

На правах рукописи

Лашин Сергей Александрович

**МНОГОУРОВНЕВЫЕ КОМПЬЮТЕРНЫЕ МОДЕЛИ
ЭВОЛЮЦИОННЫХ ПРОЦЕССОВ**

1.5.8 – математическая биология, биоинформатика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

доктора биологических наук

Новосибирск – 2024

Работа выполнена в секторе биоинформатики и информационных технологий ФГБНУ “Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук” (ИЦиГ СО РАН), г. Новосибирск, Россия.

Научный консультант: **Колчанов Николай Александрович**, доктор биологических наук, профессор, академик РАН, заведующий Отделом системной биологии, ИЦиГ СО РАН, научный руководитель Института (г. Новосибирск)

Официальные оппоненты:

1. **Макеев Всеволод Юрьевич**, доктор физико-математических наук, профессор, член-корреспондент РАН, заведующий лабораторией системной биологии и вычислительной генетики ФГБУН «Институт общей генетики им. Н.И. Вавилова Российской академии наук», г. Москва.
2. **Фрисман Ефим Яковлевич**, доктор биологических наук, профессор, член-корреспондент РАН, ФГБУН Институт комплексного анализа региональных проблем Дальневосточного отделения Российской академии наук, г. Биробиджан.
3. **Щербаков Дмитрий Юрьевич**, доктор биологических наук, главный научный сотрудник, заведующий лабораторией геносистематики ФГБУН Лимнологический институт Сибирского отделения Российской академии наук (г. Иркутск)

Ведущая организация:

Федеральное бюджетное учреждение науки «Научно-исследовательский институт системной биологии и медицины» Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека (ФБУН НИИ СБМ Роспотребнадзора), г. Москва.

Защита диссертации состоится «__» _____ 2024 г. на утреннем заседании диссертационного совета 24.1.239.01 по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук в ИЦиГ СО РАН по адресу: 630090 г. Новосибирск, пр. ак. Лаврентьева, 10. Тел/факс: (383)3634906; Факс: (383)3331278, e-mail: dissov@bionet.nsc.ru

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте Института: www.icgbio.ru

Автореферат разослан «__» _____ 2024 г.

Ученый секретарь
диссертационного совета
доктор биологических наук

Т.М. Хлебодарова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования. Биологические системы – это сложные иерархические системы, функционирующие на разных уровнях биологической организации – от молекулярно-генетического до экосистемного и биоценотического.

Ключевым свойством биологических систем является их способность к эволюции, то есть постепенному 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 популяционной генетики, эволюции и видообразования, учитывающих влияние условий окружающей среды, трофических взаимодействий, географических факторов, генетической структуры организмов и пр. является одной из основных задач для биологии, математики и информатики XXI века.

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 и др., 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 и др., 2014; Yilmaz, Walhout, 2017].

Расцвет гибридных подходов моделирования в биологии, обусловленный, с одной стороны, существенным ростом вычислительной мощности компьютеров, а, с другой, развитием методов объектно-ориентированного проектирования, вкупе с появлением соответствующих 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) или агентно-ориентированного (англ. agent-based) моделирования [Abar и др., 2017; Grimm и др., 2006].

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

Однако, несмотря на разнообразие предложенных моделей и подходов, актуальной задачей остаётся корректное и всестороннее описание таких систем с помощью математического и компьютерного моделирования на основе разработки новых методов, которые могли бы связать в единой модели данные, относящиеся если не ко всем, то хотя бы к нескольким уровням биологической организации. В рамках таких новых методов представляет научный интерес не только обеспечение соответствия новых моделей принципам, сформулированным Уолкером Гриммом и Утой Бергер, но и расширение спектра областей их применения в биологии, в частности, с акцентом на молекулярно-генетическом уровне организации биологических систем.

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:

1. 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.
2. Реализовать разработанные методы в виде универсальных программных комплексов, в том числе – высокопроизводительных версий, работающих на суперкомпьютерах.
3. Провести численные эксперименты по изучению эволюции микробных сообществ и популяций человека на основе разработанных программных комплексов.

Научная новизна. Впервые на момент опубликования были представлены методы моделирования функционирования и эволюции комплексных биологических си-

стем, 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 и др., 2015; Klimenko и др., 2019; Lashin, Matushkin, 2012; Mustafin, Matushkin, Lashin, 2012] и «Диплоидный Эволюционный Конструктор» (ДЭК) [Lashin, Matushkin, 2013]. Ключевой особенностью предложенных методов является то, что конструируемые модели обладают многослойной (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 и др., 2015; Klimenko и др., 2016; Klimenko и др., 2019; Lashin и др., 2012; Lashin и др., 2020; Лашин, Мамонтова, Матушкин, 2012];
- влияния социальных факторов на распространение генетической глухоты в популяциях человека [Posukh и др., 2016; Romanov и др., 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 и др., 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 и др., 2018; Vincent и др., 2017]; функционированию микробиомов коралловых экосистем [Morrow, Muller, Lesser, 2018]; микробиоте кишечника человека [Ghattargi и др., 2018], рубца жвачных животных Бразилии [Mantovani и др., 2017], отложений мангровых лесов [Wallace Ribeiro Nunes Neto и др., 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 работы. Основные результаты диссертации докладывались и обсуждались на 36-ти всероссийских и международных научных конференциях: «Актуальные проблемы вычислительной и прикладной математики» (2015, 2019, Новосибирск); «The International Society for Ecological Modelling Global Conference» (2019, Вена, Австрия); 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); 1st, 2nd, 3rd International Symposium Mathematical Modeling and High-Performance Computing in Bioinformatics, Biomedicine and Biotechnology (MM-HPC-BBB) (2014, 2016, 2018, Новосибирск); «The European Human Genetics Conference in conjunction with the European Meeting on Psychosocial Aspects of Genetics» (2018, Милан); «Генетика микроорганизмов: от геномики к биоэкономике» (2018, Пущино); V, VI, VII Международная конференция «Математическая биология и биоинформатика» (2014, 2016, 2018, Пущино); IV, V Международная научно-практическая конференция «Постгеномные методы анализа в биологии, лабораторной и клинической медицине» (2014, 2018, Казань); «Математика в современном мире. Международная конференция, посвящённая 60-летию Института математики им. С.Л. Соболева» (2017, Новосибирск); «Беляевские чтения,

