групп.

4. Увеличение возраста для всех агентов в популяции.

Основной единицей модели является агент, который представляет собой абстракцию отдельного индивидуума, взаимодействующего с другими агентами-индивидуумами. Агент имеет следующие атрибуты/параметры:

- Возраст;
- Пол;
- Генотип;
- Фенотип;
- Предпочтительный фенотип брачного партнёра и степень готовности вступить в брак с индивидуумом с фенотипом, отличным от предпочтительного – так называемая «разборчивость»;
- Владение жестовым/устным языком.

Генотип представлен списком пар аллелей, каждый из которых может принимать значения true/false – есть или нет повреждающая мутации в данном аллельном варианте. Модель позволяет исследовать как многолокусные варианты, так и однолокусный.

Фенотип показывает наличие глухоты у агента. Он может принимать значения true/false – глухота присутствует/отсутствует. Фенотип агента вычисляется по его генотипу с помощью **стратегии вычисления фенотипа (СВФ)**. Это позволяет менять механизм фенотипического проявления (глухота) (рецессивный тип наследования патологии – биаллельные

рецессивные мутации гена / доминантный тип наследования патологии – одна доминантная мутация) генотипа без внесения изменений в общую архитектуру модели. СВФ принимает на вход генотип, и выдает результат «да/нет», являющийся ответом на вопрос «глухой ли агент?». Возможна модификация СВФ для расчёта степени потери слуха в виде величины из заданного диапазона $[d_{low}, d_{high}]$.

Разборчивость – требовательность агента к совместимости с партнером при выборе пары, определяется **стратегией выбора партнера (СВП)**. СВП использует как общие для всех индивидуумов данные потенциальных партнеров (возраст, пол), так и специфичные для нашего проекта: фенотипы (по состоянию слуха) партнера-инициатора и фенотип потенциального партнера. Могут вводиться и другие переменные, например, этническая принадлежность потенциального партнера. Кроме того, для оценки совместимости с потенциальным партнером, наряду с фенотипом используются дополнительные параметры, такие как жестовым и устным языком. Отметим, что список параметров СВП может быть расширен любыми другими признаками, имеющими булево, дискретное или непрерывное значение. Для каждого из параметров у агента имеется собственный (индивидуальный) вес, который отображает важность соответствующего параметра при выборе партнера. Так, вес 1000 для параметра А и вес 10 для параметра В означают, что для агента значение параметра А у потенциального партнера в 100 раз важнее значения параметра В. В данной модели значения весов являются относительными, т.к. показывают только соотношение эффектов параметров на поведения агента.

Следующей структурной единицей модели служит семья. Семья состоит из двух партнеров и некоторого неотрицательного количества детей. В текущей версии модели агенты образуют семью согласно результату работы СВП и далее состоят в союзе вплоть до смерти одного из них. Основное

назначение семьи – формирование устойчивой пары агентов и, как следствие, уменьшение случайности в формировании генотипов следующего поколения.

4.2.1. Итерационный цикл модели PGD-S

Данная модель описывает жизненный цикл популяции и ее изменение со временем. Моделирование производится итеративно, где каждый шаг итерации соответствует одному году реального времени. Каждый шаг состоит из нескольких этапов, реализующих отдельные аспекты жизни популяции: выбор партнера, образование пар, рождение детей и отбор особей. Каждый этап представляет собой отдельную подмодель, которая в PGD-S описывается аналогично тому, как это сделано в ПК ДЭК – через паттерн *стратегия*.

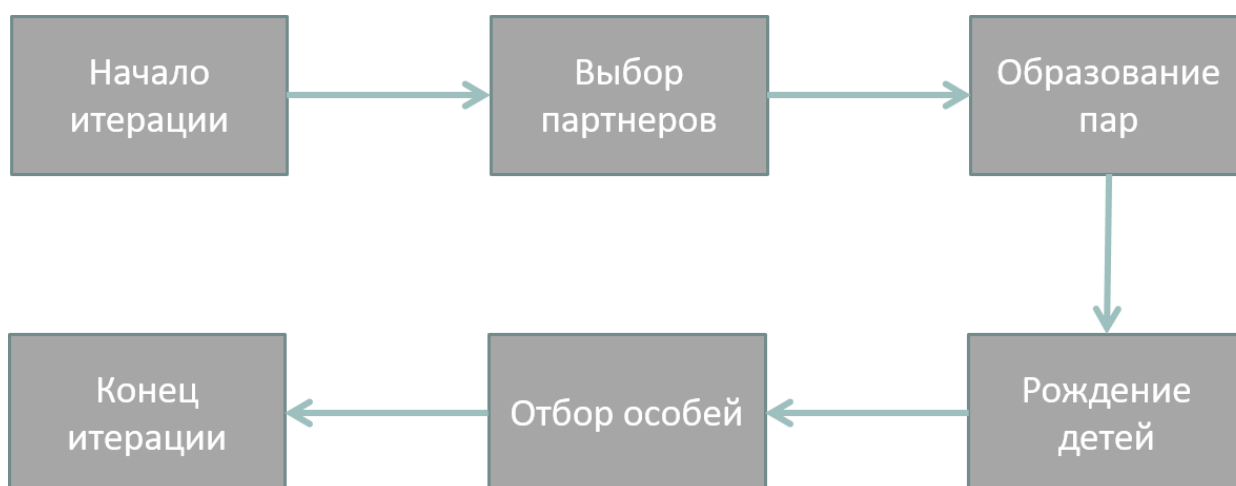


Рисунок 4.2.2. Порядок этапов в одном итерационном шаге.

Ниже приведены описания основных стратегий PGD-S.

4.2.2. Стратегия выбора партнеров (СВП)

Основным для данного алгоритма этапом является этап выбора партнеров, так как именно на нем проявляется влияние социальных факторов.

На этапе выбора партнеров в первую очередь из популяции выбирается фертильная субпопуляция. Агенты, не достигшие возраста, позволяющего иметь детей, или уже вышедшие из него (нижняя и верхняя границы этого

возраста задаются дополнительными параметрами модели для всей популяции – ключи *Reproduction parameters/Fertile age start* и *Reproduction parameters/Fertile age end* соответственно), не представляют интереса в рамках данной задачи, так как они уже не могут внести изменения в генетический состав популяции.

Из мужской части фертильной субпопуляции берется представитель, для которого запускается процесс выбора партнерши. В первую очередь проверяется его вероятность вступления в брак. Она представляет собой случайную величину, подчиняющуюся равномерному распределению, в интервале от 0 до 1 включительно. Если случайная величина оказывается меньше порогового значения склонности к браку, т.н. брачности (ключ настроек *Marriage parameters/Bachelority*), то считается, что на данной итерации этот конкретный мужчина отказывается выбирать себе партнершу, и для него процесс выбора партнера заканчивается.

Если величина превысила порог, то он начинает искать подходящую себе пару. Для каждой потенциальной жены в первую очередь проверяется разница в возрасте с мужчиной. С увеличением разницы возрастов вероятность заключения брака падает экспоненциально/

Если разница возрастов достаточно мала, то вычисляется, совместимы ли партнеры. Степень их совместимости вычисляется следующим образом: параметры мужчины и потенциальной партнерши сравниваются поэлементно, в результате чего получается общий для обоих агентов список совпадений.

$s_i = p_{woman_i} \sim p_{man_i}, \forall i$, где s_i – совпадение i -х параметров, p_{woman_i} - i -й параметр женщины, p_{man_i} - i -й параметр мужчины,

Далее у каждого агента берется список весов параметров и на основе этого списка считается метрика оценки партнера. Метрика считается по следующей формуле:

$$m = \sum_i w_i * s'_i, \text{ где } w_i - i\text{-й вес, а } s'_i = \begin{cases} 1, s_i = true \\ 0, s_i = false \end{cases}$$

Полученная метрика сравнивается с разборчивостью агента, выступающей как пороговое значение. Если метрика оказывается больше разборчивости, то считается, что потенциальный партнер устраивает агента.

Если оба агента устраивают друг друга, то они формируют пару и выбывают из дальнейшего процесса выбора партнеров. Если нет – подбор партнеров продолжается. Если подходящий партнер так и не найден – считается, что на этой итерации агент остается без партнера.

Моделирование рождения детей

Каждую итерацию агенты, состоящие в браке, могут принять решение о рождении детей. Этот этап, в свою очередь, состоит из нескольких подэтапов. Так, сначала проверяется наличие в семье малолетних детей. Если младшему из детей семь лет или меньше, то семья примет решение заводить детей на текущей итерации с вероятностью $\frac{youngestAge}{7}$, где *youngestAge* – возраст младшего ребенка. Если младший ребенок старше 7 лет, то считается, что его возраст не оказывает влияние на принятие решения о рождении следующего.

Далее идет проверка количества детей в семье. С ростом размера семьи вероятность рождения следующего ребенка уменьшается экспоненциально. Математическое ожидание количества детей берется из настроек приложения (ключ *Birth parameters/Children expected value*).

Если обе предыдущих проверки дали положительный результат, то с некоторой вероятностью семья заводит ребенка. Вероятность берется из настроек (ключ *Birth parameters/Birth rate*) и отражает общую рождаемость по популяции.

Алгоритм создания агента

При рождении ребенка создается новый агент, который добавляется в список детей семьи и в общий список агентов программы.

Новый агент конфигурируется генотипами своих отца и матери. Для каждого гена в генотипе агента берутся случайно выбранные аллели данного гена у отца и матери, из которых формируются аллели гена у агента-ребенка. В процессе формирования генотипа выбранные аллели могут с взятой из настроек модели (ключ *Birth parameters/Mutation rate*) вероятностью подвергнуться мутации.

Далее с помощью стратегии СВФ на основе полученного генотипа вычисляется фенотип агента: будет ли он слышащим или нет. Если СВФ дает «слышащий» фенотип, то с некоторой вероятностью агент может потерять слух по негенетическим причинам. В этом случае его социальное поведение соответствует поведению глухого по генетическим причинам агента, но вклад в генетическое разнообразие он вносит сообразно своему генотипу.

В зависимости от полученного фенотипа инициализируются дополнительные параметры и веса предпочтений агента. Значения параметров определяются случайным образом сообразно вероятностям, заданным для каждого параметра в настройках.

Алгоритм отбора индивидов

На каждой итерации для каждого агента проверяется, выживает ли он. Агент может умереть по двум причинам: случайно или от старости.

Случайная смерть моделирует возможность смерти агента по некоторой внешней причине (стихийное бедствие, автомобильная авария и т.д.). Реализуется она путем сравнения случайной величины, задаваемой равномерным распределением $U(0;1)$ с пороговым значением, взятым из настроек (ключ *Death parameters/Accident probability*). Это значение означает

среднюю смертность по внешним причинам, взятую из статистических данных. Средняя смертность различна для глухой и слышащей субпопуляций в силу их разной адаптированности к условиям окружающей среды.

Алгоритм смерти от старости параметризуется двумя величинами: средней продолжительностью жизни (ключ настроек *Death parameters/Death age*) и отклонением продолжительности жизни от среднего (ключ настроек *Death parameters/Death delta*). Вероятность смерти от старости возрастает линейно от 0 до 1 на промежутке [*deathAge – deathDelta; deathAge + deathDelta*], где *deathAge* – средняя продолжительности жизни, а *deathDelta* – отклонение от среднего. Отсюда получается, что к возрасту *deathAge + deathDelta* вероятность становится равной 1, и агентов старше *deathAge + deathDelta* в модели нет.

4.2.3. Программный комплекс PGD-S

Общая архитектура программного комплекса «Population Genetics of Deafness – Simulator» (PGD-S) приведена в Приложении 5. Для представления состояния системы разработан графический интерфейс пользователя, содержащий ряд мониторов, отображающих состояние системы.

Мониторы можно разделить на две категории: мониторы текущего состояния системы и мониторы изменения состояния. Мониторы первой категории показывают моментальный срез состояния, обновляющийся на каждой моделирования. Они представлены популяционными пирамидами, показывающими половозрастное распределение агентов в популяции, распределение фенотипов и распределение генотипов (Рис. 4.2.3). Мониторы второй категории отображают состояние системы за все время моделирования (Рис. 4.2.4). Технически они реализованы обычными графиками и представляют размеры всей популяции, женской и мужской субпопуляции, а также глухой и слышащей субпопуляций. Отдельно представлен график поддержки глухоты среди глухих агентов – он показывает среднюю ассортативность среди фенотипически глухих агентов.

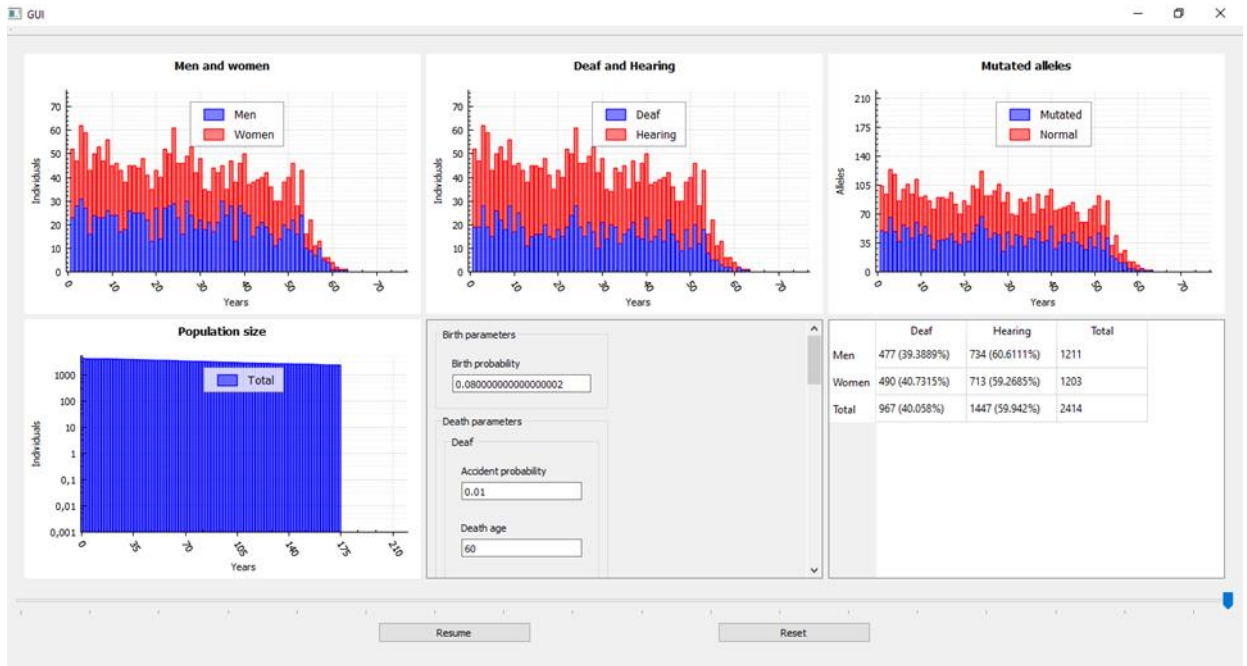


Рис. 4.2.3. Графический интерфейс пользователя программы PGD-S. Мониторы первой категории – половозрастное распределение агентов в популяции, распределения генотипов и фенотипов.

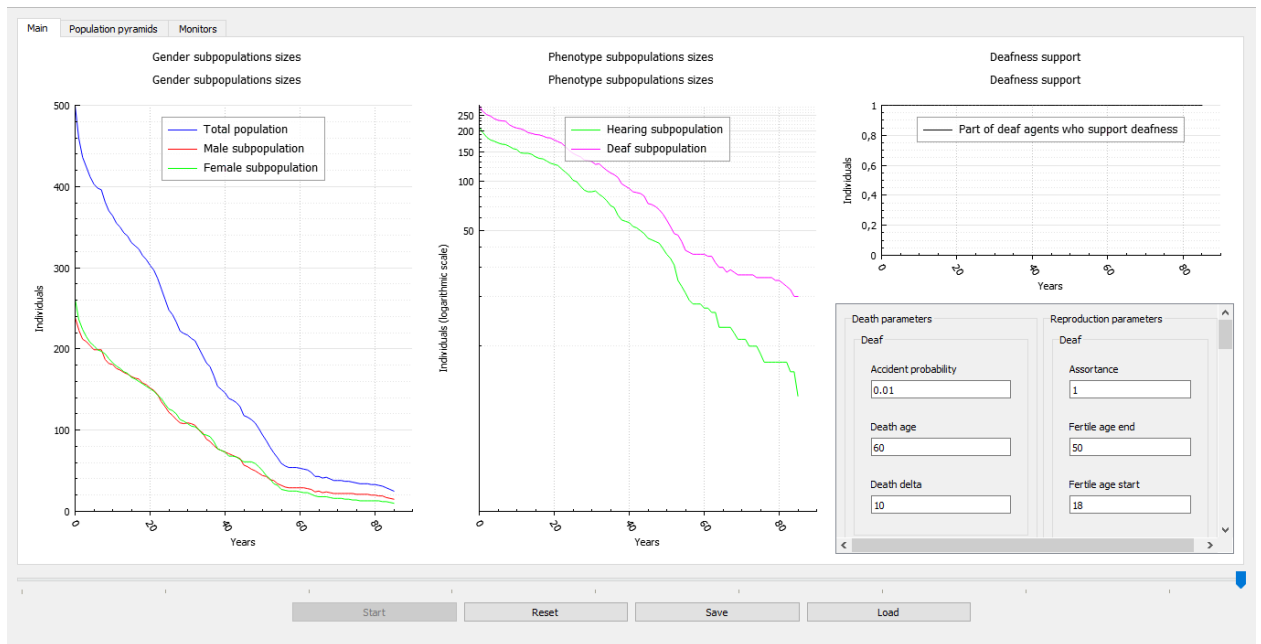


Рис. 4.2.4. Графический интерфейс пользователя программы PGD-S. Мониторы второй категории, динамика системы за всё время моделирования.

В программе PGD-S предусмотрен также режим мультизапуска. В этом режиме происходит пакетный запуск численных экспериментов с одинаковыми параметрами. В силу стохастического подхода к

моделированию, результаты разных запусков могут отличаться, усреднение же некоторого множества результатов позволяет отфильтровать стохастические выбросы и получить более статистически обоснованную картину.

4.2.4. Моделирование трендов распространения глухоты в программе PGD-S

С использованием программы PGD-S была проведена серия вычислительных экспериментов по моделированию трендов распространения глухоты. Были рассмотрены три базовых варианта модели – со стартовой численностью популяции 10000 индивидуумов, 50000 и 100000.

