

**Иголкина Анна Андреевна**

**Реконструкция эволюционной истории нута с применением  
моделирования сложных событий смещений и  
композиционного анализа данных**

Специальность 1.5.8 —  
«Математическая биология, биоинформатика»

**Автореферат**  
диссертации на соискание учёной степени  
кандидата биологических наук

Работа выполнена в Федеральном государственном бюджетном научном учреждении «Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии» .

Научный руководитель: **Андронов Евгений Евгеньевич**  
доктор биологических наук, главный научный сотрудник лаборатории микробиологического мониторинга и биоремедиации почв, ФГБНУ ВНИИСХМ

Официальные оппоненты: **Щур Владимир Львович,**  
доктор физико-математических наук, заведующий лабораторией Международная лаборатория статистической и вычислительной геномики, НИУ ВШЭ

**Матвеева Татьяна Валерьевна,**  
доктор биологических наук, профессор кафедры генетики, СПбГУ

Ведущая организация: Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений имени Н.И. Вавилова (ВИР)

Защита состоится 29 сентября 2025 г. в 15.00 на заседании диссертационного совета Д 24.1.101.01 при Институте проблем передачи информации имени А. А. Харкевича Российской академии наук (ИППИ РАН) по адресу: 127051, г. Москва, Большой Каретный переулок, д.19 стр. 1.

С диссертацией можно ознакомиться в библиотеке ИППИ РАН, а также на сайте ИППИ РАН по адресу:  
[http://iitp.ru/upload/content/1728/igolkina\\_dissertation\\_2025\\_04\\_08.pdf](http://iitp.ru/upload/content/1728/igolkina_dissertation_2025_04_08.pdf).

Автореферат разослан «\_\_» \_\_\_\_\_ 2025 г..

Ученый секретарь  
диссертационного совета  
Д 24.1.101.01,  
доктор биологических наук

\_\_\_\_\_ Казенников Олег Васильевич

## Общая характеристика работы

### Актуальность

Доместикация (введение в культуру) растений представляет собой биологическую коэволюцию с участием человека, направленную на создание новых форм с улучшенными и полезными свойствами. Активная доместикация началась примерно 10–12 тысяч лет назад и продолжалась до 1-2-го тысячелетия до н.э. (Meyer et al., 2012). После первичного этапа окультуривания растений продолжалось их дальнейшее поддержание и улучшение в соответствии с потребностями человека: адаптация к разнообразным условиям окружающей среды, повышение урожайности и улучшение вкусовых качеств, неприхотливость в хранении и приспособленность к методам выращивания. Все это осуществлялось методами народной селекции. Исследования в области доместикации в первую очередь мотивированы экономическими интересами. Однако, помимо этого, они предоставляют ответы на фундаментальные вопросы культурно-исторического значения, такие как определение местоположения и времени начала доместикации. Научная ценность таких исследований заключается в том, что знания, полученные о введении в культуру одного вида растений, могут быть применены к другим благодаря сходству процессов.

На протяжении тысячелетий оседлые сообщества, а затем крестьяне и местные фермеры занимались доместикацией и улучшением растений. Такая примитивная или народная селекция сохранялась до начала XIX века, точки начала интенсивного развития методов научной селекции и системного подхода к улучшению растений. Эти достижения к середине XX века привели к так называемой ”Зелёной революции”— периоду разработки и применения централизованных сельскохозяйственных программ, направленных на повышение производительности и глобальное обеспечение существования больших масс населения. Последствия этого масштабного перехода оказались неоднозначными: произошло сокращение генетического разнообразия сельскохозяйственных культур и массовое использование современных сортов, которые не редко экологически неустойчивы, уязвимы к изменениям климата и бедны питательными веществами (Pingali, 2019). Восстановлению разнообразия может помочь обращение к генетическому материалу, не затронутому научной селекцией и Зелёной революцией, например, к так называемым староместным сортам или сортам народной селекции (англ. landraces) Такой генетический материал хранится в банках семян, крупнейшим из которых в России является коллекция генетических ресурсов растений, расположенная в Федеральном исследовательском центре — Всероссийском институте генетических ресурсов растений имени Н. И. Вавилова (ВИР). Основным источником староместных сортов стали многочисленные ботанико-агрономические экспедиции, проведённые в первой половине XX века. Для правильного внедрения этого материала в селекционную разработку при создании новых сортов важно знать их генетическое родство, а

именно эволюционную историю формирования генетических пулов в регионах доместикации.