Международная конференция, посвященная 100-летию со дня рождения академика АН СССР Д.К. Беляева» (2017, Новосибирск); Международный конгресс «Биотехнология: состояние и перспективы развития» (2011, 2017, Москва); «Systems Biology and Bioinformatics (SBBI'2016)» (2016, Санкт-Петербург); 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); VIII Сибирский форум индустрии информационных систем (SIIS 2016) (2016, Новосибирск); FAIRDOM Barcelona user meeting (2016, Барселона, Испания); EMBO | EMBL Symposium “New Approaches and Concepts in Microbiology” (2015, Гейдельберг, Германия); Седьмая международная молодежная научная школа-конференция «Теория и численные методы решения обратных и некорректных задач» (2015, Новосибирск); VI и VII съезды Вавиловского общества генетиков и селекционеров (2014, 2019, Ростов-на-Дону и Санкт-Петербург, соответственно); Federation of European biochemical societies congress (2013, Санкт-Петербург); The International Moscow Conference On Computational Molecular Biology (MCCMB'11) (2011, Москва); III International Conference “Biosphere Origin and Evolution” (2011, Ретимно, Греция); ERASysBio Summer School Data Management for Systems Biology III (2010, Тенерифе, Испания).

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.

Публикации

По материалам диссертационной работы опубликовано 106 научных работ (включая тезисы), в том числе 27 статей в журналах из списка изданий, входящих в международные базы цитирования (Web of Science и Scopus), 4 главы в монографиях, 5 авторских свидетельств. Все работы опубликованы в соавторстве.

Структура и объем 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 в проведении исследования

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: научному консультанту академику Н.А. Колчанову; к.б.н. Ю. Матушкину, д.б.н. В.А. Лихошваю; д.б.н. Д.А. Афонникову и Д.А. Рассказову; к.б.н. О.Л. Посух, к.б.н. Н.А. Барашкову и к.б.н. Г.П. Романову; к.б.н. Ф.В. Казанцеву, к.б.н. А.И. Клименко, к.б.н. З.С. Мустафину, Т.Н. Лаховой.

СОДЕРЖАНИЕ РАБОТЫ

Во **Введении** обоснована актуальность диссертационной работы, сформулированы её цели и задачи, представлены методы исследований, изложены научная новизна полученных результатов, их практическая значимость, личный вклад автора.

Глава 1 «Обзор литературы: о проблемах и перспективах моделирования эволюционно-биологических процессов» посвящена описанию основных иерархических уровней биологической организации (молекулярно-генетический, клеточный, тканевый/органный, организменный, популяционный, экологический/биоценотический). В главе также приводится обзор математических и компьютерных моделей, используемых для исследований этих уровней. 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 [Handbook of Ecological Modelling and Informatics, 2009] и эволюционных [Ferrer, Prats, López, 2008; Jong de, 2002] 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.

В заключение **Главы 1** показывается, что 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.

Глава 2 «Моделирование микробных сообществ» посвящена описанию результатов 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).

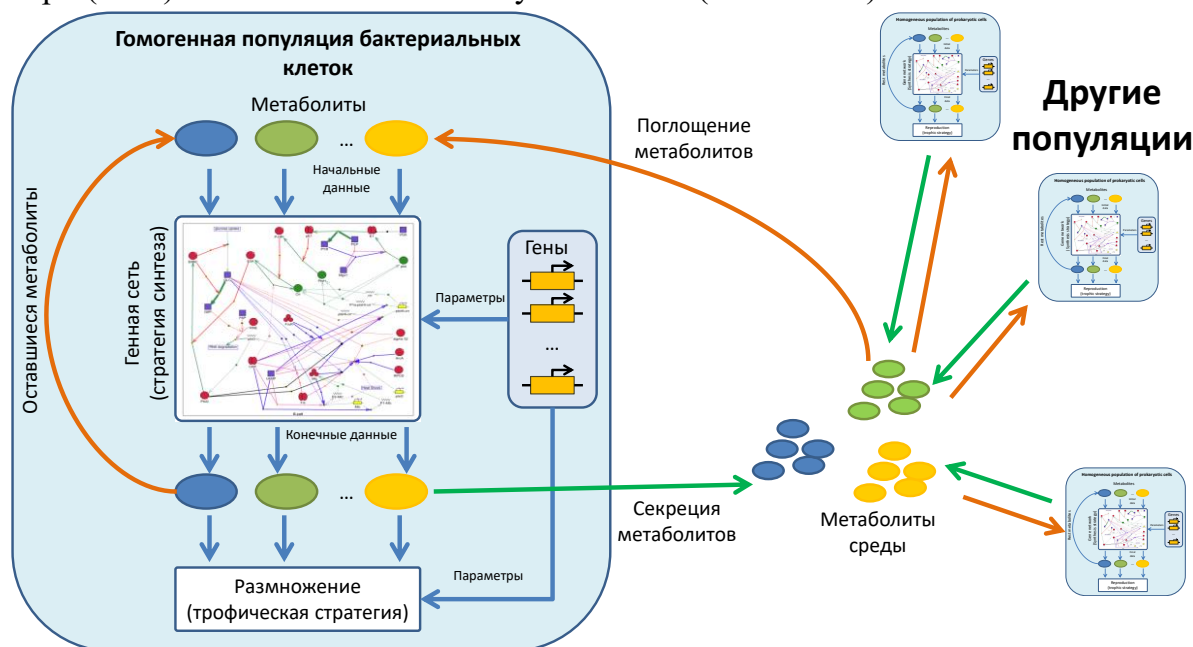


Рисунок 1. Основные объекты и процессы ГЭК.

В разделе 2.1 изложены теоретические основы методики моделирования микробных сообществ, основанной на концепции супериндивида. Ключевой единицей моделирования является популяция метаболически однородных клеток, взаимодействующая с окружающей средой и с другими популяциями (рис. 1). В рамках данной методики рассматриваются следующие объекты: *Среда* – жидкая проточная среда с равномерно распределёнными клетками и веществами; *субстраты* – вещества, необходимое клеткам для их жизнедеятельности, которые могут оказывать как ускоряющее, так и замедляющее (вплоть до отрицательного) воздействие на рост популяций; *Популяция* – характеризуется набором метаболических систем (модельных *генов*), отвечающих за то, какие субстраты и с какой скоростью клетки данной популяции способны потреблять и производить. Процессы: *транспорт веществ* – поглощение субстратов из среды и секреция продуктов клетками популяций; *самовоспроизведение*

популяций и естественный отбор; *мутации* – изменения 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 генов* – клетки различных популяций могут обмениваться между собой генами, образуя новые популяции. В рамках разработанного 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 множество неспецифических и специфических субстратов; (2) продуцируют одинаковое множество 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.