Для каждого варианта рассчитывалось четыре численных эксперимента, различавшихся между собой значениями параметров разборчивости в субпопуляциях глухих (*as_deaf*) и слышащих (*as_hearing*) индивидуумов (Exp1: *as_deaf* = 0.3, *as_hearing* = 0.3; Exp2: *as_deaf* = 0.3, *as_hearing* = 0.6; Exp3: *as_deaf* = 0.6, *as_hearing* = 0.3; Exp4: *as_deaf* = 0.9, *as_hearing* = 0.6). Другие важные параметры модели (общие для всех вариантов): владение жестовым языком среди глухих – 50%, веса *Weight Sign language Deaf*, *Weight Phenotype Deaf* и *Weight Speaking ability Deaf* равны, соответственно, 20, 10 и 1; веса *Weight Sign language Hearing*, *Weight Phenotype Hearing* и *Weight Speaking ability Hearing* равны, соответственно, 10, 10 и 10. Наличие мутантного аллеля в популяции в момент старта моделирования – 10% в гетерозиготном состоянии, 0.25% в гомозиготном состоянии.

Каждый численный эксперимент был рассчитан в нескольких репликах (число реплик варьировало от 5 до 20). Общее время моделирования составило 500 модельных лет (время расчёта на компьютере Intel Core i5-3570K, 16 Gb RAM варьировало от 5 до 30 минут на реплику). После окончания моделирования оценивалась частота «глухих» аллелей в популяции. Суммирующие результаты моделирования данные приведены ниже на рисунке 4.2.5. Из рисунка видно, что увеличение ассортативности в субпопуляции глухих может приводить к уменьшению частоты мутантного

аллеля в популяции в тех случаях, если высокая ассортативность сопряжена с относительно высоким весом владения жестовым языком при обще невысокой доле владеющих им.

Полученные контринтуитивные данные моделирования отличаются от модельных данных Нэнса и соавторов, главным образом именно из-за введения в нашу модель дополнительных параметров (владение жестовым языком и др.). Детальное изучение влияния всех перечисленных выше параметров на прогнозируемую частоту мутантного аллеля в популяции требует дополнительных усилий и временных ресурсов. Тем не менее, программно-методический базис для таких исследований создан в рамках выполнения данного проекта.

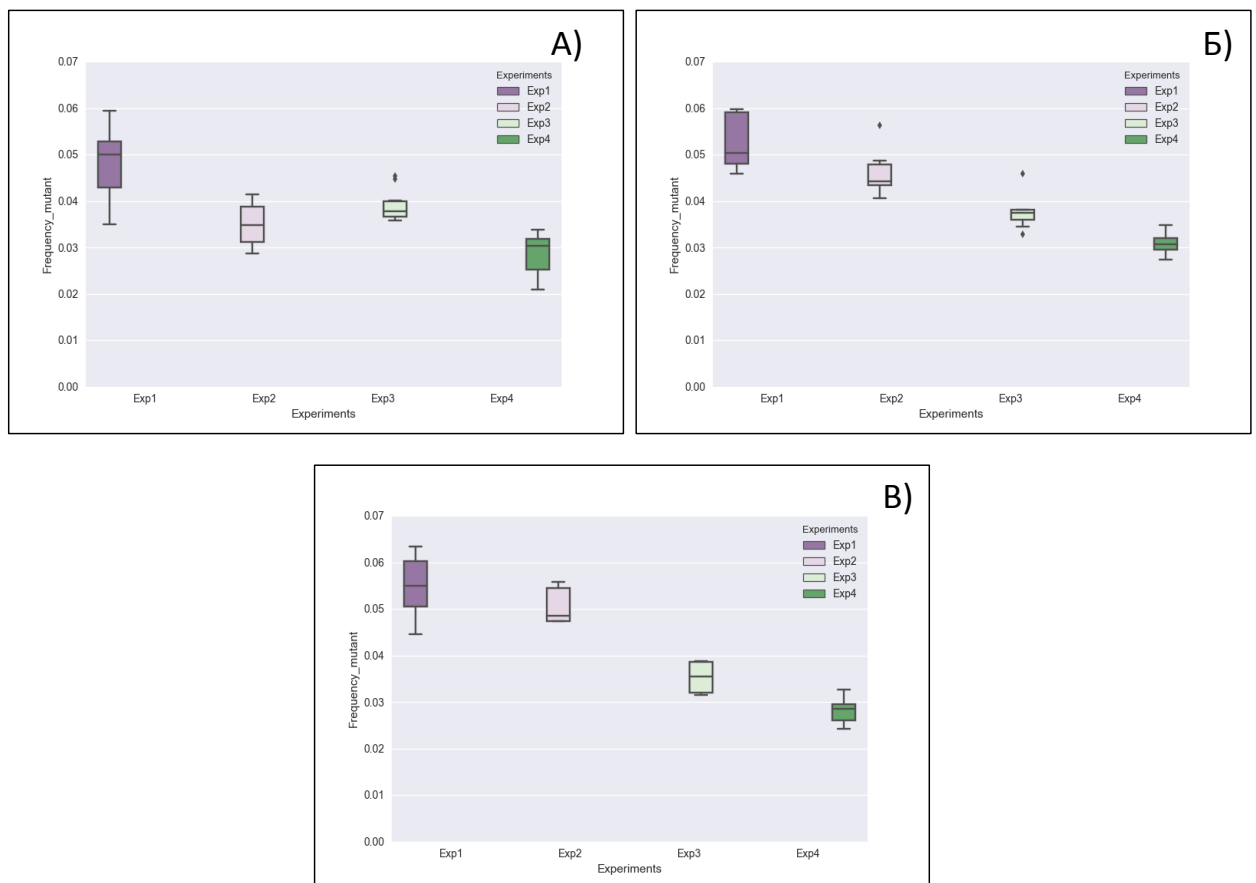


Рисунок 4.2.5. Распределения частот мутантного «глухого» аллеля в популяции для четырёх численных экспериментов (Exp1-Exp4 – пояснения в тексте) по прошествии 500 лет модельного времени. Стартовый размер популяции: А) 10000 человек; Б) 50000 человек; В) 100000 человек.

Заключение к главе 4

В данной главе приведены методы моделирования биологических систем с учётом социальных факторов, а также результаты исследования моделей социобиологических систем «Агрессор-Защитник» и её расширенной версии, модели распространения генетической глухоты в популяциях человека. Приведено описание программного комплекса Population Genetics of Deafness – Simulator (PGD-S), который является расширением программного комплекса ДЭК, описанного в третьей главе диссертации.

Получены контринтуитивные данные, отличающиеся от данных, приведённых в литературе (в частности, в (Nance et al., 2000; Nance, Kearsey, 2004)). Мы считаем, что наши результаты, расширяющие понимание популяционной генетики глухоты (Romanov et al., 2022), главным образом именно из-за введения в модель PGD-S таких дополнительных параметров к владение жестовым языком и др. (т.е. «социально-ориентированных» параметров). Детальное изучение влияния всех перечисленных выше параметров на прогнозируемую частоту мутантного аллеля в популяции требует дополнительных усилий и временных ресурсов. Дальнейшее развитие модели PGD-S может быть связано с увеличением количества рассматриваемых генетических локусов, встречающихся в разных популяциях, и представляющих интерес, в отношении брачной структуры. Расширение географии моделируемых регионов, мегаполисов и т.п. не может быть реализовано простым увеличением числа агентов в моделируемой популяции. Для этого необходимы данные о взаимодействии больших групп людей. И хотя наши вычислительные эксперименты были посвящены относительно небольшим сообществам, программно-методческая база для таких расширенных исследований уже создана в рамках данной диссертации. После того, как мы получим более четкое понимание того, как функционируют взаимодействия внутри больших популяций, можно будет изучать более крупные сообщества.

Заключение

В диссертации сформулированы и решены постановки новых задач математического и компьютерного моделирования сложных, иерархически организованных биологических систем. Разработанные модели учитывают такие уровни биологической организации как метаболический, генетический, клеточный, организменный, популяционный и экологический. Методы моделирования, представленные в диссертации, вносят значимый вклад в развитие т.н. «моделирования нового поколения» (англ. next-generations modelling), принципы которого были сформулированы классиками экологического моделирования Уолкера Гримма и Уты Бергер в работе (Grimm, Berger, 2016). Среди главных принципов – структурный реализм, эмерджентность и предсказательность, построение модели «от первых принципов», гетерогенность и динамичность среды обитания, микроэволюция, взаимосвязь иерархических уровней и использование стандартизированных подмоделей. Модели, представленные в нашей работе, не просто удовлетворяют вышеописанным свойствам, но и расширяют спектр областей их применения в биологии, учитывая дополнительные уровни системной организации, не упомянутые Гриммом и Бергер, в частности, молекулярно-генетический и социальный уровни. Отметим, что первые статьи с этими результатами были опубликованы нами за четыре года до программной статьи Грима и Бергер (Lashin et al., 2012a, 2012b; Lashin, Matushkin, 2012).

Разработанные методы моделирования были реализованы в виде программных комплексов «Гаплоидный Эволюционный Конструктор» (ГЭК), «Microcosm», «Диплоидный Эволюционный Конструктор» (ДЭК), «Population Genetics of Deafness – Simulator» (PGD-S), представленных в данной диссертации, а также в ряде других программ и программных комплексов, не представленных в данной работе.

Разработанные в диссертации методы построения и численного анализа иерархических многослойных моделей биологических систем, а также реализованные на их основе программные комплексы являются эффективными инструментами для решения содержательных биологических задач как фундаментальной, так и прикладной направленности.

Программный комплекс (ГЭК) предназначен для моделирование микробных сообществ, населяющих гетерогенные среды обитания (в одно-, двух- и трёхмерном случае), их функционирование и эволюцию. При этом рассматриваются такие процессы как диффузия и перенос под действием потока для метаболитов, а для клеток – пассивное (с помощью потока) и активное (с помощью хемотаксиса) перемещение. Созданы специальный язык описания моделей ГЭК, графический интерфейс пользователя и веб-приложение, включающее в себя библиотеку стандартных моделей ГЭК.

Разработаны модули ГЭК для высокопроизводительного моделирования на многоядерных кластерах. Комплексное тестирование разработанных высокопроизводительных модулей ГЭК показало следующие результаты: эффективность распараллеливания до 180% с использованием небольшого числа процессов на MPI на сложных задачах; ускорение до 40 раз с использованием большого числа процессов MPI; ускорение до 3 раз с использованием QtConcurrent в графическом приложении на 6-ядерном процессоре.

Выполнена интеграция ГЭК с существующими средствами моделирования генных сетей и базами данных (Biomodels, CellML, SabioRK, MAMMOTH) с помощью стандарта SBML, а также с базой данных наименований химических веществ ChEBI.

Программный комплекс Microcosm, реализованный в агентно-ориентированной парадигме (с разрешением до отдельной клетки), является дальнейшим развитием ПК ГЭК и предназначен для исследования микробных

сообществ относительно небольшого размера, в которых возможно появление сложноорганизованных паттернов, связанных с рассинхронизацией клеточных процессов и/или особых сигнальных путей, участвующих в формировании т.н. чувства кворума.

Программный комплекс ДЭК предназначен для моделирования эволюционно-генетических процессов в популяциях высокоорганизованных диплоидных организмов, в частности, людей. Реализованные на архитектуре ДЭК расширения программного комплекса позволяют моделировать в популяциях человека влияние социальных аспектов на распространение генетических болезней. В частности, для моделирования распространения генетической глухоты на базе ДЭК был создан программный комплекс PGD-S. Так же на базе ДЭК были реализованы модель «Агрессор – Защитник» и её расширенная версия, которые могут быть использованы для исследования влияния социальных факторов, таких как стресс, в биологических системах.

Для программных комплексов ДЭК и PGD-S также были разработаны соответствующие языки описания моделей и графические интерфейсы пользователя. Разработаны модули поддержки высокопроизводительных вычислений. Версия ПК ДЭК с использованием OpenMP была протестирована на различных компьютерах и показала значительное ускорение как на обычном настольном компьютере (до 13 раз), так и на кластере Сибирского Суперкомпьютерного Центра СО РАН (до 60-70 раз), что позволяет значительно сократить время, затрачиваемое на проведение расчетов ДЭК.

Использование подходов «многослойного» моделирования позволяет получать нетривиальные и контринтуитивные биологические результаты, касающиеся влияния социальных и экологических факторов на динамику генетического разнообразия популяций и сообществ, и наоборот, влияния генетической структуры популяций на эволюцию экосистем. Так, например, в отмеченное нами в результатах моделирования микробных сообществ схожее влияние локальных минимумов, относящихся к совершенно разным

иерархическим уровням биологической организации (генетическому и экологическому), было получено не за счёт использования одинаковых или даже похожих уравнений, а за счёт построения механистической компьютерной модели (т.е. основанное на упоминаемых выше так называемых первых принципах), что свидетельствует о существовании фундаментальных паттернов развития биологических систем. Эти паттерны могут успешно выявляться с помощью методов комплексного моделирования биологических систем, представленных в данной диссертации.

Разработанные в диссертации методы и программные комплексы, предназначенные для конструирования и численного исследования моделей комплексных иерархически организованных биологических систем, **рекомендуются** для использования специалистам-биологам в таких областях как микробиология (исследование почв, расчёт биореакторов, особенно многовидовых), археология (численное моделирование различных эволюционных сценариев для популяций древних людей), эпидемиология (моделирование распространения заболеваний, в т.ч. генетических).

В качестве **перспективы дальнейшей разработки темы** предполагается расширение функциональности разработанных программных комплексов, прежде всего, в части интеграции с современными базами математических моделей биологических систем, с современными источниками данных. Необходима интеграция современных методов решения обратных задач моделирования для более полной верификации разрабатываемых моделей. Планируется расширение спектра применения разработанных методов моделирования и программ, прежде всего, в части усиления акцента на решение прикладных задач биологии, а именно – микробной биотехнологии, биомедицины и эпидемиологии.

Выводы

1. Разработана методика компьютерного моделирования микробных сообществ и на её основе - программный комплекс «Гаплоидный Эволюционный Конструктор» (ГЭК), позволяющие моделировать пространственно-распределённые микробные сообщества, их функционирование и эволюцию, в одно-, двух- и трёхмерном случае, соответственно. При этом рассматриваются, в том числе, такие процессы как диффузия и перенос под действием протока для метаболитов, а для клеток – пассивное (с помощью протока) и активное (с помощью хемотаксиса) перемещение.

2. С помощью ГЭК проведено сравнительное моделирование изменения сложности генома в прокариотических сообществах в различных условиях среды обитания:

а. Показано, что пространственная организация среды обитания формирует ландшафт давления отбора, опосредуя тем самым эволюционные процессы, происходящие в экосистеме.

б. Установлено, что экологические паттерны самоорганизации микробных сообществ способствуют поддержанию различных стратегий, лежащих в основе сценариев как усложнения, так и упрощения метаболизма.

с. Показано, что разница в приспособленности для различных аллельных комбинаций, определяя форму ландшафта приспособленности, задает оптимальную для преодоления его долин частоту мутаций. Более того, в зависимости от характера ландшафта приспособленности оптимальными для достижения пика (наибольшей средней приспособленности) оказываются либо градуальный, либо сальтационный эволюционные режимы.

3. С помощью ГЭК проведено моделирование влияния бактериофагов на эволюцию бактериальных сообществ в пространственно-распределённых средах обитания:

а. Показано, что фаговая инфекция влияет на эволюцию микробного сообщества, замедляя видообразование и стабилизируя систему в целом.

б. Установлено, что эффективность влияния варьирует в зависимости от пространственно-экологических факторов, а также от состояния сообщества в момент инвазии фага.

4. С помощью ГЭК исследованы эволюционная конкуренция между бактериями с различным типом движения и эволюционные преимущества/недостатки хемотаксиса в зависимости от экологических условий:

а. Показано, что хемотаксис является фактором, препятствующим внедрению новых форм и снижающим общую биомассу сообщества.

б. Показано, что баланс энергетических затрат на движение и доступности питательных веществ определяет, получают ли бактерии с высокой хемотаксической активностью преимущество над низкоподвижными.

с. Для низкоподвижных бактерий показана возможность сохранения преимуществ при заселении пространства вблизи источника питательных веществ, и создания вокруг себя «биотической пустыни», которую не могут преодолеть бактерии с высокой хемотаксической активностью.

5. Разработан программный комплекс «Диплоидный Эволюционный Конструктор» (ДЭК), позволяющий моделировать эволюционно-генетические процессы в популяциях диплоидных организмов с половым размножением. С помощью ДЭК:

а. Построены и проанализированы компьютерные модели процессов инбридинга и аутбридинга в популяции. Показано, что давление инбридинга на группу генов, составляющих одну генную сеть, способно влиять на эволюцию и генетическое разнообразие другой группы генов. Причём этот эффект в диплоидных геномах сильно зависит от типа генетического доминирования.

б. Построены и исследованы модели возникновения и эволюции регуляторных контуров генных сетей (в частности, петель отрицательной и

положительной обратной связи). Показано, что популяции, содержащие организмы с отрицательной обратной связью, сохраняют биоразнообразие значительно дольше (по сравнению с популяциями без обратной связи), что делает их, с одной стороны, более стабильными и устойчивыми, а с другой - обладающими большим потенциалом изменчивости, который может быть изменен в будущем.

б. На основе программной архитектуры ДЭК разработан программный комплекс «PGD-S» («Популяционная генетика глухоты – симулятор»), позволяющий моделировать влияние социальных аспектов на распространение генетической глухоты (в частности, аутосомно-рецессивной глухоты 1A (DFNB1A)) в популяциях человека. С помощью PGD-S:

а. Показано, что распространенность DFNB1A в изолированной популяции может быть резко увеличена при частых ассортативных браках за относительно короткий промежуток времени под давлением ослабленного отбора.

б. Показано, что в современных условиях доля рецессивных гомозигот быстро (за несколько поколений) достигает максимума и затем непрерывно снижается.

с. Показано, что в долгосрочной перспективе этот эффект может быть нивелирован ростом социального равенства глухих людей, о чем свидетельствуют результаты моделирования нейтрального отбора.