Одной из сельскохозяйственных культур в коллекции ВИР является нут, бобовое растение, служащее незаменимым источником высококачественного белка (Abbo et al., 2003), и занимающее второе место в мире среди бобовых по производству зерна (Dutta et al., 2022). Ввиду значимости нута в сельском хозяйстве исследования его глобального генетического разнообразия сохраняют свою актуальность. Несмотря на то, что история окультуривания нута частично установлена с помощью археологических данных, в этом вопросе все еще присутствуют значительные пробелы. Во-первых, генетико-географические группы разнообразия нута описаны не полно. Во-вторых, остается неизвестной эволюционная история взаимосвязи между этими группами, включая потенциальные события смешения.

Для ответа на поставленные вопросы необходимо составить репрезентативную выборку мирового разнообразия нута. В нее должны войти староместные сорта собранные по всему ареалу распространения нута. Соответствующая этим требованиям выборка из 421 образца староместных сортов была ранее сформирована из коллекции ВИР (Sokolova et al., 2020). Хотя географическое распределение этой выборки ожидаемо неравномерно, она достаточно полна для выделения генетико-географических групп нута.

Реконструкция эволюционной истории возникновения этих генетико-географических групп подразумевает построение графа смешений (англ. admixture graph). В то время как формирование новых групп в результате миграций обычно представимо в виде бинарного дерева, граф смешений является направленным ациклическим графом более сложной сетевидной структуры, которая учитывает дополнительные сценарии: события смешения могут быть множественными, вложенными и включать более двух популяций-источников. Существующие модели реконструкции графов смешений не охватывают все из перечисленных сценариев, что делает актуальной необходимость разработки новых, более полных моделей.

Дополнительной сложностью в задачах реконструкции является необходимость индивидуального подхода к хромосомным блокам. В то время как на уровне общих закономерностей в рамках всего генома эволюционная история популяций может быть весьма однозначной, на уровне отдельных хромосомных блоков она может иметь отклонения. Это происходит из-за того, что разные участки ДНК могут подвергаться различным независимым воздействиям — например, естественному отбору или генетическому дрейфу — и соединяться или перестраиваться в процессе рекомбинации. В данном контексте особую важность приобретает концепция графа предковых рекомбинаций (англ. ancestral recombination graph). Эта структура позволяет проследить уникальную эволюционную историю каждого конкретного участка хромосомы, представляя её в виде бинарного дерева. По аналогии с этим можно предположить, что каждому

хромосомному блоку также может соответствовать свой уникальный граф смешений. Таким образом, имеет место задача реконструкции графов смешений в режиме скользящего окна вдоль хромосом.

Большинство методов реконструкции эволюционной истории популяций базируются на модели Райта-Фишера. Эта модель описывает изменения частот аллелей во времени в популяции фиксированного размера и основывается на предположениях о случайном скрещивании в отсутствие мутаций, миграций и естественного отбора. Модель Райта-Фишера является дискретной по времени, поэтому для в методах реконструкции она используется в непрерывной аппроксимации. Все известные аппроксимации предполагают, что каждый генетический локус имеет два аллельных варианта, и описывают изменение частоты только одного аллеля, считая второй зависимым. Во-первых, такой подход является несимметричным. Во-вторых, он не является масштабируемым на случай мультиаллельных локусов. Поэтому необходимо разработать аппроксимации модели Райта-Фишера, которые обходят эти ограничения.