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. 1), некомпенсаторная (ур. 2) и ингибиторная (ур. 3) трофические стратегии.

$$F(n_0, \vec{S}, r_0, \vec{C}, P, G) = \sqrt{r_0 n_0(P) \sum_{i \in I_{consumed}} c_i s_i(P) - k_{death} P^2} \quad (1)$$

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

$$F(n_0, \vec{S}, r_0, \vec{C}, P, G) = a_{basal}(n_0)P - \sqrt{\sum_{i \in I_{consumed}} c_i s_i(P) - k_{death} P^2} \quad (3)$$

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.

В разделе 2.2 описана версия ГЭК для моделирования пространственно-распределённых сообществ, расширяющая методику, изложенную в 2.1. В разделе подробно описывается моделирование процессов (1) протока; (2) диффузии; (3) хемотаксиса. Кроме того, обсуждаются описание в модели факторов пространственного перераспределения и энергетических резервов клеток. В разделе 2.3 приведены подходы к моделированию питания и синтеза питательных веществ. Эти процессы описываются в методике ГЭК опираясь на следующие предположения: (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.

Раздел 2.4 содержит описание разработанных численных методов моделирования динамики генетического разнообразия в микробных популяциях, реализованных в ГЭК. Разработанные методы базируются на двух подходах. Первый основан использовании т.н. «генетических спектров популяции» - многомерных распределений частот аллелей генов и соответствует принципам моделирования микробной популяции в виде «супериндивида». Второй метод основан на непосредственном хранении всех аллельных комбинаций, присутствующих в популяции и идейно близок агентно-ориентированному подходу. По результатам сравнения двух подходов, были получены следующие рекомендации по оптимальному моделированию генетического разнообразия в ГЭК: (1) Для расчета моделей с большим количеством уникальных аллельных комбинаций ($>10^8$) лучше всего использовать модель с хранением генетического спектра популяции, так как данная модель не потребует больших затрат 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. Отметим, что разработанные методы допускают эффективную реализацию на суперкомпьютерах с поддержкой MPI и OpenMP, которая была описана в подразделе 2.6.3.

Раздел 2.5 посвящён описанию моделирования фаговой инфекции в ГЭК. Моделирование инфекционного процесса включает следующие фазы: (1) заражение здоровых клеток путем проникновения фагов из окружающей среды в некую часть популяции клеток; (2) размножение фагов внутри инфицированных клеток; (3) 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, приобретая иммунитет к данному типу фага. Выбор лизогенного или литического сценария зависит от условий, в которых в момент заражения находятся клетки популяции.

Раздел 2.6 содержит описание программного комплекса ГЭК, разработанного в диссертации. Описаны: Общая архитектура ПК ГЭК (2.6.1); Организация потока вычислений ПК ГЭК (2.6.2); Модули высокопроизводительных вычислений с параллелизацией по генетическому разнообразию сообщества (2.6.3) и по пространственной структуре среды обитания сообщества (2.6.4); Модули интеграции ПК ГЭК с существующими средствами моделирования генных сетей и базами данных (2.6.5); Графический интерфейс пользователя ПК ГЭК и веб-приложения (2.6.6). 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 процессорами; (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 (рис. 2).

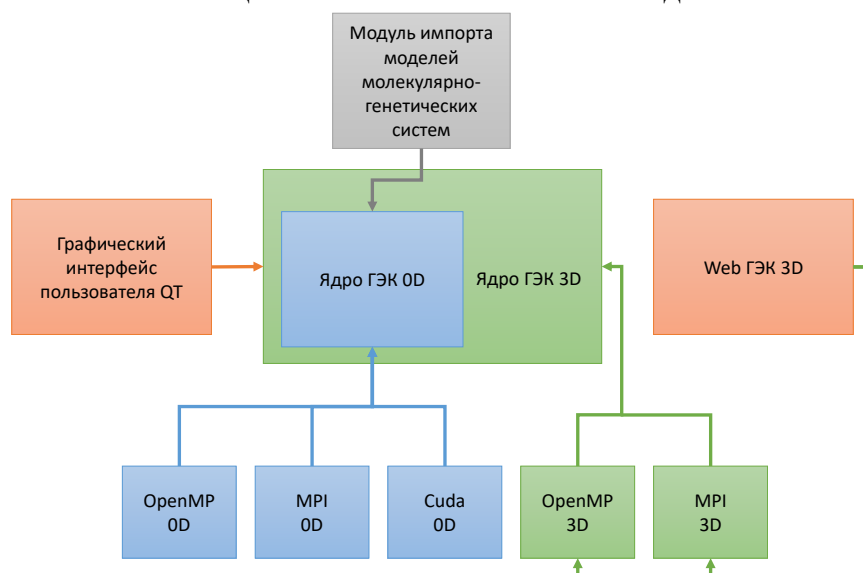


Рисунок 2. Общая схема взаимодействия модулей программного комплекса ГЭК.

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 потока для метаболитов, а для клеток – пассивное (с помощью потока) 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.

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.

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 (рис. 3). Кроме того, выполнена интеграция базы данных наименований химических веществ ChEBI. Для всех модулей выполнена верификация и показано соответствие их функциональности заявленным 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 WebNEC.