Список литературы

1. Abar S., Theodoropoulos G.K., Lemarinier P., O'Hare G.M.P. Agent Based Modelling and Simulation tools: A review of the state-of-art software. *Computer Science Review*. 2017;24:13–33.
<https://doi.org/10.1016/j.cosrev.2017.03.001>
2. Adler J. Chemotaxis in bacteria. *Journal of supramolecular structure*. 1976;4:305–317. <https://doi.org/10.1146/annurev.bi.44.070175.002013>
3. Afonnikov D.A., Oshchepkov D.Y., Kolchanov N.A. Detection of conserved physico-chemical characteristics of proteins by analyzing clusters of positions with co-ordinated substitutions. *Bioinformatics*. 2001;17:1035–1046. <https://doi.org/10.1093/bioinformatics/17.11.1035>
4. Aita T., Husimi Y. Fitness Spectrum Among Random Mutants on Mt. Fuji-Type Fitness Landscape. *Journal of Theoretical Biology*. 1996;182:469–485.
<https://doi.org/10.1006/jtbi.1996.0189>
5. Akberdin I.R.I.R., Ozonov E.A.E.A., Mironova V.V.V., Omelyanchuk N.A.N.A., Likhoshvai V.A.V.A., Gorpinchenko D.N.D.N., Kolchanov N.A.N.A. A cellular automaton to model the development of primary shoot meristems of *Arabidopsis thaliana*. *Journal of Bioinformatics and Computational Biology*. 2007;05:641–650.
<https://doi.org/10.1142/S0219720007002862>
6. Aktipis C.A. Is cooperation viable in mobile organisms? Simple Walk Away rule favors the evolution of cooperation in groups. *Evolution and Human Behavior*. 2011;32:263–276.
<https://doi.org/10.1016/j.evolhumbehav.2011.01.002>
7. Allen A.P., Gillooly J.F. Towards an integration of ecological stoichiometry and the metabolic theory of ecology to better understand nutrient cycling. *Ecology letters*. 2009;12:369–84. <https://doi.org/10.1111/j.1461-0248.2009.01302.x>
8. Amann R., Ludwig W., Schleifer K. Phylogenetic identification and in situ

- detection of individual microbial cells without cultivation. *Microbiol. Rev.* 1995;59:143–169.
9. Amores A., Catchen J., Ferrara A., Fontenot Q., Postlethwait J.H. Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication. *Genetics.* 2011;188:799–808. <https://doi.org/10.1534/genetics.111.127324>
 10. Arnos K.S., Welch K.O., Tekin M., Norris V.W., Blanton S.H., Pandya A., Nance W.E. A Comparative Analysis of the Genetic Epidemiology of Deafness in the United States in Two Sets of Pedigrees Collected More than a Century Apart. *American Journal of Human Genetics.* 2008;83:200–207. <https://doi.org/10.1016/j.ajhg.2008.07.001>
 11. Asakura S., Honda H. Two-state model for bacterial chemoreceptor proteins. The role of multiple methylation. *Journal of Molecular Biology.* 1984;176:349–367. [https://doi.org/10.1016/0022-2836\(84\)90494-7](https://doi.org/10.1016/0022-2836(84)90494-7)
 12. Axelrod R. The Dissemination of Culture. *Journal of Conflict Resolution.* 1997;41:203–226. <https://doi.org/10.1177/0022002797041002001>
 13. Barbuti R., Gori R., Milazzo P., Nasti L. A survey of gene regulatory networks modelling methods: from differential equations, to Boolean and qualitative bioinspired models. *Journal of Membrane Computing.* 2020;2:207–226. <https://doi.org/10.1007/s41965-020-00046-y>
 14. Barnabas J., Raj P. The human body: A digital twin of the cyber physical systems. pp. 219–246. <https://doi.org/10.1016/bs.adcom.2019.09.004>
 15. Bataillon T., Mailund T., Thorlacius S., Steingrímsson E., Rafnar T., Halldorsson M.M., Calian V., Schierup M.H. The effective size of the Icelandic population and the prospects for LD mapping: Inference from unphased microsatellite markers. *European Journal of Human Genetics.* <https://doi.org/10.1038/sj.ejhg.5201669>
 16. Beardmore R.E., Gudelj I., Lipson D.A., Hurst L.D. Metabolic trade-offs and the maintenance of the fittest and the flattest. *Nature.* 2011;472:342–6. <https://doi.org/10.1038/nature09905>

17. Beerli P., Palczewski M. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*. <https://doi.org/10.1534/genetics.109.112532>
18. Berg H.C., Purcell E.M. Physics of chemoreception. *Biophysical Journal*. 1977;20:193–219. [https://doi.org/10.1016/S0006-3495\(77\)85544-6](https://doi.org/10.1016/S0006-3495(77)85544-6)
19. Bergholz P.W., Wiedmann M. Toward agent-based models for pre-harvest food safety. *IBM Journal of Research and Development*. 2016;60:8:1-8:13. <https://doi.org/10.1147/JRD.2016.2596378>
20. Beslon G., Parsons D.P., Sanchez-Dehesa Y., Peña J.-M., Knibbe C. Scaling laws in bacterial genomes: a side-effect of selection of mutational robustness? *Bio Systems*. 2010;102:32–40. <https://doi.org/10.1016/j.biosystems.2010.07.009>
21. Bezzi M., Celada F., Ruffo S., Seiden P.E. The transition between immune and disease states in a cellular automaton model of clonal immune response. *Physica A: Statistical Mechanics and its Applications*. 1997;245:145–163. [https://doi.org/10.1016/S0378-4371\(97\)00290-2](https://doi.org/10.1016/S0378-4371(97)00290-2)
22. Biess A., Nagurka M., Flash T. Simulating discrete and rhythmic multi-joint human arm movements by optimization of nonlinear performance indices. *Biological cybernetics*. 2006;95:31–53. <https://doi.org/10.1007/s00422-006-0067-7>
23. Bleidorn C. Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Systematics and Biodiversity*. 2016;14:1–8. <https://doi.org/10.1080/14772000.2015.1099575>
24. Blinov M.L., Faeder J.R., Goldstein B., Hlavacek W.S. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*. 2004;20:3289–3291. <https://doi.org/10.1093/bioinformatics/bth378>
25. Blyth C.R. On Simpson’s Paradox and the Sure-Thing Principle. *Journal of the American Statistical Association*. 1972;67:364. <https://doi.org/10.2307/2284382>

26. Bonabeau E. Agent-based modeling: methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;99 Suppl 3:7280–7.
<https://doi.org/10.1073/pnas.082080899>
27. Bordbar A., Monk J.M., King Z.A., Palsson B.O. Constraint-based models predict metabolic and associated cellular functions. *Nature Reviews Genetics*. 2014;15:107–120. <https://doi.org/10.1038/nrg3643>
28. Bornstein B., Keating S., Jouraku A., Hucka M. LibSBML: an API library for SBML. *Bioinformatics*.
29. Bray D., Bourret R.B. Computer analysis of the binding reactions leading to a transmembrane receptor-linked multiprotein complex involved in bacterial chemotaxis. *Molecular Biology of the Cell*. 1995;6:1367–1380.
<https://doi.org/10.1091/mbc.6.10.1367>
30. Bray D., Bourret R.B., Simon M.I. Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis. *Molecular Biology of the Cell*. 1993;4:469–482. <https://doi.org/10.1091/mbc.4.5.469>
31. Bray D., Levin M.D., Lipkow K. The Chemotactic Behavior of Computer-Based Surrogate Bacteria. *Current Biology*. 2007;17:12–19.
<https://doi.org/10.1016/j.cub.2006.11.027>
32. Bren A., Eisenbach M. How Signals Are Heard during Bacterial Chemotaxis : Protein-Protein Interactions in Sensory Signal Propagation. *Journal of bacteriology*. 2000;182:6865–6873.
<https://doi.org/10.1128/JB.182.24.6865-6873.2000>. Updated
33. Bridier A., Piard J.-C., Pandin C., Labarthe S., Dubois-Brissonnet F., Briandet R. Spatial Organization Plasticity as an Adaptive Driver of Surface Microbial Communities. *Frontiers in Microbiology*. 2017;8.
<https://doi.org/10.3389/fmicb.2017.01364>
34. Brüningk S.C., Ziegenhein P., Rivens I., Oelfke U., Haar G. ter. A cellular automaton model for spheroid response to radiation and hyperthermia treatments. *Scientific Reports*. 2019;9:17674.

- <https://doi.org/10.1038/s41598-019-54117-x>
35. Buchanan D. *Social Epidemiology*. New York: Oxford University Press, 2003.
 36. Callahan B.J., Fukami T., Fisher D.S. Rapid evolution of adaptive niche construction in experimental microbial populations. *Evolution*. 2014;68:3307–3316. <https://doi.org/10.1111/evo.12512>
 37. Carroll S.B. Chance and necessity: the evolution of morphological complexity and diversity. *Nature*. 2001;409:1102–1109. <https://doi.org/10.1038/35059227>
 38. Chen L., Kulasiri D., Samarasinghe S. A Novel Data-Driven Boolean Model for Genetic Regulatory Networks. *Frontiers in Physiology*. 2018;9. <https://doi.org/10.3389/fphys.2018.01328>
 39. Chew S.C., Kundukad B., Seviour T., van der Maarel J.R.C., Yang L., Rice S.A., Doyle P., Kjelleberg S. Dynamic Remodeling of Microbial Biofilms by Functionally Distinct Exopolysaccharides. *mBio*. 2014;5:e01536-14-e01536-14. <https://doi.org/10.1128/mBio.01536-14>
 40. Christodoulou F., Raible F., Tomer R., Simakov O., Trachana K., Klaus S., Snyman H., Hannon G.J., Bork P., Arendt D. Ancient animal microRNAs and the evolution of tissue identity. *Nature*. 2010;463:1084–1088. <https://doi.org/10.1038/nature08744>
 41. Clifford P., Sudbury A. A model for spatial conflict. *Biometrika*. 1973;60:581–588. <https://doi.org/10.1093/biomet/60.3.581>
 42. Colasanti R.L., Hunt R., Watrud L. A simple cellular automaton model for high-level vegetation dynamics. *Ecological Modelling*. 2007;203:363–374. <https://doi.org/10.1016/j.ecolmodel.2006.12.039>
 43. Dayhoff M.O., Schwartz R.M., Orcutt B.C. A model of evolutionary change in proteins, in: Dayhoff M.O. (Ed.), *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3. Washington, DC, DC: Natl. Biomed. Res. Found., 1978, pp. 345–352.
 44. de Jong H. *Modeling and Simulation of Genetic Regulatory Systems: A*

- Literature Review. *Journal of Computational Biology*. 2002;9:67–103.
<https://doi.org/10.1089/10665270252833208>
45. DeAngelis D.L., Mooij W.M. Individual-Based Modeling of Ecological and Evolutionary Processes 1. *Annual Review of Ecology, Evolution, and Systematics*. 2005;36:147–168.
<https://doi.org/10.1146/annurev.ecolsys.36.102003.152644>
46. Deffuant G., Neau D., Amblard F., Weisbuch G. Mixing beliefs among interacting agents. *Advances in Complex Systems*. 2000;03:87–98.
<https://doi.org/10.1142/S0219525900000078>
47. Deschamps S., Campbell M.A. Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. *Molecular Breeding*. 2010;25:553–570. <https://doi.org/10.1007/s11032-009-9357-9>
48. Dobretsov N., Kolchanov N., Rozanov A., Zavarzin G. (Eds.). *Biosphere Origin and Evolution*, *Biosphere Origin and Evolution*. Boston, MA: Springer US, 2008. <https://doi.org/10.1007/978-0-387-68656-1>
49. Doebeli M., Dieckmann U. Adaptive Dynamics of Speciation: Spatial Structure, in: *Adaptive Speciation*. Cambridge: Cambridge University Press, 2004, pp. 140–167.
50. Doebeli M., Dieckmann U. Speciation along environmental gradients. *Nature*. 2003;421:259–264. <https://doi.org/10.1038/nature01274>
51. Domingo E., Holland J.J. RNA VIRUS MUTATIONS AND FITNESS FOR SURVIVAL. *Annual Review of Microbiology*. 1997;51:151–178.
<https://doi.org/10.1146/annurev.micro.51.1.151>
52. Drake J.W., Charlesworth B., Charlesworth D., Crow J.F. Rates of Spontaneous Mutation. *Genetics*. 1998;148:1667–1686.
53. Drasdo D., Bode J., Dahmen U., Dirsch O., Dooley S., Gebhardt R., Ghallab A., Godoy P., Häussinger D., Hammad S., Hoehme S., Holzhütter H.-G., Klingmüller U., Kuepfer L., Timmer J., Zerial M., Hengstler J.G. The virtual liver: state of the art and future perspectives. *Archives of Toxicology*. 2014;88:2071–2075. <https://doi.org/10.1007/s00204-014-1384-6>

54. El-Sayed A.M., Scarborough P., Seemann L., Galea S. Social network analysis and agent-based modeling in social epidemiology. *Epidemiologic perspectives & innovations : EP+I*. 2012;9:1. <https://doi.org/10.1186/1742-5573-9-1>
55. Emonet T., Macal C.M., North M.J., Wickersham C.E., Cluzel P. AgentCell: a digital single-cell assay for bacterial chemotaxis. *Bioinformatics (Oxford, England)*. 2005;21:2714–21. <https://doi.org/10.1093/bioinformatics/bti391>
56. Esteban P.G., Rodríguez-Patón A. Simulating a Rock-Scissors-Paper Bacterial Game with a Discrete Cellular Automaton, in: Ferrández J.M., Álvarez Sánchez J.R., de la Paz F., Toledo F.J. (Eds.), *New Challenges on Bioinspired Applications, Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 363–370. <https://doi.org/10.1007/978-3-642-21326-7>
57. Ewens W. Mathematical population genetics: I. Theoretical introduction.
58. Fagundes N.J.R., Ray N., Beaumont M., Neuenschwander S., Salzano F.M., Bonatto S.L., Excoffier L. Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences of the United States of America*. 2007;104:17614–9. <https://doi.org/10.1073/pnas.0708280104>
59. Falke J.J., Bass R.B., Butler S.L., Chervitz S. a., Danielson M. a. THE TWO-COMPONENT SIGNALING PATHWAY OF BACTERIAL CHEMOTAXIS: A Molecular View of Signal Transduction by Receptors, Kinases, and Adaptation Enzymes. *Annu Rev Cell Dev Biol*. 1997;457–512. <https://doi.org/10.1146/annurev.cellbio.13.1.457>. THE
60. Fath B.D., Scharler U.M., Ulanowicz R.E., Hannon B. Ecological network analysis: network construction. *Ecological Modelling*. 2007;208:49–55. <https://doi.org/10.1016/j.ecolmodel.2007.04.029>
61. Feng Z., Dieckmann U., Levin S. (Eds.). *DIMACS Series in Discrete Mathematics and Theoretical Computer Science, DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. American

- Mathematical Society, 2006.
62. Ferrer J., Prats C., López D. Individual-based modelling: an essential tool for microbiology. *Journal of biological physics*. 2008;34:19–37.
<https://doi.org/10.1007/s10867-008-9082-3>
63. Fisher R.A. The genetical theory of natural selection. Oxford: Clarendon Press, 1930.
64. Foster P.L. Stress responses and genetic variation in bacteria. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 2005;569:3–11. <https://doi.org/10.1016/j.mrfmmm.2004.07.017>
65. France S.C., Hessler R.R., Vrijenhoek R.C. Genetic differentiation between spatially-disjunct populations of the deep-sea, hydrothermal vent-endemic amphipod *Ventiella sulfuris*. *Marine Biology*. 1992;114:551–559.
<https://doi.org/10.1007/BF00357252>
66. Frankel N.W., Pontius W., Dufour Y.S., Long J., Hernandez-Nunez L., Emonet T. Adaptability of non-genetic diversity in bacterial chemotaxis. *eLife*. 2014;3. <https://doi.org/10.7554/eLife.03526>
67. Galam S. The September 11 attack: A percolation of individual passive support. *The European Physical Journal B*. 2002;26:269–272.
<https://doi.org/10.1140/epjb/e20020089>
68. Gatenby R.A., Vincent T.L. Application of quantitative models from population biology and evolutionary game theory to tumor therapeutic strategies. *Molecular cancer therapeutics*. 2003;2:919–27.
69. Gavrillets S., Li H., Vose M.D. PATTERNS OF PARAPATRIC SPECIATION. *Evolution*. 2000;54:1126–1134.
<https://doi.org/10.1111/j.0014-3820.2000.tb00548.x>
70. Gevertz J.L., Torquato S. Modeling the effects of vasculature evolution on early brain tumor growth. *Journal of Theoretical Biology*. 2006;243:517–531. <https://doi.org/10.1016/j.jtbi.2006.07.002>
71. Ghattargi V.C., Nimonkar Y.S., Sape K., Prakash O., Suryavanshi M. V., Shouche Y.S., Meti B.S., Pawar S.P. Functional and Comparative Genomics

- of Niche-Specific Adapted Actinomycetes *Kocuria rhizophila* Strain D2 Isolated from Healthy Human Gut. *bioRxiv*.
72. Ginovart M., López D., Valls J. INDISIM, an individual-based discrete simulation model to study bacterial cultures. *Journal of theoretical biology*. 2002;214:305–19. <https://doi.org/10.1006/jtbi.2001.2466>
73. Giovannoni S.J. SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Annual Review of Marine Science*. 2017;9:231–255. <https://doi.org/10.1146/annurev-marine-010814-015934>
74. Giovannoni S.J. Genome Streamlining in a Cosmopolitan Oceanic Bacterium. *Science*. 2005;309:1242–1245. <https://doi.org/10.1126/science.1114057>
75. Glazier J.A., Graner F. Simulation of the differential adhesion driven rearrangement of biological cells. *Physical Review E*. 1993;47:2128–2154. <https://doi.org/10.1103/PhysRevE.47.2128>
76. Gogarten J.P., Townsend J.P. Horizontal gene transfer, genome innovation and evolution. *Nature reviews. Microbiology*. 2005;3:679–87. <https://doi.org/10.1038/nrmicro1204>
77. Gokhale C.S., Iwasa Y., Nowak M.A., Traulsen A. The pace of evolution across fitness valleys. *Journal of Theoretical Biology*. 2009;259:613–620. <https://doi.org/10.1016/j.jtbi.2009.04.011>
78. Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., Li H., Zhai W., Fritz M.H.Y., Hansen N.F., Durand E.Y., Malaspina A.S., Jensen J.D., Marques-Bonet T., Alkan C., Prufer K., Meyer M., Burbano H. a, Good J.M., Schultz R., Aximu-Petri A., Butthof A., Hober B., Hoffner B., Siegemund M., Weihmann A., Nusbaum C., Lander E.S., Russ C., Novod N., Affourtit J., Egholm M., Verna C., Rudan P., Brajkovic D., Kucan Z., Gusic I., Doronichev V.B., Golovanova L. V, Lalueza-Fox C., de la Rasilla M., Fortea J., Rosas A., Schmitz R.W., Johnson P.L.F., Eichler E.E., Falush D., Birney E., Mullikin J.C., Slatkin M., Nielsen R., Kelso J., Lachmann M., Reich D., Paabo S. A Draft

- Sequence of the Neandertal Genome. *Science*. 2010;328:710–722.
<https://doi.org/10.1126/science.1188021>
79. Grimm V. Ecological Models: Individual-Based Models, in: Encyclopedia of Ecology. Elsevier, 2019, pp. 65–73. <https://doi.org/10.1016/B978-0-12-409548-9.11144-3>
80. Grimm V., Berger U. Structural realism, emergence, and predictions in next-generation ecological modelling: Synthesis from a special issue. *Ecological Modelling*. 2016;326:177–187.
<https://doi.org/10.1016/j.ecolmodel.2016.01.001>
81. Grimm V., Berger U., Bastiansen F., Eliassen S., Ginot V., Giske J., Goss-Custard J., Grand T., Heinz S.K., Huse G., Huth A., Jepsen J.U., Jørgensen C., Mooij W.M., Müller B., Pe'er G., Piou C., Railsback S.F., Robbins A.M., Robbins M.M., Rossmanith E., Rüger N., Strand E., Souissi S., Stillman R. a., Vabø R., Visser U., DeAngelis D.L. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*. 2006;198:115–126. <https://doi.org/10.1016/j.ecolmodel.2006.04.023>
82. Grimm V., Berger U., DeAngelis D.L., Polhill J.G., Giske J., Railsback S.F. The ODD protocol: A review and first update. *Ecological Modelling*. 2010;221:2760–2768. <https://doi.org/10.1016/j.ecolmodel.2010.08.019>
83. Grimm V., Revilla E., Berger U., Jeltsch F., Mooij W.M., Railsback S.F., Thulke H.-H., Weiner J., Wiegand T., DeAngelis D.L. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science (New York, N.Y.)*. 2005;310:987–91.
<https://doi.org/10.1126/science.1116681>
84. Grossart H.P., Riemann L., Azam F. Bacterial motility in the sea and its ecological implications. *Aquatic Microbial Ecology*. 2001;25:247–258.
<https://doi.org/10.3354/ame025247>
85. Guan W., Liang L., Boehnke M., Abecasis G.R. Genotype-based matching to correct for population stratification in large-scale case-control genetic association studies. *Genetic Epidemiology*.

- <https://doi.org/10.1002/gepi.20403>
86. Gunbin K.V., Afonnikov D.A., Kolchanov N.A., Derevianko A.P. The Importance of Changes to Microna in the Evolution of Homo Neanderthalensis and Homo Denisova. *Archaeology, Ethnology and Anthropology of Eurasia*. 2012;40:22–30.
<https://doi.org/10.1016/j.aeae.2012.11.004>
87. Gunbin K.V., Lashin S.A. The gene flow between ancient and modern humans: how often the hybridization occurred and what are population consequences?, in: Proceedings of the Moscow Conference on Computational Molecular Biology (MCCMB'13). Moscow:, p. 95.
88. Haarsma L., Nelesen S., VanAndel E., Lamine J., VandeHaar P. Simulating evolution of protein complexes through gene duplication and co-option. *Journal of Theoretical Biology*. 2016;399:22–32.
<https://doi.org/10.1016/j.jtbi.2016.03.028>
89. Haeckel E. Generelle Morphologie der Organismen. Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von C. Darwin reformirte Descendenz-Theorie. Berlin: Verlag Von Georg Reimer, 1866.
90. Haldane J.B.S. A mathematical theory of natural and artificial selection—I. *Bulletin of Mathematical Biology*. 1990;52:209–240.
<https://doi.org/10.1007/BF02459574>
91. Haldane J.B.S. A mathematical theory of natural and artificial selection. *Mathematical Proceedings of the Cambridge Philosophical Society*. 1926;23:363–372. <https://doi.org/10.1017/S0305004100015176>
92. Haldane J.B.S. A MATHEMATICAL THEORY OF NATURAL AND ARTIFICIAL SELECTION. PART II THE INFLUENCE OF PARTIAL SELF-FERTILISATION, INBREEDING, ASSORTATIVE MATING, AND SELECTIVE FERTILISATION ON THE COMPOSITION OF MENDELIAN POPULATIONS, AND ON NATURAL SELECTION. *Biological Reviews*. 1924;1:158–163. <https://doi.org/10.1111/j.1469->

- 185X.1924.tb00546.x
93. Handel A., Rozen D.E. The impact of population size on the evolution of asexual microbes on smooth versus rugged fitness landscapes. *BMC Evolutionary Biology*. 2009;9:236. <https://doi.org/10.1186/1471-2148-9-236>
94. Harris L.A., Hogg J.S., Tapia J.-J.J., Sekar J.A.P.P., Gupta S., Korsunsky I., Arora A., Barua D., Sheehan R.P., Faeder J.R. BioNetGen 2.2: advances in rule-based modeling. *Bioinformatics*. 2016;32:3366–3368. <https://doi.org/10.1093/bioinformatics/btw469>
95. Hastings J., de Matos P., Dekker A., Ennis M., Harsha B., Kale N., Muthukrishnan V., Owen G., Turner S., Williams M., Steinbeck C. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Research*. 2013;41:D456–D463. <https://doi.org/10.1093/nar/gks1146>
96. Hegselmann R., Krause U. Opinion dynamics and bounded confidence: models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*. 2002;5.
97. Hein A.M., Carrara F., Brumley D.R., Stocker R., Levin S. a. Natural search algorithms as a bridge between organisms, evolution, and ecology. *Proceedings of the National Academy of Sciences of the United States of America*. 2016;113:9413–9420. <https://doi.org/10.1073/pnas.1606195113>
98. Hellweger F.L., Clegg R.J., Clark J.R., Plugge C.M., Kreft J.-U. Advancing microbial sciences by individual-based modelling. *Nature Reviews Microbiology*. 2016;14:461–471. <https://doi.org/10.1038/nrmicro.2016.62>
99. Herring C.D., Raghunathan A., Honisch C., Patel T., Applebee M.K., Joyce A.R., Albert T.J., Blattner F.R., van den Boom D., Cantor C.R., Palsson B.Ø. Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nature Genetics*. 2006;38:1406–1412. <https://doi.org/10.1038/ng1906>
100. Hey J. Isolation with Migration Models for More Than Two Populations. *Molecular Biology and Evolution*. 2010;27:905–920.

- <https://doi.org/10.1093/molbev/msp296>
101. Hoehme S., Drasdo D. A cell-based simulation software for multi-cellular systems. *Bioinformatics*. 2010;26:2641–2642.
<https://doi.org/10.1093/bioinformatics/btq437>
 102. Hoffmann A.A., Hercus M.J. Environmental Stress as an Evolutionary Force. *BioScience*. 2000;50:217. [https://doi.org/10.1641/0006-3568\(2000\)050\[0217:ESAAEF\]2.3.CO;2](https://doi.org/10.1641/0006-3568(2000)050[0217:ESAAEF]2.3.CO;2)
 103. Holley R., Liggett T. Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*. 1975;3:643–663.
 104. Holmes E.E., Lewis M.A., Banks J.E., Veit R.R. Partial Differential Equations in Ecology: Spatial Interactions and Population Dynamics. *Ecology*. 1994;75:17–29. <https://doi.org/10.2307/1939378>
 105. Holzhütter H.-G., Drasdo D., Preusser T., Lippert J., Henney A.M. The virtual liver: a multidisciplinary, multilevel challenge for systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*. 2012;4:221–235. <https://doi.org/10.1002/wsbm.1158>
 106. Hucka M., Bergmann F.T., Hoops S., Keating S.M., Sahle S., Schaff J.C., Smith L.P., Wilkinson D.J. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core. *Journal of Integrative Bioinformatics (JIB)*. 2015;12:382–549.
<https://doi.org/10.2390/biecoll-jib-2015-266>
 107. Hucka M., Finney A., Novère N. Le. Systems biology markup language (SBML) level 2: structures and facilities for model definitions.
 108. Hucka M., Finney A., Sauro H.M., Bolouri H., Doyle J.C., Kitano H., Arkin A.P., Bornstein B.J., Bray D., Cornish-Bowden A., Cuellar A.A., Dronov S., Gilles E.D., Ginkel M., Gor V., Goryanin I.I., Hedley W.J., Hodgman T.C., Hofmeyr J.-H., Hunter P.J., Juty N.S., Kasberger J.L., Kremling A., Kummer U., Le Novère N., Loew L.M., Lucio D., Mendes P., Minch E., Mjolsness E.D., Nakayama Y., Nelson M.R., Nielsen P.F., Sakurada T., Schaff J.C., Shapiro B.E., Shimizu T.S., Spence H.D., Stelling

- J., Takahashi K., Tomita M., Wagner J., Wang J. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics (Oxford, England)*. 2003;19:524–31.
109. Hudson R.R. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*. 2002;18:337–338.
<https://doi.org/10.1093/bioinformatics/18.2.337>
110. Ilyina T.S. Bacterial superintegrons, a source of new genes with adaptive functions. *Russian Journal of Genetics*. 2006;42:1294–1302.
<https://doi.org/10.1134/S1022795406110111>
111. Ishii N., Robert M., Nakayama Y., Kanai A., Tomita M. Toward large-scale modeling of the microbial cell for computer simulation. *Journal of biotechnology*. 2004;113:281–94.
<https://doi.org/10.1016/j.jbiotec.2004.04.038>
112. Jørgensen S.E. Ecosystem ecology. London: Elsevier, 2009.
113. Jørgensen S.E., Chon T.-S., Recknagel F.A. (Eds.). Handbook of Ecological Modelling and Informatics. WIT Press, 2009.
114. Kaizu K., de Ronde W., Pajmans J., Takahashi K., Tostevin F., ten Wolde P.R. The Berg-Purcell Limit Revisited. *Biophysical Journal*. 2014;106:976–985. <https://doi.org/10.1016/j.bpj.2013.12.030>
115. Karr J.R., Sanghvi J.C., MacKlin D.N., Gutschow M. V., Jacobs J.M., Bolival B., Assad-Garcia N., Glass J.I., Covert M.W. A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell*. 2012;150:389–401. <https://doi.org/10.1016/j.cell.2012.05.044>
116. Kauffman S.A. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of theoretical biology*. 1969;22:437–67.
117. Kawashima S., Pokarowski P., Pokarowska M., Kolinski A., Katayama T., Kanehisa M. AAindex: amino acid index database, progress report 2008. *Nucleic Acids Research*. 2007;36:D202–D205.
<https://doi.org/10.1093/nar/gkm998>

118. Kazantsev F.V., Akberdin I.R., Bezmaternikh K.D., Likhoshvai V.A. The tool for automatic generation of gene networks mathematical models. *Vogis herald*. 2009;13:163–169.
119. Kazantsev F. V., Akberdin I.R., Lashin S.A., Ree N.A., Timonov V.S., Ratushnyi A. V., Khlebodarova T.M., Likhoshvai V.A., Ratushny A., Khlebodarova T.M., Likhoshvai V.A. MAMMOTH: A new database for curated mathematical models of biomolecular systems. *Journal of Bioinformatics and Computational Biology*. 2018;16:1740010 (16 pages). <https://doi.org/10.1142/S0219720017400108>
120. Khatri B.S., Free A., Allen R.J. Oscillating microbial dynamics driven by small populations, limited nutrient supply and high death rates. *Journal of theoretical biology*. 2012;314:120–9. <https://doi.org/10.1016/j.jtbi.2012.08.013>
121. Kimura M. The role of compensatory neutral mutations in molecular evolution. *Journal of Genetics*. 1985;64:7–19. <https://doi.org/10.1007/BF02923549>
122. Kinnison M.T., Unwin M.J., Hendry A.P., Quinn T.P. Migratory costs and the evolution of egg size and number in introduced and indigenous salmon populations. *Evolution*. 2001;55:1656–1667. <https://doi.org/10.1111/j.0014-3820.2001.tb00685.x>
123. Klimenko A.I., Matushkin Y.G., Kolchanov N.A., Lashin S.A. Bacteriophages affect evolution of bacterial communities in spatially distributed habitats: a simulation study. *BMC Microbiology*. 2016;16:S10. <https://doi.org/10.1186/s12866-015-0620-4>
124. Klimenko A.I., Matushkin Y.G., Kolchanov N.A., Lashin S.A. Modeling evolution of spatially distributed bacterial communities : a simulation with the haploid evolutionary constructor. *BMC Evolutionary Biology*. 2015;15:S3. <https://doi.org/10.1186/1471-2148-15-S1-S3>
125. Klimenko A.I.A.I., Matushkin Y.G.Y.G., Kolchanov N.A.N.A.N.A., Lashin S.A.S.A. Spatial heterogeneity promotes antagonistic evolutionary

- scenarios in microbial community explained by ecological stratification: a simulation study. *Ecological Modelling*. 2019;399:66–76.
<https://doi.org/10.1016/j.ecolmodel.2019.02.007>
126. Knibbe C., Fayard J.-M., Beslon G. The topology of the protein network influences the dynamics of gene order: from systems biology to a systemic understanding of evolution. *Artificial life*. 2008;14:149–56.
<https://doi.org/10.1162/artl.2008.14.1.149>
127. Knibbe C., Mazet O., Chaudier F., Fayard J.-M., Beslon G. Evolutionary coupling between the deleteriousness of gene mutations and the amount of non-coding sequences. *Journal of theoretical biology*. 2007;244:621–30. <https://doi.org/10.1016/j.jtbi.2006.09.005>
128. Komarov A., Palenova M., Smirnova O.. The concept of discrete description of plant ontogenesis and cellular automata models of plant populations. *Ecological Modelling*. 2003;170:427–439.
[https://doi.org/10.1016/S0304-3800\(03\)00243-6](https://doi.org/10.1016/S0304-3800(03)00243-6)
129. Korogodin V.I., Korogodina V.L., Fajsz C., Chepurnoy A.I., Mikhova-Tsenova N., Simonyan N. V. On the dependence of spontaneous mutation rates on the functional state of genes. *Yeast*. 1991;7:105–117.
<https://doi.org/10.1002/yea.320070204>
130. Koskella B., Brockhurst M.A. Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiology Reviews*. 2014;38:916–931. <https://doi.org/10.1111/1574-6976.12072>
131. Koskella B., Parr N. The evolution of bacterial resistance against bacteriophages in the horse chestnut phyllosphere is general across both space and time. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2015;370:20140297.
<https://doi.org/10.1098/rstb.2014.0297>
132. Krause J., Fu Q., Good J.M., Viola B., Shunkov M. V, Derevianko A.P., Pääbo S. The complete mitochondrial DNA genome of an unknown

- hominin from southern Siberia. *Nature*. 2010;464:894–897.
<https://doi.org/10.1038/nature08976>
133. Kreft J.-U., Booth G., Wimpenny J.W.T. BacSim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology*. 1998;144:3275–3287. <https://doi.org/10.1099/00221287-144-12-3275>
 134. Kreft J., Picioreanu C., Wimpenny J.W.T., Loosdrecht M.C.M. Van. Individual-based modelling of biofilms. *Microbiology*. 2001;147:2897–2912.
 135. Kuchner O., Arnold F.H. Directed evolution of enzyme catalysts. *Trends in Biotechnology*. 1997;15:523–530. [https://doi.org/10.1016/S0167-7799\(97\)01138-4](https://doi.org/10.1016/S0167-7799(97)01138-4)
 136. Kuo S.C., Koshland D.E. Multiple kinetic states for the flagellar motor switch. *Journal of Bacteriology*. 1989;171:6279–6287.
 137. Kotalik Z., Razaz M., Baranyi J. Connection between stochastic and deterministic modelling of microbial growth. *Journal of theoretical biology*. 2005;232:285–99. <https://doi.org/10.1016/j.jtbi.2004.08.013>
 138. Lander E.S. Initial impact of the sequencing of the human genome. *Nature*. 2011;470:187–197. <https://doi.org/10.1038/nature09792>
 139. Lashin S., Matushkin Y. Multi-layer computer models of gene network evolution in diploid populations, in: FEBS JOURNAL. p. 563.
 140. Lashin S.A., Klimenko A.I., Mustafin Z.S., Kolchanov N.A., Matushkin Y.G. HEC 2.0: improved simulation of the evolution of prokaryotic communities. *Mathematical Biology and Bioinformatics*. 2014;9:585–596. <https://doi.org/10.17537/2014.9.585>
 141. Lashin S.A., Matushkin Y.G. Haploid evolutionary constructor: new features and further challenges. *In silico biology*. 2012;11:125–35. <https://doi.org/10.3233/ISB-2012-0447>
 142. Lashin S.A., Matushkin Y.G., Suslov V. V., Kolchanov N.A. Computer modeling of genome complexity variation trends in prokaryotic communities under varying habitat conditions. *Ecological Modelling*.

- 2012a;224:124–129. <https://doi.org/10.1016/j.ecolmodel.2011.11.004>
143. Lashin S.A., Mustafin Z.S., Klimenko A.I., Afonnikov D.A., Matushkin Y.G. Computer Simulation of the Evolution of Microbial Population: Overcoming Local Minima When Reaching a Peak on the Fitness Landscape. *Russian Journal of Genetics*. 2020;56:242–252. <https://doi.org/10.1134/S1022795420020076>
144. Lashin S.A., Suslov V. V., Matushkin Y.G. Theories of biological evolution from the viewpoint of the modern systemic biology. *Russian Journal Of Genetics*. 2012b;48:481–496. <https://doi.org/10.1134/S1022795412030064>
145. Lashin S.A., Suslov V.V., Matushkin Y.G. Comparative modeling of coevolution in communities of unicellular organisms: adaptability and biodiversity. *Journal of Bioinformatics and Computational Biology*. 2010;08:627–643. <https://doi.org/10.1142/S0219720010004653>
146. Laspidou C.S., Rittmann B.E. Evaluating trends in biofilm density using the UMCCA model. *Water research*. 2004a;38:3362–72. <https://doi.org/10.1016/j.watres.2004.04.051>
147. Laspidou C.S., Rittmann B.E. Modeling the development of biofilm density including active bacteria, inert biomass, and extracellular polymeric substances. *Water research*. 2004b;38:3349–61. <https://doi.org/10.1016/j.watres.2004.04.037>
148. Latané B. The psychology of social impact. *American Psychologist*. 1981;36:343–356. <https://doi.org/10.1037/0003-066X.36.4.343>
149. Lavaud R., Filgueira R., Nadeau A., Steeves L., Guyondet T. A Dynamic Energy Budget model for the macroalga *Ulva lactuca*. *Ecological Modelling*. 2020;418:108922. <https://doi.org/10.1016/j.ecolmodel.2019.108922>
150. Levin B.R., Bergstrom C.T. Bacteria are different: Observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. *Proceedings of the National Academy of Sciences*.