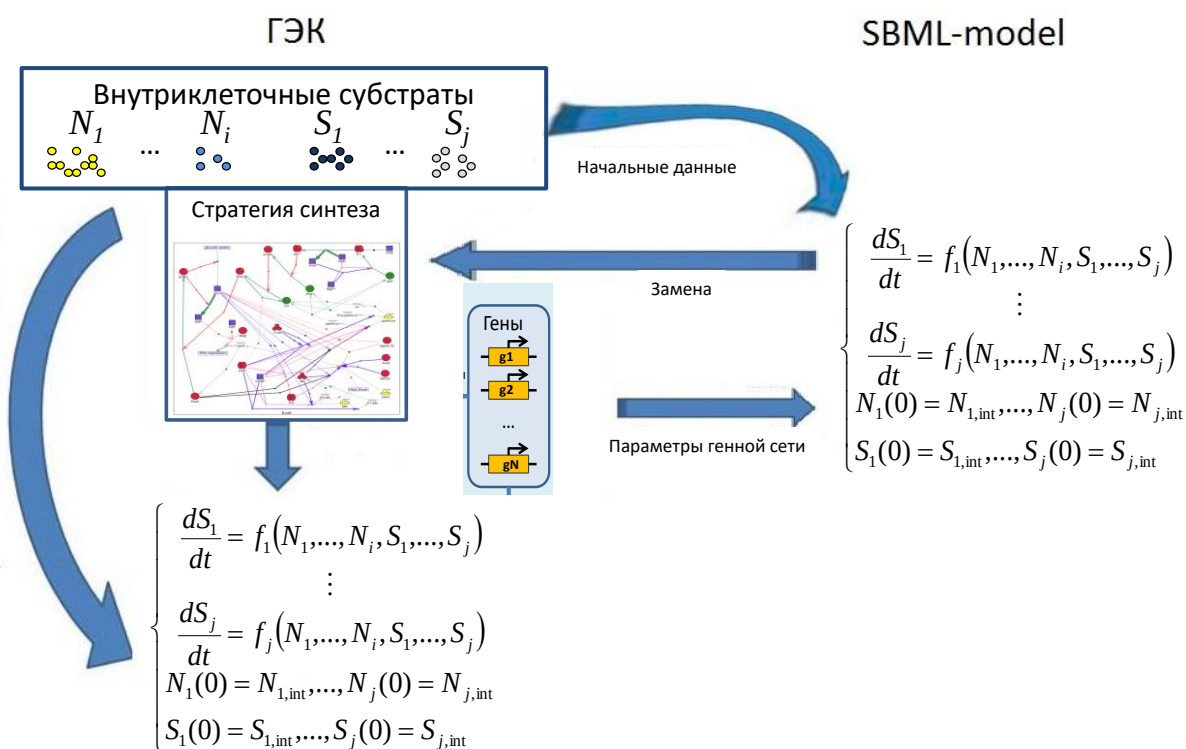


Рисунок 3. Общая схема интеграции ГЭК и SBML.

В разделе 2.7 приведены содержательные биологические результаты, полученные с использованием ПК ГЭК. В подразделе 2.7.1 приведены результаты компьютерного моделирования эволюции микробной популяции по преодолению локальных минимумов при достижении пика на ландшафте приспособленности. Была построена модель эволюционной динамики в популяции микроорганизмов, состоящей из клеток

одного вида, населяющих резервуар с постоянным притоком неспецифического субстрата S , являющегося источником энергии для живущих в нём клеток (Рис. 4А). Приспособленность клеток определялась размером прироста популяции в единицу времени и зависела от эффективности утилизации субстрата, доступности субстрата и смертности. Популяционная динамика (ΔP) описывается уравнением 4:

$$\Delta P = P \cdot K_u \cdot S_{uptaken} - k_{flow} \cdot P - k_{dens} \cdot P^2 \quad (4),$$

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 :

$$S \xrightarrow{K_u} S' + energy \quad (5)$$

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. 4Б). Таким образом, K_u зависит от аллельных вариантов обоих локусов; обозначим эти варианты e_1 и e_2 : $K_u = F(e_1, e_2)$. 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].

Были исследованы эволюционные ландшафты приспособленности, задаваемые в модели функцией K_u [Лашин и др., 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, Суслов, Матушкин, 2012]), либо сальтационный (быстрые и значительные преобразования предковых форм [Иорданский, 2001; Лашин, Суслов, Матушкин, 2012]) эволюционные режимы.

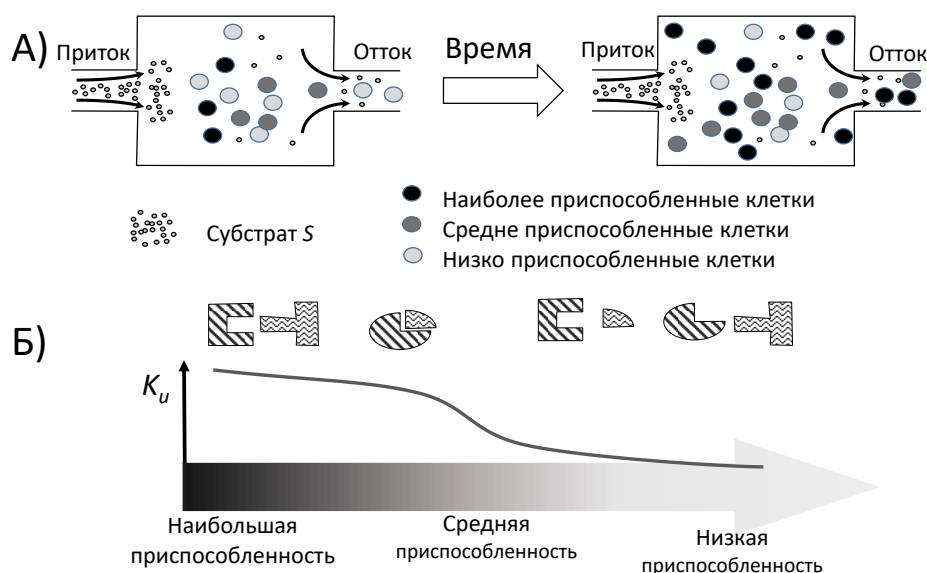


Рисунок 4. А) Схема моделируемой системы: проточный резервуар, на вход которого подаётся неспецифический субстрат – источник энергии. Резервуар населяет гетерогенная микробная популяция (клетки разных генотипов показаны разным цветом). С течением времени меняется как размер популяции, так и соотношения между представленностями клеток различных генотипов. Б) Схема функции K_u , демонстрирующая изменение приспособленности в зависимости от комбинации аллельных вариантов.

Подраздел 2.7.2 посвящён описанию результатов моделирования формирования и конкуренции популяций подвижных микроорганизмов с различными адаптационными стратегиями в периодически изменяющейся среде. 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 (рис. 5а). Рассматривалась 2D-среда обитания в виде квадрата размером 10×10 ячеек с замкнутыми границами, в которую периодически с двух противоположных углов подавался питательный субстрат (рис. 5б). 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. 5в). 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. 5а).

В каждой пространственной ячейке численность популяции подчиняется следующему уравнению 6, 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. Рассматривалось два типа смертности: квадратичная (плотностно-зависимая) и линейная (пропорциональная).