- 2000;97:6981–6985. <https://doi.org/10.1073/pnas.97.13.6981>
151. Lewenstein M., Nowak A., Latané B. Statistical mechanics of social impact. *Physical Review A*. 1992;45:763–776.
<https://doi.org/10.1103/PhysRevA.45.763>
152. Liang L., Zöllner S., Abecasis G.R. GENOME: A rapid coalescent-based whole genome simulator. *Bioinformatics*.
<https://doi.org/10.1093/bioinformatics/btm138>
153. Liu H., Nolla H.A., Campbell L. Prochlorococcus growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. *Aquatic Microbial Ecology*. 1997;12:39–47.
<https://doi.org/10.3354/ame012039>
154. Liu W., Røder H.L., Madsen J.S., Bjarnsholt T., Sørensen S.J., Burmølle M. Interspecific Bacterial Interactions are Reflected in Multispecies Biofilm Spatial Organization. *Frontiers in Microbiology*. 2016;7. <https://doi.org/10.3389/fmicb.2016.01366>
155. Logofet D.O., Alexandrov G.A. Modelling of matter cycle in a mesotrophic bog ecosystem I. Linear analysis of carbon environs. *Ecological Modelling*. 1984;21:247–258. [https://doi.org/10.1016/0304-3800\(84\)90062-0](https://doi.org/10.1016/0304-3800(84)90062-0)
156. Lotka A.J. *Elements of Physical Biology*. Baltimore: Williams and Wilkins, 1925.
157. Lotka A.J. UNDAMPED OSCILLATIONS DERIVED FROM THE LAW OF MASS ACTION. *Journal of the American Chemical Society*. 1920;42:1595–1599. <https://doi.org/10.1021/ja01453a010>
158. Lotka A.J. Contribution to the Theory of Periodic Reactions. *The Journal of Physical Chemistry*. 1909;14:271–274.
<https://doi.org/10.1021/j150111a004>
159. Lynch M. Evolution of the mutation rate. *Trends in Genetics*. 2010;26:345–352. <https://doi.org/10.1016/j.tig.2010.05.003>
160. Machado D., Costa R.S., Rocha M., Ferreira E.C., Tidor B., Rocha I.

- Modeling formalisms in Systems Biology. *AMB Express*. 2011;1:45.
<https://doi.org/10.1186/2191-0855-1-45>
161. Mackay T.F.C. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. *Nature Reviews Genetics*. 2013;15:22–33. <https://doi.org/10.1038/nrg3627>
162. Macnab R.M., Koshland D.E. The Gradient-Sensing Mechanism in Bacterial Chemotaxis. *Proceedings of the National Academy of Sciences of the United States of America*. 1972;69:2509–2512.
<https://doi.org/10.1073/pnas.69.9.2509>
163. Mailund T., Schierup M.H., Pedersen C.N.S., Mechlenborg P.J.M., Madsen J.N., Schausser L. CoaSim: A flexible environment for simulating genetic data under coalescent models. *BMC Bioinformatics*.
<https://doi.org/10.1186/1471-2105-6-252>
164. Malik-Sheriff R.S., Glont M., Nguyen T.V.N., Tiwari K., Roberts M.G., Xavier A., Vu M.T., Men J., Maire M., Kananathan S., Fairbanks E.L., Meyer J.P., Arankalle C., Varusai T.M., Knight-Schrijver V., Li L., Dueñas-Roca C., Dass G., Keating S.M., Park Y.M., Buso N., Rodriguez N., Hucka M., Hermjakob H. BioModels—15 years of sharing computational models in life science. *Nucleic Acids Research*.
<https://doi.org/10.1093/nar/gkz1055>
165. Malthus T.R. An essay on the principle of population. London: Johnson, 1978.
166. Mantovani H.C., Lopes D.R.G., Bento C.B.P., de Oliveira M.N. Microbiomes Associated with Animals: Implications for Livestock and Animal Production, in: The Brazilian Microbiome. Cham: Springer International Publishing, 2017, pp. 41–63. https://doi.org/10.1007/978-3-319-59997-7_4
167. Marée A.F.M., Grieneisen V.A., Hogeweg P. The Cellular Potts Model and Biophysical Properties of Cells, Tissues and Morphogenesis, in: Single-Cell-Based Models in Biology and Medicine. Basel: Birkhäuser

- Basel, 2007, pp. 107–136. https://doi.org/10.1007/978-3-7643-8123-3_5
168. Markel A.L. Stress and evolution. *Russian Journal of Genetics: Applied Research*. 2008;12:206–215.
169. Markus M., Böhm D., Schmick M. Simulation of vessel morphogenesis using cellular automata. *Mathematical Biosciences*. 1999;156:191–206. [https://doi.org/10.1016/S0025-5564\(98\)10066-4](https://doi.org/10.1016/S0025-5564(98)10066-4)
170. Matushkin Y.G., Lashin S.A. Spatially distributed models of evolution in symbiotic/antagonistic prokaryotic communities. *FEBS JOURNAL*. 2013;280:563–564.
171. Mburu N., Rousseau D.P.L., Stein O.R., Lens P.N.L. Simulation of batch-operated experimental wetland mesocosms in AQUASIM biofilm reactor compartment. *Journal of environmental management*. 2014;134:100–8. <https://doi.org/10.1016/j.jenvman.2014.01.005>
172. Meer M. V., Kondrashov A.S., Artzy-Randrup Y., Kondrashov F.A. Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. *Nature*. 2010;464:279–282. <https://doi.org/10.1038/nature08691>
173. Mehinto A.C., Martyniuk C.J., Spade D.J., Denslow N.D. Applications for next-generation sequencing in fish ecotoxicogenomics. *Frontiers in Genetics*. 2012;3. <https://doi.org/10.3389/fgene.2012.00062>
174. Mieke R., Horbelt J., Baumgarten Y., Bauernhansl T. Reprint of: Basic considerations for a digital twin of biointelligent systems: Applying technical design patterns to biological systems. *CIRP Journal of Manufacturing Science and Technology*. 2021;34:133–145. <https://doi.org/10.1016/j.cirpj.2021.06.004>
175. Mironova V. V., Omelyanchuk N.A., Yosiphon G., Fadeev S.I., Kolchanov N.A., Mjolsness E., Likhoshvai V.A. A plausible mechanism for auxin patterning along the developing root. *BMC Systems Biology*. 2010;4. <https://doi.org/10.1186/1752-0509-4-98>
176. Mironova V.V., Omelyanchuk N.A., Novoselova E.S., Doroshkov A.V., Kazantsev F.V., Kochetov A.V., Kolchanov N.A., Mjolsness E.,

- Likhoshvai V.A. Combined in silico/in vivo analysis of mechanisms providing for root apical meristem self-organization and maintenance. *Annals of Botany*. 2012;110:349–360. <https://doi.org/10.1093/aob/mcs069>
177. Mitchell J.G., Kogure K. Bacterial motility: Links to the environment and a driving force for microbial physics. *FEMS Microbiology Ecology*. 2006;55:3–16. <https://doi.org/10.1111/j.1574-6941.2005.00003.x>
178. Mitchell J.G., Pearson L., Bonazinga A., Dillon S., Khouri H., Paxinos R. Long lag times and high velocities in the motility of natural assemblages of marine bacteria. *Applied and Environmental Microbiology*. 1995;61:877–882. <https://doi.org/10.3354/ame018227>
179. Monod J., Wyman J., Changeux J.-P. On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology*. 1965;12:88–118. [https://doi.org/10.1016/S0022-2836\(65\)80285-6](https://doi.org/10.1016/S0022-2836(65)80285-6)
180. Mooij W.M., Hülsmann S., Vijverberg J., Veen A., Lammens E.H.R.R. Modeling Daphnia population dynamics and demography under natural conditions. *Hydrobiologia*. 2003;491:19–34. <https://doi.org/10.1023/A:1024451512597>
181. Moore J.H., Williams S.M. (Eds.). *Epistasis, Methods in Molecular Biology*. New York, NY, NY: Springer New York, 2015. <https://doi.org/10.1007/978-1-4939-2155-3>
182. Moran P.A.P. Some remarks on animal population dynamics. *Biometrics*. 1950;6:250–258.
183. Morrow K.M., Muller E., Lesser M.P. How Does the Coral Microbiome Cause, Respond to, or Modulate the Bleaching Process? pp. 153–188. https://doi.org/10.1007/978-3-319-75393-5_7
184. Murphy J.T., Walshe R. *Understanding the Dynamics of Biological Systems*. New York, NY: Springer New York, 2011. <https://doi.org/10.1007/978-1-4419-7964-3>
185. Murray J.D. *Mathematical Biology II: Spatial Models and Biomedical Applications*, Third Edition, Third. ed. New York: Springer, 2003.

186. Mustafin Z.S., Matushkin Y.G., Lashin S.A. Haploid Evolutionary Constructor: parallelization and high performance simulations of evolution of prokaryotic communities. *Russian Journal of Genetics: Applied Research*. 2012;16:825–829.
187. Nance W.E., Kearsey M.J. Relevance of connexin deafness (DFNB1) to human evolution. *American journal of human genetics*. 2004;74:1081–7. <https://doi.org/10.1086/420979>
188. Nance W.E., Liu X.Z., Pandya A. Relation between choice of partner and high frequency of connexin-26 deafness. *Lancet*. 2000;356:500–501. [https://doi.org/10.1016/S0140-6736\(00\)02565-4](https://doi.org/10.1016/S0140-6736(00)02565-4)
189. Nesbo C.L., L'Haridon S., Stetter K.O., Doolittle W.F. Phylogenetic analyses of two “archaeal” genes in *thermotoga maritima* reveal multiple transfers between archaea and bacteria. *Molecular biology and evolution*. 2001;18:362–375.
190. Ni B., Colin R., Link H., Endres R.G., Sourjik V. Growth-rate dependent resource investment in bacterial motile behavior quantitatively follows potential benefit of chemotaxis. *Proceedings of the National Academy of Sciences*. 2020;117:595–601. <https://doi.org/10.1073/pnas.1910849117>
191. Ni B., Ghosh B., Paldy F.S., Colin R., Heimerl T., Sourjik V. Evolutionary Remodeling of Bacterial Motility Checkpoint Control. *Cell Reports*. 2017;18:866–877. <https://doi.org/10.1016/j.celrep.2016.12.088>
192. Niehus R., Mitri S., Fletcher A.G., Foster K.R. Migration and horizontal gene transfer divide microbial genomes into multiple niches. *Nature Communications*. 2015;6:8924. <https://doi.org/10.1038/ncomms9924>
193. Novoselova E.S., Mironova V.V., Omelyanchuk N.A., Kazantsev F.V., Likhoshvai V.A. Mathematical modeling of auxin transport in protoxylem and protophloem of *arabidopsis thaliana* root tips. *Journal of Bioinformatics and Computational Biology*. 2013;11. <https://doi.org/10.1142/S0219720013400106>

194. Nowak A., Szamrej J., Latané B. From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review*. 1990;97:362–376. <https://doi.org/10.1037/0033-295X.97.3.362>
195. O’Beirne G.A., Patuzzi R.B. Mathematical model of outer hair cell regulation including ion transport and cell motility. *Hearing Research*. 2007;234:29–51. <https://doi.org/10.1016/j.heares.2007.09.008>
196. O’Brien E.J., Monk J.M., Palsson B.O. Using Genome-scale Models to Predict Biological Capabilities. *Cell*. 2015;161:971–987. <https://doi.org/10.1016/j.cell.2015.05.019>
197. Oldewurtel E.R., Kouzel N., Dewenter L., Henseler K., Maier B. Differential interaction forces govern bacterial sorting in early biofilms. *eLife*. 2015;4. <https://doi.org/10.7554/eLife.10811>
198. Olson M. V., Varki A. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nature Reviews Genetics*. 2003;4:20–28. <https://doi.org/10.1038/nrg981>
199. Omholt S.W., Gjuvsland A.B., Vik J.O. The Digital Salmon - Key to Both More Science and More Profit, in: Grue J., Almås K.A. (Eds.), *Food from the Ocean - Norway’s Opportunities*. Oslo, Norway:, pp. 81–112.
200. Orth J.D., Thiele I., Palsson B.Ø. What is flux balance analysis? *Nature Biotechnology*. 2010;28:245–248. <https://doi.org/10.1038/nbt.1614>
201. Owolabi K.M., Patidar K.C. Numerical simulations of multicomponent ecological models with adaptive methods. *Theoretical Biology and Medical Modelling*. 2016;13:1. <https://doi.org/10.1186/s12976-016-0027-4>
202. Pałubicki W., Kokosza A., Burian A. Formal description of plant morphogenesis. *Journal of Experimental Botany*. 2019;70:3601–3613. <https://doi.org/10.1093/jxb/erz210>
203. Pasternak T., Groot E.P., Kazantsev F. V., Teale W., Omelyanchuk N., Kovrizhnykh V., Palme K., Mironova V. V. Salicylic Acid Affects Root Meristem Patterning via Auxin Distribution in a Concentration-Dependent

- Manner. *Plant Physiology*. 2019;180:1725–1739.
<https://doi.org/10.1104/pp.19.00130>
204. Phillips S.J., Anderson R.P., Schapire R.E. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*. 2006;190:231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
205. Posth C., Renaud G., Mittnik A., Drucker D.G., Rougier H., Cupillard C., Valentin F., Thevenet C., Furtwängler A., Wißing C., Francken M., Malina M., Bolus M., Lari M., Gigli E., Capecchi G., Crevecoeur I., Beauval C., Flas D., Germonpré M., van der Plicht J., Cottiaux R., Gély B., Ronchitelli A., Wehrberger K., Grigorescu D., Svoboda J., Semal P., Caramelli D., Bocherens H., Harvati K., Conard N.J., Haak W., Powell A., Krause J. Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Current Biology*. <https://doi.org/10.1016/j.cub.2016.01.037>
206. Posukh O.L., Bady-Khoo M.S., Zytsar M. V., Mikhalskaia V.Y., Lashin S.A., Barashkov N.A., Romanov G.P. The impact of the sociodemographic structure of deaf people communities on the prevalence of hereditary hearing loss. *Russian Journal of Genetics: Applied Research*. 2016;6:854–863. <https://doi.org/10.1134/S2079059716070108>
207. Price N.D., Reed J.L., Palsson B.Ø. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology*. 2004;2:886–897. <https://doi.org/10.1038/nrmicro1023>
208. Reich D., Patterson N., Kircher M., Delfin F., Nandineni M.R., Pugach I., Ko A.M.-S., Ko Y.-C., Jinam T.A., Phipps M.E., Saitou N., Wollstein A., Kayser M., Pääbo S., Stoneking M. Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania. *The American Journal of Human Genetics*. 2011;89:516–528.
<https://doi.org/10.1016/j.ajhg.2011.09.005>
209. Ricker W.E. Handbook of computations of biological statistics of fish populations. Ottawa: Queen’s printer and controller of stationery, 1958.

210. Rivera M.C., Lake J.A. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature*. 2004;431:152–5. <https://doi.org/10.1038/nature02848>
211. Roff D. a. The evolution of migration and some life history parameters in marine fishes. *Environmental Biology of Fishes*. 1988;22:133–146. <https://doi.org/10.1007/BF00001543>
212. Romanov G.P., Barashkov N.A., Teryutin F.M., Lashin S.A., Solovyev A.V., Pshennikova V.G., Bondar A.A., Morozov I.V., Sazonov N.N., Tomsky M.I., Dzhemileva L.U., Khusnutdinova E.K., Posukh O.L., Fedorova S.A. Marital Structure, Genetic Fitness, and the GJB2 Gene Mutations among Deaf People in Yakutia (Eastern Siberia, Russia). *Russian Journal of Genetics*. 2018;54. <https://doi.org/10.1134/S1022795418050071>
213. Romanov G.P., Smirnova A.A., Zamyatin V.I., Mukhin A.M., Kazantsev F. V., Pshennikova V.G., Teryutin F.M., Solovyev A. V., Fedorova S.A., Posukh O.L., Lashin S.A., Barashkov N.A. Agent-Based Modeling of Autosomal Recessive Deafness 1A (DFNB1A) Prevalence with Regard to Intensity of Selection Pressure in Isolated Human Population. *Biology*. 2022;11:257. <https://doi.org/10.3390/biology11020257>
214. Salem R.M., Wessel J., Schork N.J. A comprehensive literature review of haplotyping software and methods for use with unrelated individuals. *Human Genomics*. 2005;2:39. <https://doi.org/10.1186/1479-7364-2-1-39>
215. Salguero A.G., Capel M.I., Tomeu A.J. Parallel Cellular Automaton Tumor Growth Model. pp. 175–182. https://doi.org/10.1007/978-3-319-98702-6_21
216. Sarma G.P., Lee C.W., Portegys T., Ghayoomie V., Jacobs T., Alicea B., Cantarelli M., Currie M., Gerkin R.C., Gingell S., Gleeson P., Gordon R., Hasani R.M., Idili G., Khayrulin S., Lung D., Palyanov A., Watts M., Larson S.D. OpenWorm: overview and recent advances in integrative biological simulation of *Caenorhabditis elegans*. *Philosophical Transactions*

- of the Royal Society B: Biological Sciences*. 2018;373:20170382.
<https://doi.org/10.1098/rstb.2017.0382>
217. Schaffner S.F., Foo C., Gabriel S., Reich D., Daly M.J., Altshuler D. Calibrating a coalescent simulation of human genome sequence variation. *Genome Research*. 2005;15:1576–83. <https://doi.org/10.1101/gr.3709305>
218. Schiettekatte O., Vincent A.T., Malosse C., Lechat P., Chamot-Rooke J., Veyrier F.J., Picardeau M., Bourhy P. Characterization of LE3 and LE4, the only lytic phages known to infect the spirochete *Leptospira*. *Scientific Reports*. 2018;8:11781. <https://doi.org/10.1038/s41598-018-29983-6>
219. Shendure J., Porreca G.J., Reppas N.B., Lin X., McCutcheon J.P., Rosenbaum A.M., Wang M.D., Zhang K., Mitra R.D., Church G.M. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science (New York, N.Y.)*. 2005;309:1728–32.
<https://doi.org/10.1126/science.1117389>
220. Silveira C.B., Rohwer F.L. Piggyback-the-Winner in host-associated microbial communities. *npj Biofilms and Microbiomes*. 2016;2:16010.
<https://doi.org/10.1038/npjbiofilms.2016.10>
221. Slimi R., El Yacoubi S., Dumonteil E., Gourbière S. A cellular automata model for Chagas disease. *Applied Mathematical Modelling*. 2009;33:1072–1085. <https://doi.org/10.1016/j.apm.2007.12.028>
222. Sloan D.B., Alverson A.J., Wu M., Palmer J.D., Taylor D.R. Recent Acceleration of Plastid Sequence and Structural Evolution Coincides with Extreme Mitochondrial Divergence in the Angiosperm Genus *Silene*. *Genome Biology and Evolution*. 2012;4:294–306.
<https://doi.org/10.1093/gbe/evs006>
223. Soucy S.M., Huang J., Gogarten J.P. Horizontal gene transfer: building the web of life. *Nature Reviews Genetics*. 2015;16:472–482.
<https://doi.org/10.1038/nrg3962>
224. Stocker R., Seymour J.R. Ecology and physics of bacterial chemotaxis in the ocean. *Microbiology and molecular biology reviews : MMBR*.