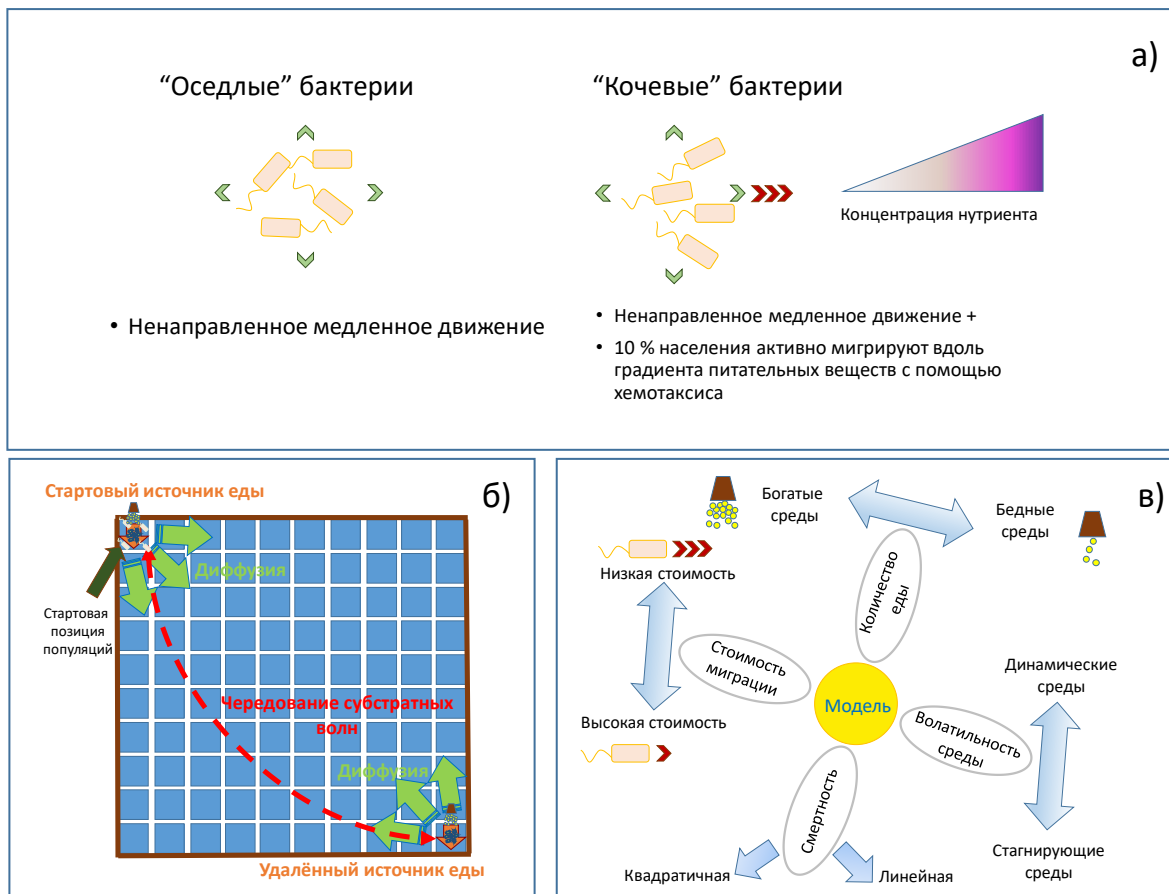


Рисунок 5. Схема модели конкуренции оседлой и кочевой популяции: а) трофическая структура сообщества; б) схема пространственной организации среды обитания сообщества; в) различные сценарии, рассмотренные в модели.

$$P_{n+1} = P_n * \left(1 + \frac{x/K}{1 + x/(C*K+N)} \right) - mort(P) \quad (6)$$

Мы моделировали адаптивную миграцию клеток через хемотаксис, предполагая, что подвижные клетки, способные к хемотаксису, распространяются через окружающую среду как за счет случайного плавления, так и за счет градиентов питательных веществ, тогда как клетки без хемотаксической способности распространяются только случайным образом без какого-либо определенного направления (подробнее см. [Klimenko et al., 2021]). Для учёта миграционных издержек нами была предложена подмодель *штрафа за миграцию* зависящая от двух параметров. Первый параметр определяет степень нелинейности функции миграционных издержек, и он был зафиксирован; 2-й 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.

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. Для модели, учитывающей затраты энергии на миграцию, также варьировалась величину миграционного штрафа. Результаты моделирования показали, что стратегия «оседлых» может иметь преимущество перед стратегией «кочевников» в различных условиях, даже в отсутствие штрафов на миграции. 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, что наблюдается в ряде численных экспериментов (рис. 6).

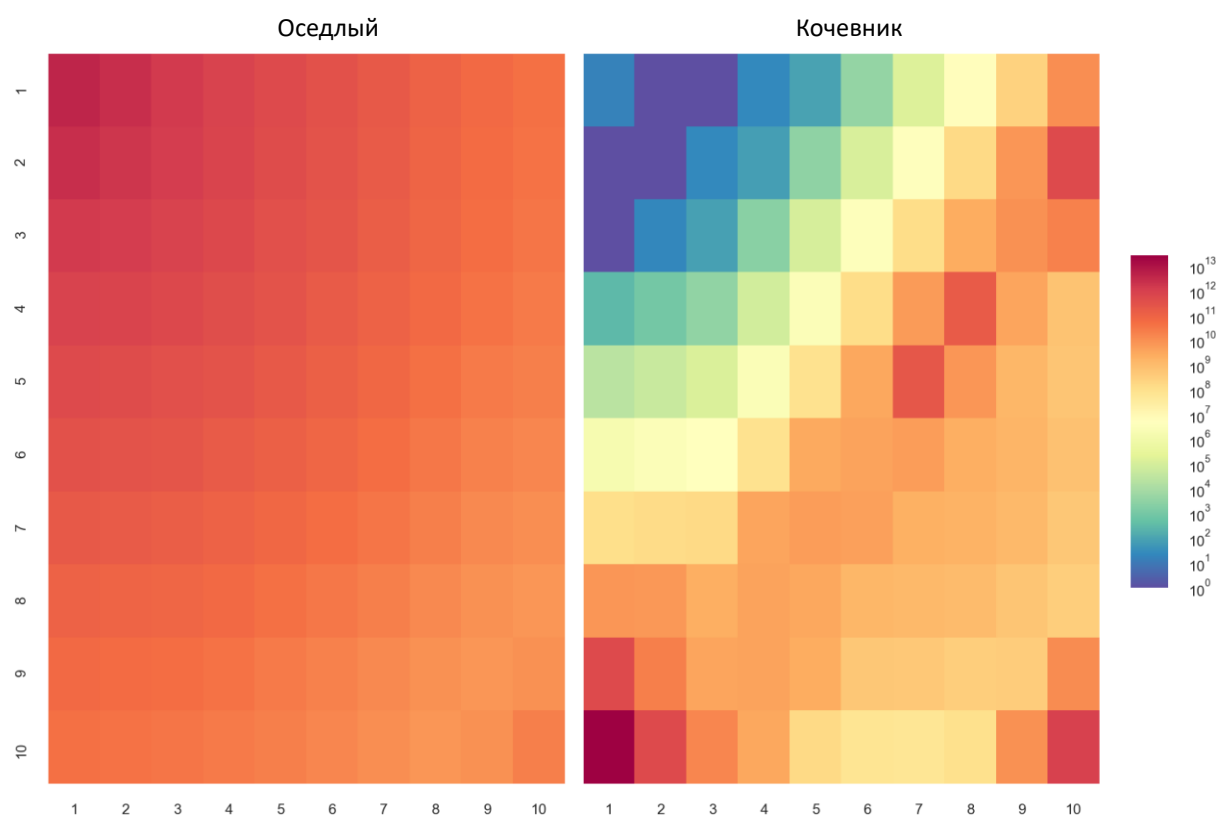


Рисунок 6. Тепловые карты пространственных распределений временных снимков клеток «оседлых» и «кочевников» в среде обитания, сделанные в конце численного эксперимента для богатой нутриентами среды обитания. Шкала тепловой карты представляет собой численность популяции (количество клеток) в конечный момент времени моделирования (9000 поколений).

Поскольку клетки «оседлой» популяции уменьшают локальную доступность питательных веществ в непосредственной близости от источника, клетки популяции «кочевника» не могут пройти 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, гонящиеся за лучшими условиями (иногда прирост может быть совсем небольшим), терпят неудачу при длительном стрессе, связанном с пересечением долин субоптимальной доступности питательных веществ.

Рассмотренная в подразделе 2.7.2 модель «Оседлый-кочевник» была дополнительно исследована с точки зрения эволюции признаков эффективности утилизации нутриентов и эффективности миграции (**подраздел 2.7.3**). 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 микроорганизмов, эволюционирующих в пространственно-распределённой среде с диффузией и периодическим во времени и пространстве субстратом, являющейся аналогом застойного водоёма (рис. 5). 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 эффективность утилизации субстрата и интенсивность миграции (мобильность). Исследовались следующие вопросы: (1) как соотносятся по степени влияния на приспособленность популяций такие факторы как эффективность утилизации субстрата и мобильность популяции; (2) в каких условиях больший выигрыш для приспособленности популяции будет давать более эффективная утилизация субстрата, а в каких – более высокая интенсивность миграции. Было 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.

Подраздел 2.7.4 посвящён моделированию влияния фаговой инфекции на эволюцию микробных сообществ при различных экологических условиях. Было показано, 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. 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.

Таким образом, **раздел 2.7** показывает, что главное преимущество моделей ГЭК состоит в том, что *с их помощью можно проследить причинно-следственные связи, возникающие на разных уровнях организации биологической системы.*

В **разделе 2.8** представлены методика, а в **разделе 2.9** – программный комплекс Microcosm, реализованные в агентно-ориентированной парадигме (с разрешением до отдельной клетки), являющийся дальнейшим развитием ПК ГЭК и предназначенный 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. В **разделе 2.9** приводятся содержательные биологические результаты, полученные с использованием ПК Microcosm. Проведены вычислительные эксперименты по воспроизведению классических моделей динамики и генетики микробных популяций, а также вычислительные эксперименты по сравнению эволюционной эффективности различных моделей перемещения клеток в среде обитания, по оценке влияния диффузионных параметров нутриентов на эволюцию сообществ и другие.

Глава 3 посвящена моделированию популяций диплоидных сложноорганизованных организмов. Описываются результаты работ по разработке инструментария для теоретического исследования популяций 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. Описана разработанная методика моделирования (**раздел 3.1**) и программный комплекс «Диплоидный эволюционный конструктор» – ДЭК (**раздел 3.2**), ориентированный 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. 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 7:

окружающая среда может содержать несколько популяций, каждая из которых может содержать несколько особей (характерные размеры популяции – 100-10000 особей).

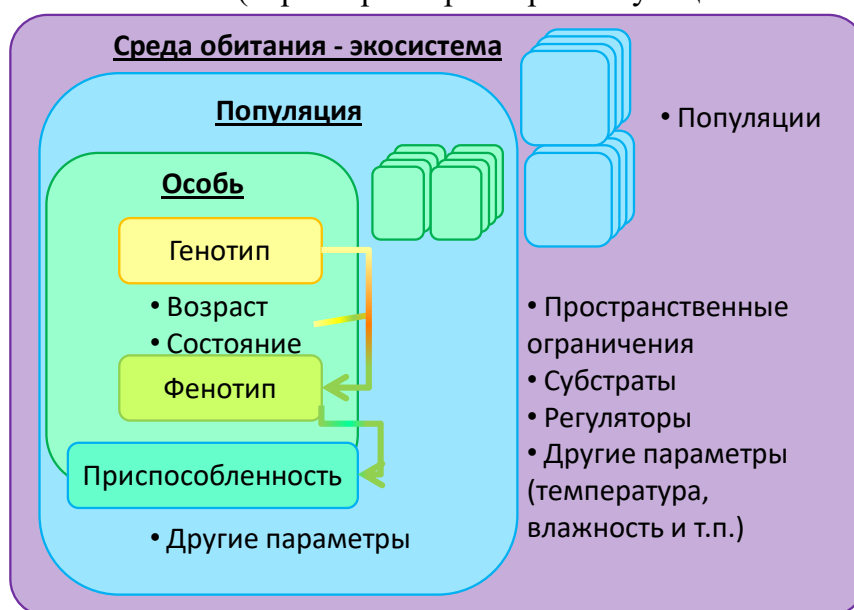


Рисунок 7. Основные объекты методики ДЭК и их свойства.

Взаимодействие между объектами модели в методике осуществляется по принципу многослойного моделирования, предложенному нами ранее для описания эволюционно-популяционных процессов в бактериальных сообществах. 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 основным 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 Н.А., Гончаров С.С., Лихошвай В.А., 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, Мирнова, 2014; Колчанов Н.А., Гончаров С.С., Лихошвай В.А., 2008; Ратушный и др., 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 **разделе 3.1** 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 в **разделе 3.2**, тоже получил название ДЭК, с помощью которого моделируются эволюционные, популяционно-генетические и популяционно-экологические процессы в популяциях сложноорганизованных организмов (особей). В разделе приводятся разработанные спецификации использования программного комплекса ДЭК и его архитектура. Выделены следующие модули: 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.