- 2012;76:792–812. <https://doi.org/10.1128/MMBR.00029-12>
225. Sundararaj S., Guo A., Habibi-Nazhad B., Rouani M., Stothard P., Ellison M., Wishart D.S. The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of Escherichia coli. *Nucleic acids research*. 2004;32:D293-5. <https://doi.org/10.1093/nar/gkh108>
226. Sutterlin T., Kolb C., Dickhaus H., Jager D., Grabe N. Bridging the scales: semantic integration of quantitative SBML in graphical multi-cellular models and simulations with EPISIM and COPASI. *Bioinformatics*. 2013;29:223–229. <https://doi.org/10.1093/bioinformatics/bts659>
227. Sütterlin T., Tsingos E., Bensaci J., Stamatas G.N., Grabe N. A 3D self-organizing multicellular epidermis model of barrier formation and hydration with realistic cell morphology based on EPISIM. *Scientific Reports*. 2017;7:43472. <https://doi.org/10.1038/srep43472>
228. Swat M.H., Thomas G.L., Belmonte J.M., Shirinifard A., Hmeljak D., Glazier J.A. Multi-Scale Modeling of Tissues Using CompuCell3D. pp. 325–366. <https://doi.org/10.1016/B978-0-12-388403-9.00013-8>
229. Szabó A., Merks R.M.H. Cellular Potts Modeling of Tumor Growth, Tumor Invasion, and Tumor Evolution. *Frontiers in Oncology*. 2013;3. <https://doi.org/10.3389/fonc.2013.00087>
230. Szendro I. G., Franke J., de Visser J.A.G.M., Krug J. Predictability of evolution depends nonmonotonically on population size. *Proceedings of the National Academy of Sciences*. 2013;110:571–576. <https://doi.org/10.1073/pnas.1213613110>
231. Szendro Ivan G, Schenk M.F., Franke J., Krug J., de Visser J.A.G.M. Quantitative analyses of empirical fitness landscapes. *Journal of Statistical Mechanics: Theory and Experiment*. 2013;2013:P01005. <https://doi.org/10.1088/1742-5468/2013/01/P01005>
232. Szigeti B., Gleeson P., Vella M., Khayrulin S., Palyanov A., Hokanson J., Currie M., Cantarelli M., Idili G., Larson S. OpenWorm: an

- open-science approach to modeling *Caenorhabditis elegans*. *Frontiers in Computational Neuroscience*. 2014;8.
<https://doi.org/10.3389/fncom.2014.00137>
233. Sznajd-Weron K., Sznajd J. OPINION EVOLUTION IN CLOSED COMMUNITY. *International Journal of Modern Physics C*. 2000;11:1157–1165. <https://doi.org/10.1142/S0129183100000936>
234. Taft R.J., Pheasant M., Mattick J.S. The relationship between non-protein-coding DNA and eukaryotic complexity. *BioEssays*. 2007;29:288–299. <https://doi.org/10.1002/bies.20544>
235. Talaminos-Barroso A., Reina-Tosina J., Roa-Romero L.M. Models based on cellular automata for the analysis of biomedical systems, in: *Control Applications for Biomedical Engineering Systems*. Elsevier, 2020, pp. 405–445. <https://doi.org/10.1016/B978-0-12-817461-6.00014-7>
236. Tamames J. Evolution of gene order conservation in prokaryotes. *Genome Biology*. 2001;2:research0020.1. <https://doi.org/10.1186/gb-2001-2-6-research0020>
237. Tansley A.G. The Use and Abuse of Vegetational Concepts and Terms. *Ecology*. 1935;16:284–307. <https://doi.org/10.2307/1930070>
238. Tellechea-Luzardo J., Winterhalter C., Widera P., Kozyra J., de Lorenzo V., Krasnogor N. Linking Engineered Cells to Their Digital Twins: A Version Control System for Strain Engineering. *ACS Synthetic Biology*. 2020;9:536–545. <https://doi.org/10.1021/acssynbio.9b00400>
239. Tindall M.J., Porter S.L., Maini P.K., Gaglia G., Armitage J.P. Overview of Mathematical Approaches Used to Model Bacterial Chemotaxis I: The Single Cell. *Bulletin of Mathematical Biology*. 2008;70:1525–1569. <https://doi.org/10.1007/s11538-008-9321-6>
240. Tomita M. Whole-cell simulation: a grand challenge of the 21st century. *Trends in Biotechnology*. 2001;19:205–210.
[https://doi.org/10.1016/S0167-7799\(01\)01636-5](https://doi.org/10.1016/S0167-7799(01)01636-5)
241. Tomita M., Hashimoto K., Takahashi K., Shimizu T., Matsuzaki Y.,

- Miyoshi F., Saito K., Tanida S., Yugi K., Venter J., Hutchison C. E-CELL: software environment for whole-cell simulation. *Bioinformatics*. 1999;15:72–84. <https://doi.org/10.1093/bioinformatics/15.1.72>
242. Tran Q.N. Finding Long-Term Influence and Sensitivity of Genes Using Probabilistic Genetic Regulatory NetworksaaA preliminary version of this article was presented at the ACM-BCB 2015 conference in Atlanta, GA., in: *Emerging Trends in Applications and Infrastructures for Computational Biology, Bioinformatics, and Systems Biology*. Elsevier, 2016, pp. 107–120. <https://doi.org/10.1016/B978-0-12-804203-8.00008-0>
243. Trut L.N., Herbek Y.E., Trapezov O.V., Lashin S.A., Matushkin Y.G., Markel A.L., Kolchanov N.A. The animal domestication experiment as a model of the evolutionary process: A new insight into evolution under selection targeting regulatory systems, *Genetics, Evolution and Radiation: Crossing Borders, The Interdisciplinary Legacy of Nikolay W. Timofeeff-Ressovsky*. https://doi.org/10.1007/978-3-319-48838-7_37
244. Turelli M., Barton N.H., Coyne J.A. Theory and speciation. *Trends in Ecology & Evolution*. 2001;16:330–343. [https://doi.org/10.1016/S0169-5347\(01\)02177-2](https://doi.org/10.1016/S0169-5347(01)02177-2)
245. Turing A.M. The chemical theory of morphogenesis. *Phil. Trans. Roy. Soc.* 1952;13:1.
246. van Oven M., Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*. 2009;30:E386–E394. <https://doi.org/10.1002/humu.20921>
247. Verhulst J.H. Notice sur la loi que population suit dans son accroissement. *Correspondance mathématique et physique*. 1838;10:113–121.
248. Vincent A.T., Paquet V.E., Bernatchez A., Tremblay D.M., Moineau S., Charette S.J. Characterization and diversity of phages infecting *Aeromonas salmonicida* subsp. *salmonicida*. *Scientific Reports*. 2017;7:7054. <https://doi.org/10.1038/s41598-017-07401-7>

249. Vladimirov N., Løvdok L., Lebiez D., Sourjik V. Dependence of Bacterial Chemotaxis on Gradient Shape and Adaptation Rate. *PLoS Computational Biology*. 2008;4:e1000242. <https://doi.org/10.1371/journal.pcbi.1000242>
250. Vladimirov N., Sourjik V. Chemotaxis: how bacteria use memory. *Biological Chemistry*. 2009;390. <https://doi.org/10.1515/BC.2009.130>
251. Vol'terra V. Математическая теория борьбы за существование. *Uspekhi Fizicheskikh Nauk*. 1928;8:13–34. <https://doi.org/10.3367/UFNr.0008.192801b.0013>
252. Voss-Böhme A. Multi-Scale Modeling in Morphogenesis: A Critical Analysis of the Cellular Potts Model. *PLoS ONE*. 2012;7:e42852. <https://doi.org/10.1371/journal.pone.0042852>
253. Wadhams G.H., Armitage J.P. Making Sense of it All: Bacterial Chemotaxis. *Nature reviews. Molecular cell biology*. 2004;5:1024–1037. <https://doi.org/10.1038/nrm1524>
254. Waite A.J., Frankel N.W., Dufour Y.S., Johnston J.F., Long J., Emonet T. Non- genetic diversity modulates population performance. *Molecular Systems Biology*. 2016;12:895. <https://doi.org/10.15252/msb.20167044>
255. Waite A.J., Frankel N.W., Emonet T. Behavioral Variability and Phenotypic Diversity in Bacterial Chemotaxis. *Annual Review of Biophysics*. 2018;47:595–616. <https://doi.org/10.1146/annurev-biophys-062215-010954>
256. Wakeley J. The Coalescent in an Island Model of Population Subdivision with Variation among Demes. *Theoretical Population Biology*. 2001;59:133–144. <https://doi.org/10.1006/tpbi.2000.1495>
257. Wallace Ribeiro Nunes Neto, Jessica Mayara Mendes Araujo, Darlan Ferreira da Silva, Joveliane de Melo Monteiro, Wallace Borges Pachêco, Jessflan Rafael Nascimento Santos, Maria Raimunda Chagas Silva, Maria Rosa Quaresma Bomfim, Andrea de Souza Monteiro. Detection of Gram-Negative Bacteria Carrying the blaKPC-2 Gene from Mangrove Sediments.

- Journal of Pharmacy and Pharmacology*. 2019;7.
<https://doi.org/10.17265/2328-2150/2019.08.004>
258. Wanner O., Morgenroth E. Biofilm modeling with AQUASIM. *Water science and technology : a journal of the International Association on Water Pollution Research*. 2004;49:137–44.
259. Wei Y., Wang X., Liu J., Nememan I., Singh A.H., Weiss H., Levin B.R. The population dynamics of bacteria in physically structured habitats and the adaptive virtue of random motility. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108:4047–4052.
<https://doi.org/10.1073/pnas.1013499108>
260. Weissman D.B., Desai M.M., Fisher D.S., Feldman M.W. The rate at which asexual populations cross fitness valleys. *Theoretical Population Biology*. 2009;75:286–300. <https://doi.org/10.1016/j.tpb.2009.02.006>
261. Weissman D.B., Feldman M.W., Fisher D.S. The Rate of Fitness-Valley Crossing in Sexual Populations. *Genetics*. 2010;186:1389–1410.
<https://doi.org/10.1534/genetics.110.123240>
262. Whitlock M.C., Griswold C.K., Peters A.D. Compensating for the meltdown: The critical effective size of a population with deleterious and compensatory mutations. *Annales Zoologici Fennici*. 2003;40:169–183.
263. Whitlock M.C., Phillips P.C., Moore F.B.-G., Tonsor S.J. Multiple Fitness Peaks and Epistasis. *Annual Review of Ecology and Systematics*. 1995;26:601–629.
264. Widder S., Allen R.J., Pfeiffer T., Curtis T.P., Wiuf C., Sloan W.T., Cordero O.X., Brown S.P., Momeni B., Shou W., Kettle H., Flint H.J., Haas A.F., Laroche B., Kreft J.-U., Rainey P.B., Freilich S., Schuster S., Milferstedt K., van der Meer J.R., Großkopf T., Huisman J., Free A., Picioreanu C., Quince C., Klapper I., Labarthe S., Smets B.F., Wang H., Soyer O.S. Challenges in microbial ecology: building predictive understanding of community function and dynamics. *The ISME Journal*.
<https://doi.org/10.1038/ismej.2016.45>

265. Wilke C. Dynamic fitness landscapes in molecular evolution. *Physics Reports*. 2001;349:395–446. [https://doi.org/10.1016/S0370-1573\(00\)00118-6](https://doi.org/10.1016/S0370-1573(00)00118-6)
266. Wittig U., Kania R., Golebiewski M., Rey M., Shi L., Jong L., Algae E., Weidemann A., Sauer-Danzwith H., Mir S., Krebs O., Bittkowski M., Wetsch E., Rojas I., Muller W. SABIO-RK--database for biochemical reaction kinetics. *Nucleic Acids Research*. 2012;40:D790–D796. <https://doi.org/10.1093/nar/gkr1046>
267. Won Y., Young C.R., Lutz R.A., Vrijenhoek R.C. Dispersal barriers and isolation among deep-sea mussel populations (*Mytilidae*: *Bathymodiolus*) from eastern Pacific hydrothermal vents. *Molecular Ecology*. 2002;12:169–184. <https://doi.org/10.1046/j.1365-294X.2003.01726.x>
268. Wong-Ng J., Melbinger A., Celani A., Vergassola M. The Role of Adaptation in Bacterial Speed Races. *PLOS Computational Biology*. 2016;12:e1004974. <https://doi.org/10.1371/journal.pcbi.1004974>
269. Wright S. Evolution and the genetics of populations, volume 3: experimental results and evolutionary deductions. The University of Chicago Press, 1984.
270. Wright S. THE GENETICAL STRUCTURE OF POPULATIONS. *Annals of Eugenics*. 1949;15:323–354. <https://doi.org/10.1111/j.1469-1809.1949.tb02451.x>
271. Wright S. EVOLUTION IN MENDELIAN POPULATIONS. *Genetics*. 1931;16:97–159. <https://doi.org/10.1093/genetics/16.2.97>
272. Xavier J.B., Picioreanu C., van Loosdrecht M.C.M. A framework for multidimensional modelling of activity and structure of multispecies biofilms. *Environmental microbiology*. 2005;7:1085–103. <https://doi.org/10.1111/j.1462-2920.2005.00787.x>
273. Xu X., Wen X., Veltman D.M., Keizer-Gunnink I., Pots H., Kortholt A., Jin T. GPCR-controlled membrane recruitment of negative regulator

- C2GAP1 locally inhibits Ras signaling for adaptation and long-range chemotaxis. *Proceedings of the National Academy of Sciences*. 2017;114:E10092–E10101. <https://doi.org/10.1073/pnas.1703208114>
274. Yawata Y., Cordero O.X., Menolascina F., Hehemann J.H., Polz M.F., Stocker R. Competition-dispersal tradeoff ecologically differentiates recently speciated marine bacterioplankton populations. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111:5622–5627. <https://doi.org/10.1073/pnas.1318943111>
275. Yilmaz L.S., Walhout A.J. Metabolic network modeling with model organisms. *Current Opinion in Chemical Biology*. 2017;36:32–39. <https://doi.org/10.1016/j.cbpa.2016.12.025>
276. Zavarzin G. Microbial Cycles, in: Jørgensen S.E. (Ed.), *Global Ecology*. Academic Press, 2010, pp. 183–190.
277. Zöllner R., Oldewurtel E.R., Kouzel N., Maier B. Phase and antigenic variation govern competition dynamics through positioning in bacterial colonies. *Scientific Reports*. 2017;7:12151. <https://doi.org/10.1038/s41598-017-12472-7>
278. Акбердин И.Р., Казанцев Ф.В., Ермак Т.В., Тимонов В.С., Хлебодарова Т.М., Лихошвай В.А. «Электронная клетка»: проблемы и перспективы. *Математическая биология и биоинформатика*. 2013;8:287–307.
279. Александреску А. Современное проектирование на C++. Обобщённое программирование и прикладные шаблоны проектирования. Вильямс, 2019.
280. Алтухов Ю.П. Генетические процессы в популяциях. Москва: Наука, 1983.
281. Апонина Е.А., Апонин Ю.М., Базыкин А.Д. Анализ сложного динамического поведения в модели хищник–две жертвы, in: Проблемы Экологического Мониторинга и Моделирования Экосистем. т. 5. Ленинград: Гидрометеоиздат, 1982, pp. 163–180.

282. Афонников Д.А., Миронова В.В. СИСТЕМНАЯ БИОЛОГИЯ. *Вавиловский журнал генетики и селекции*. 2014;18:175–192.
283. Базыкин А.Д. Нелинейная динамика взаимодействующих популяций.
284. Байков К.С., Сергеев М.Г., Черемушкина В.А. Биоразнообразие: сущность и опыт создания онтологии. Популяционно-видовой уровень, in: Шумный В.К., Шокин Ю.И., Колчанов Н.А., Федотов А.М. (Eds.), Биоразнообразие и Динамика Экосистем. Новосибирск: Издательство СО РАН, 2006, pp. 35–43.
285. Беклемишев В.Н. Биоценологические основы сравнительной паразитологии.
286. Буч Г., Максимчук Р.А., Энгл М.У., Янг Б.Д., Коналлен Д., Хьюстон К.А. Объектно-ориентированный анализ и проектирование с примерами приложений, 3rd ed. Москва: Вильямс, 2017.
287. Быков Б.А. Экологический словарь. Алма-Ата: Наука, 1983.
288. Витвицкий А.А. Клеточные автоматы с динамической структурой для моделирования роста биологических тканей. *Сибирский журнал индустриальной математики*. 2014;17:315–327.
289. Вольтерра В. Математическая теория борьбы за существование. Москва: Наука, 1976.
290. Гамма Э., Хелм Р., Джонсон Р., Влиссидес Д. Приемы объектно-ориентированного проектирования. Паттерны проектирования. Санкт-Петербург: Питер, 2015.
291. Гиляров М.С., Баев А.А., Винберг Г.Г., Заварзин Г.А., Иванов А.В., Северин С.Е., Симолин А.В., Соколов В.Е., Татаринцев Л.П., Тахтаджян А.Л., Яблоков А.В. (Eds.). Биологический энциклопедический словарь. Москва: Советская энциклопедия, 1986.
292. Гунбин К.В., Афонников Д.А., Колчанов Н.А., Деревянко А.П. Важная роль изменений микроРНК в эволюции Homo Neanderthalensis и Homo Denisova. *Археология, этнография и антропология Евразии*.