В подразделе 3.2.1 описываются модули поддержки высокопроизводительных вычислений ПК ДЭК, реализованные с использованием таких технологий как MPI, OpenMP и Nvidia Cuda. Версия ПК ДЭК с использованием OpenMP была протестирована на различных компьютерах и показала значительное ускорение как на офисном компьютере – до 13 раз, так и на кластере Сибирского Суперкомпьютерного Центра СО РАН – до 60-70 раз, что позволило 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.

В подразделе 3.2.2 приведено описание графического интерфейса пользователя ПК ДЭК. Реализованы два варианта использования ПК: (1) интерактивный (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 (рис. 8) и (2) пакетный (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.

Раздел 3.3 содержит описание содержательных биологических результатов, полученных с использованием ПК ДЭК. Верификация ПК ДЭК (**подраздел 3.3.1**) проведена на классических задачах генетики популяций, в частности, смоделирована теорема Фишера. Верификация показала применимость ПК ДЭК для решения задач моделирования эволюционно-популяционных процессов. **Подраздел 3.3.2** описывает результаты моделирования инбридинга и аутбридинга в популяциях диплоидных организмов, фенотип которых определяется двумя сложными признаками – фертильностью и устойчивостью к болезни. При определении признаков рассматривались разные схемы доминирования: полное доминирование, кодоминирование, а также материнский эффект. Фертильность определялась с учётом кумулятивного действия двух простых количественных признаков (мультипликативный эффект), но имела при этом сложный ландшафт значений с двумя пиками повышенной приспособленности. Признак устойчивости к болезни определялся четырьмя простыми количественными признаками, кодируемых соответствующими генами. При этом также рассматривались разные схемы доминирования и, кроме того, аддитивные и мультипликативные эффекты.

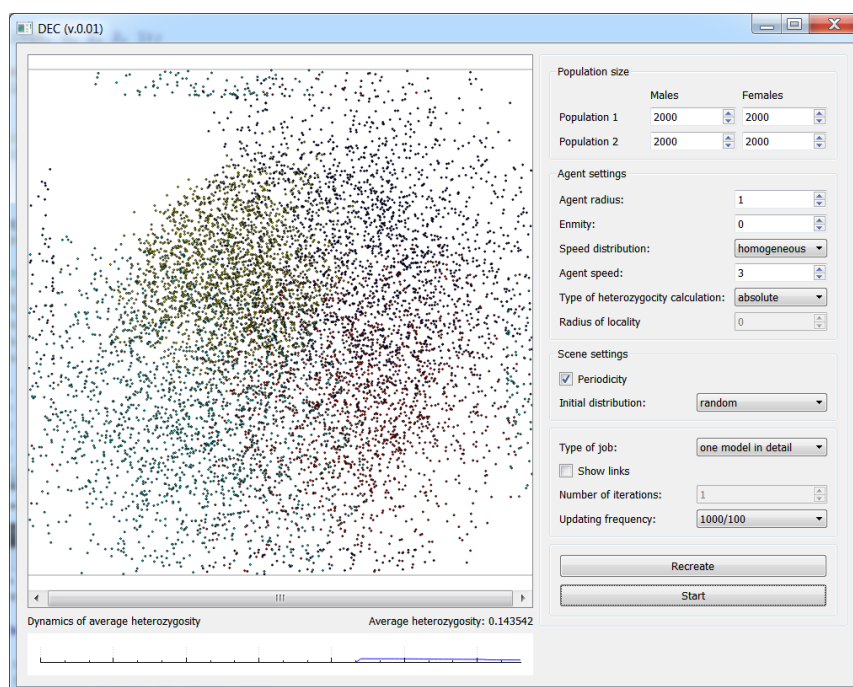


Рисунок 8. Пример графического интерфейса ПК ДЭК: модель смешивания двух популяций в результате миграций.

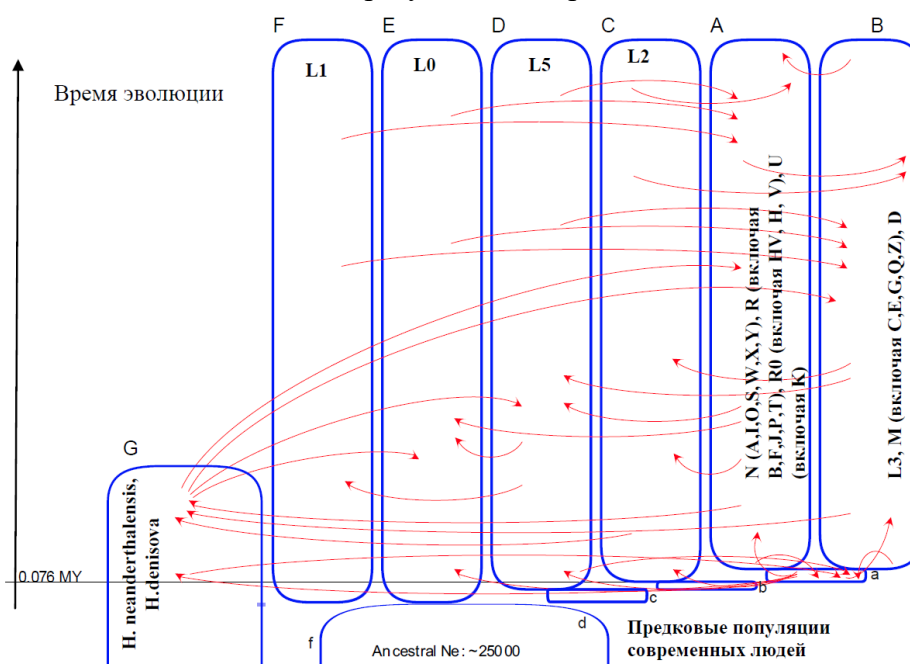


Рисунок 9. Схема миграций в популяциях человека, реконструированная на основе последовательностей мтДНК (По: [Gunbin и др., 2012]) и воспроизведённая в моделях ДЭК.

В подразделе 3.3.3 приведены результаты моделирования с помощью ПК ДЭК различных сценариев популяционно-генетических и миграционных процессов в популяциях древних людей. На основе полученных с помощью литературы оценок популяционно-генетических параметров, с помощью ПК ДЭК была построена серия компьютерных моделей эволюции предков современных людей на территории Евразии. Проанализировано несколько возможных эволюционных сценариев, представленных на рисунке 9.

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.