- 2012;51:22–30.
293. Жаботинский А.М. Концентрационные автоколебания. М., “Наука,” 1974.
294. Заварзин Г.А. Будущее отбирается прошлым. *Вестник РАН*. 2004;74:813–822.
295. Заварзин Г.А. Эволюция геосферно-биосферной системы. *Природа*. 2003;1:27–35.
296. Загульнова Л.Б., Смирнова О.В., Комаров А.С., Ханина П.Г. Мониторинг фитопопуляций. *Успехи современной биологии*. 1993;113:402–414.
297. Ильина В.Л., Корогодин В.И., Файси Ч. Зависимость частоты спонтанного возникновения реверсов разных типов у ауксотрофных по аденину дрожжей от содержания аденина в среде. *Генетика*. 1987;23:637–642.
298. Иорданский Н. Эволюция жизни. Москва: Академия, 2001.
299. Казанцев Ф.В., Акбердин И.Р., Безматерных К.Д., Лихошвай В.А. Система автоматизированной генерации математических моделей генных сетей. *Вестник ВОГиС*. 2009;13:163–169.
300. Камшилов М.М. Фенотип и генотип в эволюции. *Проблемы эволюции*. 1972;2:28–44.
301. Карев Г.П., Скоморовский Ю.И. Моделирование динамики однопородных древостоев. *Сибирский экологический журнал*. 1999;4:403–417.
302. Колобов А.Н., Фрисман Е.Я. Моделирование процессов динамической самоорганизации в пространственно распределенных растительных сообществах. *Математическая биология и биоинформатика*. 2008;3:85–102.
303. Колчанов Н., Игнатьева Е.В., Подколотная О.А., Лихошвай В.А., Матушкин Ю.Г. Генные сети. *Вавиловский журнал генетики и селекции*. 2013;17:833–850.

304. Колчанов Н.А., Гончаров С.С., Лихошвай В.А., В.А. И. Системная компьютерная биология, 14th ed. Новосибирск: СО РАН, 2008.
305. Колчанов Н.А., Шиндялов И.Н. Теоретическое исследование эволюции регуляторных контуров при различных типах отбора, in: Шумный В.К., Колчанов Н.А., Рувинский А.О. (Eds.), Проблемы Генетики и Теории Эволюции. Новосибирск: Наука, 1991, pp. 268–279.
306. Кудрина К.Н., Полетаев И.А. Модели распределённых биоценозов, in: Некоторые Проблемы Мат. Биологии. Институт гидродинамики СО АН СССР, 1974.
307. Кулагина О.С., Ляпунов А.. К вопросу о моделировании эволюционного процесса. *Проблемы кибернетики: сборник статей.* 1966;16:147–169.
308. Лашин С.А. Математическое и компьютерное моделирование коэволюции сообществ одноклеточных гаплоидных организмов. ИЦиГ СО РАН, Новосибирск, 2010.
309. Лашин С.А., Колчанов Н.А., Матушкин Ю.Г. Эволюционный конструктор : методика и комплекс программ для моделирования эволюции трофически связанных популяций одноклеточных гаплоидных. *Вычислительные технологии.* 2008;13:91–101.
310. Лашин С.А., Мамонтова Е.А., Матушкин Ю.Г. Разработка пространственно распределенной модели эволюции прокариотических сообществ. *Вавиловский журнал генетики и селекции.* 2012а;16:111–118.
311. Лашин С А, Матушкин Ю.Г., Суслов В.В., Колчанов Н.А. Эволюционные тренды в системах “прокариотическое сообщество” и “прокариотическое сообщество – фаг.” *Генетика.* 2011;47:1676–1685.
312. Лашин С. А., Матушкин Ю.Г., Суслов В.В., Колчанов Н.А. Эволюционные тренды в системах “прокариотическое сообщество” и “прокариотическое сообщество – фаг.” *Генетика.* 2011;47:1676–1685.

313. Лашин С.А., Мустафин З.С., Клименко А.И., Афонников Д.А., Матушкин Ю.Г. Компьютерное моделирование эволюции микробной популяции: преодоление локальных минимумов при достижении пика на ландшафте приспособленности. *Генетика*. 2020;56:234–246.
<https://doi.org/10.31857/S0016675820020071>
314. Лашин С.А., Суслов В.В., Матушкин Ю.Г. Теории биологической эволюции с позиций современного развития системной биологии. *Генетика*. 2012b;48:573–589.
315. Лашин С.А., Суслов В.В., Матушкин Ю.Г. Моделирование эволюции трофически замкнутых сообществ с компенсаторным и некомпенсаторным метаболизмом. *Вестник ВОГиС*. 2009;13:150–159.
316. Ли Ч. Введение в популяционную генетику. Москва: Мир, 1978.
317. Лихошвай В.А. Математическое моделирование и компьютерный анализ генных сетей.
318. Лихошвай В.А., Матушкин Ю.Г., Ватолин Ю.Н., Бажан С.И. Обобщенный химико-кинетический метод моделирования сложных биологических систем. Компьютерная модель онтогенеза бактериофага Lambda. *Вычислительные технологии*. 2000;5:87–99.
319. Лихошвай В.А., Омельянчук Н.А., Миронова В.В., Фадеев С.И., Мелснесс Э.Д., Колчанов Н.А. Математическая модель распределения ауксина в корне растения. *Онтогенез*. 2007;38:446–456.
320. Лобанов М.Л. Методы определения коэффициентов диффузии : учеб. пособие. Екатеринбург: Издательство Уральского ун-та, 2017.
321. Ляпунов А.. О кибернетических вопросах биологии, in: Проблемы Кибернетики: Сборник Статей. Москва: Наука, 1972, pp. 5–39.
322. Ляпунов А.. Об управляющих системах живой природы и общем понимании жизненных процессов. Москва: Издательство Академии Наук СССР, 1962.
323. Нетрусов А.И., Котова И.Б. Микробиология: теория и практика.

- В 2 частях. Часть 1: учебник для бакалавриата и магистратуры. Москва: Издательство Юрайт, 2018.
324. Одум Ю. Основы экологии. Москва: Мир, 1975.
325. Пальянов А.Ю. Методы и алгоритмы для решения ряда актуальных задач в области вычислительной нейробиологии, биомеханики и молекулярной биологии. Новосибирск, 2019.
326. Полетаев И.А. Использование принципа Либиха в математических моделях метаболирующих систем, in: Имитационное Моделирование и Экология. Москва.:
327. Полетаев И.А. О математических моделях элементарных процессов в биоценозах. *Проблемы кибернетики: сборник статей.* 1966;171–190.
328. Полетаев И.А. О некоторых моделях биогеоценозов, in: Труды Совещания По Применению Математических Методов в Биологии.
329. Пономаренко М.П., Суслов В.В., Гунбин К.В., Пономаренко П.М., Вишневецкий О.В., Колчанов Н.А. Выявление связи вариабельности экспрессии генов путей передачи сигналов в мозге человека со сродством ТАТА-связывающего белка к промоторам этих генов. *Вавиловский журнал генетики и селекции.* 2014;18:1219–1230.
330. Ратнер В.А. Молекулярно-Генетическая Система Управления. *Природа.* 2001;3:16–22.
331. Ратнер В.А. Блочно-модульный принцип организации и эволюции молекулярно-генетических систем управления (МГСУ). *Генетика.* 1992;28:5–13.
332. Ратнер В.А. ГЕНЕТИЧЕСКИЕ УПРАВЛЯЮЩИЕ СИСТЕМЫ.
333. Ратнер В.А., Жарких А.А., Колчанов Н.А., Родин С.Н., Соловьев В.В., Шамин В.В. Проблемы теории молекулярной эволюции. Новосибирск: Наука, 1985.
334. Ратушный А.В., Лихошвай В.А., Ананько Е.А., Владимиров Н.В., Гунбин К.В., Лашин С.А., Недосекина Е.А., Николаев С.В.,

- Омельянчук Л.В., Матушкин Ю.Г., Колчанов Н.А. Новосибирская школа системной компьютерной биологии: исторический экскурс и перспективы развития. *Вестник ВОГиС*. 2005;9:232–261.
335. Ризниченко Г.Ю. Математическое моделирование биологических процессов. Модели в биофизике и экологии. М. : Издательство Юрайт, 2017.
336. Ризниченко Г.Ю. Математические модели в биофизике и экологии. Москва-Ижевск: Институт компьютерных исследований, 2003.
337. Ризниченко Г.Ю. Лекции По Математическим Моделям В Биологии.
338. Романюха А.А., Каркач А.. Индивидуально-ориентированная модель динамики инфекционного процесса в неоднородной популяции. *Математическое моделирование*. 2003;15:95–105.
339. Семовский С. В.; Букин Ю.С., Щербаков Д.Ю. Модели симпатрического видообразования в изменяющихся условиях среды. *Сибирский экологический журнал*. 2004;621–627.
340. Смирнова О.В., Загульнова Л.Б., Попадюк Р.В. Популяционная концепция в биоценологии. *Журнал общей биологии*. 1993;54:438–447.
341. Сукачёв В.Н. Избранные труды в трёх томах. Ленинград: Наука, 1972.
342. Фадеев С.И., Лихошвай В.А. О Гипотетических Генных Сетях. *Сибжсим*. 2003;6:134–153.
343. Фадеев С.И., Лихошвай В.А., Когай В.В., Омельянчук Н.А. О Математическом Моделировании Паттерна Распределения Ауксина В Корне Растений. *Сибжсим*. 2008;48:25–41.
344. Фишер Р. Генетическая теория естественного отбора. Москва-Ижевск: РХД, 2011.
345. Цхай А.А., Пулян М., Бельдеева Л.Н., Ганулис Д., Жерелина И.В., Квон В.И., Кириллов В.В., Ляхова С.А., Нахтнебел Х.-П. Введение в

- экологическое моделирование. Барнаул: Азбука, 2001.
346. Чураев Р.Н. Генные и эпигенные сети: два уровня организации наследственной системы. *Информ. вестник* 2005;9:199–208.
347. Чураев Р.Н., Ратнер В.А. Моделирование молекулярно-генетических систем управления на языке теории автоматов. Сообщ. 1. Опероны и оперонные системы, in: Исследования По Теоретической Генетике. Новосибирск: Издательство ИЦиГ СО АН, 1972, pp. 210–228.
348. Шумный В.К., Шокин Ю.И., Колчанов Н.А., Федотов А.М. (Eds.). Биоразнообразие и динамика экосистем: информационные технологии и моделирование. Новосибирск: Издательство СО РАН, 2006.
349. Эйген М., Шустер П. Гиперцикл. Принципы самоорганизации макромолекул. Москва: Мир, 1982.

Приложение 1. Язык описания моделей ГЭК

Описание синтаксиса языка описания моделей ГЭК 0D можно найти в работах (Lashin et al., 2014; Lashin, Matushkin, 2012). Для описания пространственно-распределённых сообществ были внесены некоторые изменения в синтаксис этого языка. В этом подразделе описаны синтаксические конструкции, описывающие неоднородное пространственное распределение, размерность, факторы перераспределения и другие команды.

Размерность среды задаётся следующим образом:

```
dimension = {[{0 | 1 | 2 | 3}D:] {<длина>|<длина> x <глубина> | <длина> x <глубина> x <высота>}}
```

Например, *dimension = 200x300x400* или *dimension = 2D: 300x400*, или *dimension = 1D: 300*.

Также изменились параметры субстратов: если раньше начальные концентрации задавались как *substrates_ss = c1, c2, ...* (где *c1* и *c2* – концентрации специфических субстратов S1 и S2 соответственно), то теперь можно конкретизировать субстрат отдельным блоком подобно тому, как это делается с популяцией:

```
SUBSTRATES  
{N|S}sub=<номер_субстрата>  
[sub_name = <имя_субстрата>]  
{distribution={homogeneous | <градиент> |  
<пользовательское_распределение>}  
[concentration=<концентрация>]  
[inflow_ns=<концентрация_в_протоке>]  
end {N|S}sub(<номер_субстрата>)  
  
...  
  
END SUBSTRATES
```


Атрибут *concentration* указывается только для равномерного (*homogeneous*) и распределения по линейному градиенту. Градиент задаётся следующим образом:

$$distribution = grad:(a, b, c)$$

$$concentration = c_base$$

Т.е. концентрация субстрата в ячейке $(i, j, k) = c_base * (1 + a*i + b*j + c*k)$. В одномерном и двумерном случаях лишние орты и коэффициенты при них не учитываются.

Примеры пользовательских распределений:

Ssub=1

SUB_NAME = Oxygen

distribution(, *, *) = 1e-5*

distribution(1, 1, 1) = 1e-4

distribution(1, 1, 2) = 1e-3

distribution(1, 3, 3) = 2e-3

end Ssub(1)

Ssub=2

distribution(, *, *) = 1e-5*

*distribution(2, 2, *) = 1e-4*

*distribution(1, *, 3) = 1e-5*

end Ssub(2)

Здесь «звёздочка» означает, что по некоторой координате (допустим по координате высоты) все ячейки попадают под данное правило распределения. Запись *distribution(*, *, *) = 1e-5* означает, что во всех ячейках системы будет установлена концентрация $1e-5$ М для субстрата из текущего блока. В случае конфликтов правил распределения силу имеет последнее из перечисленных, поэтому вначале стоит перечислять наиболее общие правила, а затем – специализированные, которые будут интерпретированы как исключения.

У популяций также появился параметр *distribution*, с тем же смыслом, что и у субстратов. При этом параметр *SIZE* популяции в случае пользовательского распределения имеет смысл «фоновой» численности (т.е. численности данной популяции в ячейках, о которых специально не написано в директивах *distribution*). В случае распределения по линейному градиенту параметр *SIZE* имеет то же значение, что и *c_base* для субстратов.

Проток описывается при помощи вектора коэффициентов интенсивности протока соответствующей размерности:

$$flow = \{x \mid (x, y) \mid (x, y, z)\}$$

Во всех инструкциях, изменяющих геном клеток популяции (мутация, потеря генов и горизонтальный перенос), добавился атрибут, указывающий ячейку, в которой происходит данное событие. Например,

hgt = location:(<координата_по_длине(i)>, <координата_по_глубине(j)>, <координата_по_высоте(k)>); <команда hgt из «ГЭК 2.0»>

Есть возможность изменять направление и интенсивность протока в ходе расчёта модели. Для этого используется следующая команда:

$$set_flow = \{x \mid (x, y) \mid (x, y, z)\}$$

Также можно задать циклическую подачу какого-либо субстрата:

sub_cycle = location:(<координата_по_длине(i)>, <координата_по_глубине(j)>, <координата_по_высоте(k)>); <min_субстрата>:<номер_субстрата>; concentration:<концентрация>; duration:<длительность_подачи_субстрата>; period:<количество_итераций_между_подачами>

<min_субстрата> ::= {N | S}

Обращаем внимание, что итерация, перед которой выполнена инструкция *sub_cycle* считается для неё стартовой, и субстрат начинает циклировать прямо с неё. Замечание: если инструкция *sub_cycle* была вызвана до начала расчёта, то стартовой итерацией для неё считается первая итерация (startIter=1).

Пример:

```
SUB_CYCLE = location:(1,2,2); S:1; concentration: 1e-3; duration: 200; period: 5600
```

Расширены на пространственно-распределённый случай команды, задающие и отменяющие буферизованные субстраты (т.е. субстраты, чья концентрация остаётся на каждой итерации константной в соответствующей ячейке системы):

```
BUFFER_ON = location:(<координата_по_длине(i)>, <координата_по_глубине(j)>, <координата_по_высоте(k)>); <команда BUFFER_ON из «ГЭК 2.0»>
```

```
BUFFER_OFF = location:(<координата_по_длине(i)>, <координата_по_глубине(j)>, <координата_по_высоте(k)>); <команда BUFFER_OFF из «ГЭК 2.0»>
```

Пример:

```
BUFFER_ON=location:(1,5,5);N:1:1e-1
```

```
BUFFER_OFF=location:(1,5,5);N:1:1e-1
```

Пример сценария, описанного на расширенном языке описания моделей ГЭК 3D приведен ниже (описание модели дано в комментариях, помеченных знаком % или //).

```
//Модель конкуренции «оседлой» и «кочевой» популяции в среде обитания с периодическим во времени и пространстве субстратом  
DECLARE  
//размерность системы - решётка 5x5x5 ячеек
```

```

dimension = 3D:5x5x5
volume = 7.5e-1
//проток отсутствует
flow = (0.0, 0.0, 0.0)
//скорость диффузии
DIFFUSION_RATE = 0.001

//начальные данные по субстратам - присутствует только равномерно распределённый
неспецифический субстрат
SUBSTRATES
Nsub=1
    distribution = homogeneous
    concentration = 1e-5
    inflow_ns = 0e-2
end Nsub(1)

Ssub=1
    distribution = homogeneous
    concentration = 0e-5
end Ssub(1)

Ssub=2
    distribution = homogeneous
    concentration = 0e-5
end Ssub(2)

END SUBSTRATES

//Задаём параметры "оседлой" популяции, менее склонной к миграции, но лучше
утилизирующей неспецифический субстрат
POP=1
    POP_NAME=SETTLER
    SIZE=1e+9
    INCREASER= liebig
    K_DEATH=0E-8
    K_FLOW=1
    //интенсивность активной миграции, обусловленной хемотаксисом. Значение
соответствует доли субпопуляции мигрантов
    MOTILITY=0.01

```

```

CPROD=5E+4

CCONS_NSP=1E+7

CCONS_SP=1e+5

// genome

GENE_N=1;allele_values: 10; allele_concentrations: 1

GENE_P=1;allele_values: 3; allele_concentrations: 1

GENE_CH=1;allele_values: 0.5; allele_concentrations: 1

distribution = homogeneous

END POP(1)

//Задаём параметры "оседлой" популяции, хуже утилизирующей неспецифический субстрат,
но более склонной к миграции

POP=2

POP_NAME=NOMAD

SIZE=1e+9

INCREASER= liebig

K_DEATH=0E-8

K_FLOW=1

//интенсивность активной миграции, обусловленной хемотаксисом. Значение
соответствует доли субпопуляции мигрантов

MOTILITY=0.0105

CPROD=5E+4

CCONS_NSP=1E+7

CCONS_SP=1e+5

// genome

GENE_N=1;allele_values: 1; allele_concentrations: 1

GENE_P=2;allele_values: 3; allele_concentrations: 1

GENE_CH=1;allele_values: 0.5; allele_concentrations: 1

distribution = homogeneous

END POP(2)

END DECLARE

%Задаём циклическую подкормку субстратом N1, поступающим в различные ячейки системы в
течение 200 итераций с периодом 5600 итераций

%передняя грань

SUB_CYCLE = location:(1,1,1); N:1; concentration: 1e-3; duration: 200; period: 5600

iterate = 200

SUB_CYCLE = location:(1,2,2); N:1; concentration: 1e-3; duration: 200; period: 5600

```

```
%% (5 + 5*4 + 3) * 200 = 5600 - полный цикл %%  
iterate = 19400
```

Рис. П1.1. Пример скрипта (описание в комментариях, отмеченных знаком %
или //).