Наконец, в **подразделе 3.3.4** была исследована эволюция генных сетей с отрицательными обратными связями под действием разных типов отбора. Модель эволюции генной сети с регуляторным контуром, работающим по принципу отрицательной обратной связи (ОС), ранее описанная в [Колчанов, Шиндялов, 1991], была воспроизведена 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

$$W(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-X_0}{\sigma_X}\right)^2} \quad (7)$$

Концентрация белка X определялась следующим уравнением:

$$\frac{dX}{dt} = k_s - k_d X \quad (8)$$

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 разных типов генетического доминирования. Моделирование показало, что в условиях стабилизирующего отбора (неизменные условия окружающей среды и неизменное значение X_0 в уравнении 7) 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. 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 существенный фактор (обеспечивая пул скрытой изменчивости, которая при смене условий будет вскрываться).

Глава 4 посвящена моделированию эволюции популяций с учётом социальных факторов. Важной отличительной особенностью популяций животных и, особенно, человека, по сравнению с популяциями микробов и других простых организмов, яв-

ляется 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 и др., 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 и др., 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.

В разделе 4.1 рассмотрена модель «Агрессор — Защитник», предложенная в работе [Bonabeau, 2002]. Она представляет собой агентно-ориентированную модель динамики сообщества, члены которого (агенты, игроки) двигаются 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 (подраздел 4.1.1) и существенно расширена (подраздел 4.1.2) модель «Агрессор-Защитник» Bonabeau [Bonabeau, 2002]. Были предло-

жены 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 подразделе 4.1.3 приведены основные режимы динамики расширенной модели «агрессор — защитник», полученные в результате численных экспериментов. Показано, что: (1) динамика модели зависит от выбора типа 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; (2) 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; (3) неудовлетворённость трусов всегда выше неудовлетворённости защитников, однако увеличение доли трусов в популяции ведёт к уменьшению данной величины.

Другой содержательной моделью, рассмотренной в Главе 4, является описанная в разделе 4.2 компьютерная модель распространения генетической глухоты в популяциях человека, учитывающая как генетические, так и социальные факторы. Модель предназначена для выявления того, как на распространение генетической глухоты влияет наличие жестокости 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, предоставленные к.б.н. Барашковым Н.А., к.б.н. Романовым Г.П., к.б.н. Бады-Хоо М.С., а также курировавшей биологическую часть работы к.б.н. Посух О.Л. Модель была построена в терминах описанной в Главе 3 модели, лежащей в основе программной платформы ДЭК, блок-схема цикла моделирования приведена рис. 7а. 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 (рис. 7б). 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. Верификация и валидация модели была проведена на данных по анализу влияния языковой гомогамии на распространённость аутосомно-рецессивной наследственной

глухоты (DFNB1) в США [Nance, Kearsey, 2004; Arnos et al., 2008]. Кроме того, было проведено компьютерное моделирование динамики распространения рецессивной генетической глухоты в изолированной популяции якутов. Были исследованы различные сценарии эволюции популяции на протяжении 400 лет (20 поколений). В референсной популяции при сценарии «Отбор» наблюдалось снижение доли глухих (с 0,25% до 0,07%) и частоты патогенного аллеля (с 5,25% до 2,57%). Этот сценарий предполагал, что глухие люди не могут вступать в брак, если они не используют язык жестов для общения, и поэтому не имеют потомства. Моделирование по сценарию «Ослабленный отбор» показало, что доля глухих значительно увеличивается в течение первых 4 поколений (80 лет), а затем снижается к 20-му поколению (400 лет), что примерно в 3 раза выше по сравнению с популяцией без сообщества глухих. Частота рецессивного мутантного аллеля снизилась с 5,25% до 3,96%. Эти данные позволяют предположить, что ассортативные браки между глухими людьми, основанные на лингвистической гомогамии (язык жестов), приводят к увеличению частоты наследственной потери слуха. Моделирование по сценарию «Отсутствие отбора» ожидаемо показало, что распространенность наследственной потери слуха не изменяется и остаётся на исходном уровне. Этот сценарий предполагает, что развитие современных медицинских технологий и социальных программ реабилитации глухих приводит к их большей интеграции в общество и, как следствие, отсутствию снижения репродуктивных возможностей, несмотря на их патогенные генотипы.

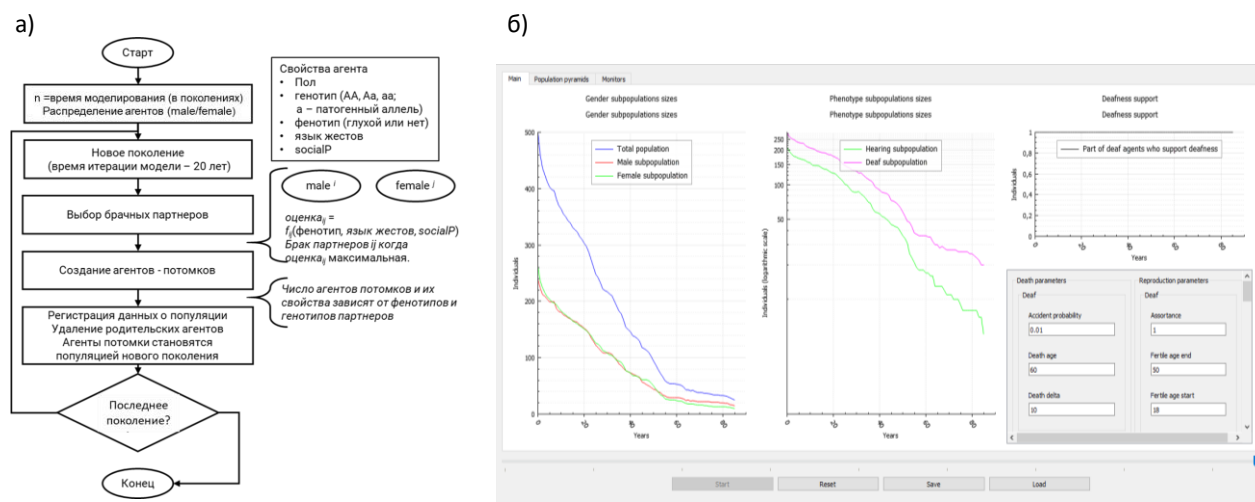


Рисунок 10. а) Блок-схема основного цикла программы PGD-S. б) Графический интерфейс пользователя программы PGD-S.

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 Бергера [Lashin и др., 2012; Lashin, Matushkin, 2012; Lashin, Suslov, 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 моделирования 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. 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 многоуровневого моделирования позволяет получать нетривиальные 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 низкомобильными.

с. Для низкомобильных 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.

6. 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 ОСНОВНЫХ ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

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–фаг” // *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. NEC 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., Tomsy 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., Pshenikova 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.