Приложение 2. Система плагинов ДЭК

Для классов стратегий в ДЭК была реализована возможность сборки в виде подключаемых модулей (плагинов, plug-in). Для реализации стратегии в виде подключаемого модуля, необходимо:

- Создать подкласс нужного класса стратегии (см. раздел 3.2);
- Реализовать необходимые виртуальные методы (например, `breed` для `PopulationBreedingStrategy`; `buildRecombinantGenotype` для `ChromosomeRearrangementStrategy`);
- Собрать проект в виде динамической библиотеки (*.dll для Microsoft Windows, *.so для Linux), добавив в него две функции – фабрику объектов плагина `Factory()` и функцию, возвращающую название плагина `PluginName()` (ниже приведён пример для Windows, для Linux аналогично):

```
extern "C" __declspec(dllexport) PopulationBreedingStrategy*
PopulationBreedingStrategy Factory()
{
    return new MyPopulationBreedingStrategy ();
}

extern "C" __declspec(dllexport) std::string PluginName()
{
    return "MySimplePopulationBreedingStrategy";
}
```

- Поместить собранную библиотеку (файл *.dll/*.so) в каталог `plugins` в рабочем каталоге программы ДЭК.

Детальное описание процесса создания плагинов для ДЭК, вместе с шаблонами проектов для всех типов стратегии (под Windows и Linux) будет размещено на сайте проекта, который в настоящее время разрабатывается.

Приложение 3. Язык описания моделей ДЭК

Разработан прототип языка описания моделей эволюционно-популяционных процессов, разработан и программно реализован транслятор этого языка. За основу был взят Си-подобный язык описания популяционно-генетических моделей, описанный в работе (Lashin, Matushkin, 2012). Описание модели состоит из двух частей: блока декларации модели и блока расчёта модели.

В блоке декларации задаётся начальное состояние модели – генетические (тип генотипа: числа, элементы перечислимого множества или строки; значения отдельных генов и т.д.), популяционные (размер и структура популяции, темпы размножения, смертности и т.д.), и другие параметры (свойства среды обитания и т.д.). На рисунке 4 показан фрагмент инициализации двух вариантов модельного гена митохондриальной ДНК человека. В качестве гена `initMtDNA` были взяты реальные образцы митохондриального генома человека с сайта проекта Генбанк (<http://www.ncbi.nlm.nih.gov/genbank/>), тогда как мутантная форма `initMtDNA_G` – искусственно получена путём замен нуклеотидов в четырёх позициях (рис. 4).

```
// Митохондриальная ДНК (популяция F)
GeneString initMtDNA = "TTCTTTCATGGGGAAGCAGATTGGGTACCACCCCAAGTATGACTCACCCATCAACAACC";
initMtDNA += "GCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACT";
initMtDNA += "TGACCACSTGTAGTACATAAAAAACCCCAATCCACATCAAAAACCCCCSTCATGCTTACAA";
initMtDNA += "GCAAGTACAGCAATCAACCTTCAASTATCACACATCAASTGCAACTCCAAAAGCCACCCST";
initMtDNA += "CACCCASTAGGATATCAACAACSTACCCATCSTTAAACAGTACATGGTACATAAAGCCAT";
initMtDNA += "TTACCGTACATAGCACATTACAGTCAAATCCSTTCTCGTCCCATGGATGACCCCCSTCA";

// Мутантная мтДНК (популяция G)
GeneString initMtDNA_G = initMtDNA;
initMtDNA_G.replace(60,1,"T");
initMtDNA_G.replace(120,1,"A");
initMtDNA_G.replace(180,1,"C");
initMtDNA_G.replace(240,1,"G");
```

Рис. 4. Пример декларации гена в ДЭК.

В блоке расчёта приводится описание сценария моделирования. Все события можно разделить на два класса – преддетерминированные и стохастические. Основные события – это мутации, рекомбинации, миграции, а также изменения условий окружающей среды, приводящие к

изменению популяционных параметров (рождаемость, смертность и т.д.).
Пример модельного сценария для конкретной биологической модели
приведён в приложении ниже.

Приложение 4. Сценарий моделирования популяционно-генетических и миграционных процессов в популяциях древних людей на территории Евразии

Общая схема сценария (упрощённая относительно схемы на рис. 3.3.4 основного текста диссертации) выглядела следующим образом:

```
// Митохондриальная ДНК (популяция f)
string initMtDNA = "ТТСТТТТКАТГГГГААГСАГАТТТГГГТАССАССААГАТТГАСТСАССАТСААСААСС";
initMtDNA += "ГСТАТГАТТТТГТАСАТТАСТГССАГССАССАТГААТАТТГАТГТАССАТАААТАСТ";
initMtDNA += "ТГАССАСТГТАГТАСАТАААААССААТССАСАТСААААССССССАТГСТТАСАА";
initMtDNA += "ГСААГТАСАГААТСААСТТТААСТАТСАСАТСААСТГСААСТССАААГССАСССТ";
initMtDNA += "САССАСТАГАТАТСААСААСТТАССАТСТТААСАГАТАТГТАСАТАААГССАТ";
initMtDNA += "ТТАССГТАСАТАГАСАТТАСАГТАААТСССТТТГТАССАТГГАТГАСССССТСА";

// Мутантная мтДНК (популяция G)
string initMtDNA_G = initMtDNA;
initMtDNA_G.replace(60,1,"Т");
initMtDNA_G.replace(120,1,"А");
initMtDNA_G.replace(180,1,"С");
initMtDNA_G.replace(240,1,"Г");

// Генерация стартовых популяций
basePopSize = 5000;
initBirthRate = 0.25f;
deltaBirthRate = 0.0002f;
BisexualPopulation pop_f = populationFactory(basePopSize, initMtDNA);
NeutralEvolutionBreedingStrategy breedStrat_f;
breedStrat_f.setBirthRate(initBirthRate);
breedStrat_f.setDeathRate(initBirthRate-deltaBirthRate);
pop_f.setBreedingStrategy(breedStrat_f);

BisexualPopulation pop_G = populationFactory(basePopSize, initMtDNA_G);
NeutralEvolutionBreedingStrategy breedStrat_G;
breedStrat_G.setBirthRate(initBirthRate);
breedStrat_G.setDeathRate(initBirthRate-deltaBirthRate);
pop_G.setBreedingStrategy(breedStrat_G);
// Эволюционный процесс
// Шаг 1.
genPerMigr_f_G = 100;
part_f_to_G = 0.01f;
part_G_to_f = 0.01f;
SetMigration(pop_f, pop_G, genPerMigr_f_G, part_f_to_G);
SetMigration(pop_G, pop_f, genPerMigr_f_G, part_G_to_f);
Evolution(pop_f, 7000);
Evolution(pop_G, 7000);

// Шаг 2.
// Вычленение популяции e и её жизнь (10 поколений) f нас больше не интересует
fraction_e_founder = 0.5f;
BisexualPopulation pop_e = pop_f.Subpopulation(fraction_e_founder);
Evolution(pop_e, 10);
Evolution(pop_G, 10);

// Шаг 3.
// Разделение популяции e на F и d
// Далее эволюция 50 поколений
// мажорная ветка - это популяция d, ей остаётся 90% популяции
fraction_F_founder = 0.1f;
BisexualPopulation pop_F = pop_e.Subpopulation(fraction_F_founder);
BisexualPopulation pop_d = pop_e; // Остаток становится d
```

```

pop_d.setDeathRate(initBirthRate-deltaBirthRate*2);
Evolution(pop_d, 50);
Evolution(pop_F, 50);
Evolution(pop_G, 50);

// Шаг 4.
// Разделение популяции d на E и c
// Далее эволюция 100 поколений
// мажорная ветка - это популяция c, ей остаётся 90% популяции
fraction_E_founder = 0.1f;
BisexualPopulation pop_E = pop_d.Subpopulation(fraction_E_founder);
BisexualPopulation pop_c = pop_d; // Остаток становится c
pop_c.setDeathRate(initBirthRate-deltaBirthRate*2.5f);
Evolution(pop_c, 100);
Evolution(pop_E, 100);
Evolution(pop_F, 100);
Evolution(pop_G, 100);

// Шаг 5.
// Разделение популяции c на D и b
// Далее эволюция 100 поколений
// мажорная ветка - это популяция b, ей остаётся 90% популяции
fraction_D_founder = 0.1f;
BisexualPopulation pop_D = pop_c.Subpopulation(fraction_D_founder);
BisexualPopulation pop_b = pop_c; // Остаток становится b
pop_b.setDeathRate(initBirthRate-deltaBirthRate*3f);
Evolution(pop_b, 100);
Evolution(pop_D, 100);
Evolution(pop_E, 100);
Evolution(pop_F, 100);
Evolution(pop_G, 100);

// Шаг 6.
// Разделение популяции b на C и a
// Далее эволюция 200 поколений
// мажорная ветка - это популяция a, ей остаётся 90% популяции
fraction_C_founder = 0.1f;
BisexualPopulation pop_C = pop_b.Subpopulation(fraction_C_founder);
BisexualPopulation pop_a = pop_b; // Остаток становится b
pop_a.setDeathRate(initBirthRate-deltaBirthRate*3.5f);
Evolution(pop_a, 100);
Evolution(pop_C, 100);
Evolution(pop_D, 100);
Evolution(pop_E, 100);
Evolution(pop_F, 100);
Evolution(pop_G, 100);

// Шаг 7.
// Разделение популяции a на A и B
// Далее эволюция 2540 поколений
// полумажорная ветка - это популяция A, ей остаётся 50% популяции,
// но чуть меньшая смертность.
// На протяжении эволюции миграции между A и B, а также из A в C, D, E, F (G выводится
из эволюции)
genPerMigr_A_B = 100;
genPerMigr_A_C = 150;
genPerMigr_A_D = 200;
genPerMigr_A_E = 250;
genPerMigr_A_F = 300;
part_A_to_B = 0.01f;
part_B_to_A = 0.01f;
part_A_to_C = 0.005f;
part_A_to_D = 0.005f;
part_A_to_E = 0.005f;
part_A_to_F = 0.005f;
SetMigration(pop_A, pop_B, genPerMigr_A_B, part_A_to_B);
SetMigration(pop_B, pop_A, genPerMigr_A_B, part_B_to_A);
SetMigration(pop_A, pop_C, genPerMigr_A_C, part_A_to_C);
SetMigration(pop_C, pop_A, genPerMigr_A_C, part_A_to_C);
SetMigration(pop_A, pop_D, genPerMigr_A_D, part_A_to_D);

```

```

SetMigration(pop_D, pop_A, genPerMigr_A_D, part_B_to_D);
SetMigration(pop_A, pop_E, genPerMigr_A_E, part_A_to_E);
SetMigration(pop_E, pop_A, genPerMigr_A_E, part_A_to_E);
SetMigration(pop_A, pop_F, genPerMigr_A_F, part_A_to_F);
SetMigration(pop_F, pop_A, genPerMigr_A_F, part_A_to_F);

fraction_B_founder = 0.5f;
BisexualPopulation pop_B = pop_a.Subpopulation(fraction_B_founder);
BisexualPopulation pop_A = pop_a; // Остаток становится А
pop_A.setDeathRate(initBirthRate-deltaBirthRate*4f);
Evolution(pop_A, 10000);
Evolution(pop_B, 10000);
Evolution(pop_C, 10000);
Evolution(pop_D, 10000);
Evolution(pop_E, 10000);
Evolution(pop_F, 10000);

// Конец эволюции - вывод статистики

```

Приложение 5. Общая архитектура программного комплекса PGD-S

Вычислительное ядро программного комплекса PGD-S построено вокруг паттерна проектирования «Стратегия». Стратегией представлен как главный цикл моделирования, так и его составляющие части: стратегия выбора партнера, рождения детей и отбора особей. Также через стратегию реализуется экспрессия генов, ответственных за проявление глухоты.

Этот подход позволяет менять сценарий моделирования без внесения изменений в архитектуру в целом. Замена одного этапа сводится к замене стратегии его реализации.

Первая из стратегий – это стратегия вычисления фенотипа (СВФ), определяющая значение фенотипа агента по его генотипу. Реализована классом `GenotypeToPhenotypeStrategy`. Стратегия задается при инициализации объекта агента, что позволяет менять механизм фенотипического проявления (глухоты) (рецессивный тип наследования патологии – биаллельные рецессивные мутации гена / доминантный тип наследования патологии – одна доминантная мутация) генотипа без внесения изменений в общую архитектуру модели. СВФ принимает на вход генотип, и выдает результат «да/нет», являющийся ответом на вопрос «глухой ли агент?».

Вторая стратегия – стратегия выбора фертильных агентов (СВФА). Реализована классом `FilterReproductibleStrategy`. Принимая на вход множество агентов, она выбирает подмножество агентов, находящихся в детородном возрасте (верхняя и нижняя границы возраста задаются в настройках).

Третья стратегия – стратегия рождения детей (СРД). Реализована классом `BirthStrategy`. Ей на вход подаются два агента (родители), и стратегия создает и инициализирует новый объект агента на основе информации о родителях (возраст, пол, генотип и т.д.).

Четвертая стратегия – стратегия отбора особей (СОО). Представлена классом `DeathStrategy`. Она реализует фактор смертности в популяции. СОО имеет три параметра: среднюю продолжительность жизни, распределение смертности и его дисперсию. Далее для каждого агента, на основе его возраста, высчитывается вероятность смерти на данной итерации. Также СОО реализует фактор случайной смертности: каждый агент может с некоторой заданной вероятностью умереть независимо от своего возраста

Стратегией можно назвать также и объект основного цикла моделирования (класс `Starter`), так как при выделении интерфейса он может быть легко заменен, тем самым меняя ход процесса моделирования. Его параметры можно разделить на две категории: параметры процесса моделирования и инициализационные параметры. К первым относится количество итераций моделирования. Ко вторым – доля мутантного аллеля в популяции и размер этой самой популяции.

Схема взаимоотношения классов PGD-S приведена на рисунке П.5.1

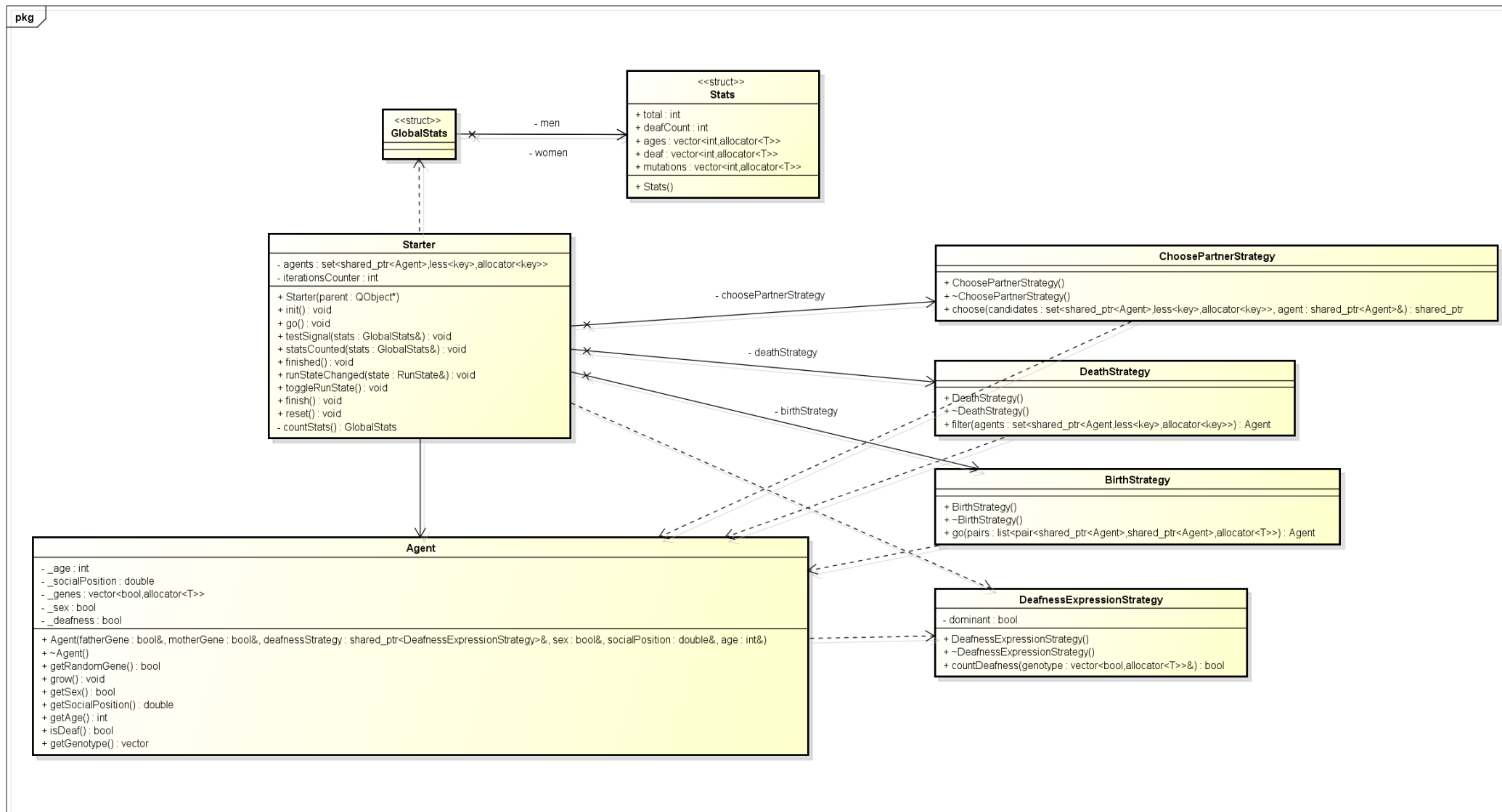


Рисунок П.5.1 Архитектура вычислительного ядра программного комплекса PGD-S.