Частоты аллелей в популяции принадлежат к композиционному типу данных. Композиционные данные представляют собой наборы значений определенной длины (так называемые композиции), которые отражают относительные частоты, доли или проценты компонентов внутри фиксированного общего объема или суммы. Например, композициями являются все наборы длины 3 следующего вида:  $x = (x_1, x_2, x_3) : x_1 + x_2 + x_3 = 1, x_i > 0$ . Все элементы в композиции не являются независимыми: увеличение одного элемента композиции влечет уменьшение всех остальных. В связи с этим многие стандартные методы статистического анализа (например, выборочная корреляция Пирсона) являются неприменимыми и дают ложные результаты. Поэтому для работы с композиционными данными была предложена теория композиционного анализа данных (Aitchison, 1982). Эту теорию начали разрабатывать в середине XX века, и она хорошо себя зарекомендовала в геологии, экологии, химии, микробиологии и других дисциплинах. Использование композиционного анализа данных в популяционной генетике пока не распространено, несмотря на то, что эта область оперирует частотами аллелей, что напрямую относится к композиционным данным.

### **Степень разработанности темы исследования**

Существует ряд исследований популяционной структуры нута, крупнейшим из которых является работа на 3366 образцах, включая 2436 староместных сортов и 197 дикорастущих образцов (Varshney et al., 2021). В этом исследовании все образцы были генотипированы по однонуклеотидным полиморфизмам (ОНП, англ. single nucleotide polymorphism, SNP) и распределены на семь групп. По части анализа этих групп работа имеет ряд недостатков. Во-первых, группирование было сделано исключительно по географическому принципу, без учета генетических сходств и различий. Во-вторых, не были получены генетические характеристики для групп - вектора характеристических частот аллелей

ОНП. В-третьих, гипотезы о распространении популяций нута были сформулированы только на основании визуального анализа представления генетических данных в двух первых главных компонентах. Следовательно, более детальный анализ популяций нута необходим. Наилучшим способом подтверждения или опровержения наличия выделенных групп и сформулированных гипотез распространения является независимый анализ на отдельной, но сравнимой по структуре выборке. Отобранная из коллекции ВИР и генотипированная по ОНП панель из 421 староместного сорта нута как раз является таковой. Серьезным улучшением в понимании популяционной структуры нута станет детальный анализ популяций и получение их характеристик на базе отобранной панели как на независимом источнике, а предыдущие исследования будут использоваться для сравнения результатов на качественном уровне.

В отсутствие четко определенных генетико-географических групп нута ряд вопросов, касающихся их происхождения, может оставаться нераскрытым и даже несформулированным. Несмотря на неполноту описания популяций нута, для одной из них была продемонстрирована географическая и генетическая обособленность — это относится к нуту дези из Эфиопского региона. В исследовании (Varshney et al., 2021) Эфиопская популяция занимает изолированное положение на графике первых двух главных компонент, располагаясь как бы в центре между всеми остальными популяциями (Рисунок 1). Так как Эфиопская популяция не является центром первичной доместикикации, такое ее положение потенциально может свидетельствовать о сложном и смешанном происхождении. Этот результат был интерпретирован как то, что Эфиопская популяция произошла из Ближневосточного или Среднеазиатского региона. Однако вопрос о том, произошла ли эта популяция в результате событий смешения или просто в результате миграции, остается открытым.

Существуют две основных культивируемых группы сортов нута, отличающихся по внешнему виду, вкусу и другим характеристикам: дези и кабули. Место первичной доместикикации дези было установлено с помощью археологических находок и подтверждено совпадением с ареалом произрастания дикого типа, который фенотипически схож с дези. Регион первичной доместикикации кабули остается неизвестным. В литературе были выдвинуты предположения, что он связан с Кабулом, исходя из лингвистического сходства между словами "кабули" и "Кабул". Однако попыток количественных подтверждений или опровержений этой гипотезы не было предпринято.

Существуют более точные методы выделения генетических групп в наборе образцов, чем простой анализ главных компонент. К таким методам относятся программы типа STRUCTURE и ADMIXTURE [0]. Принцип их работы похож на метод кластеризации k-средних, поэтому эти методы косвенно определяют векторы частот аллелей ОНП для каждой группы. Однако ни один из существующих методов выделения генетических групп в данных не учитывает их географическую неравномерность. Следовательно, требуется разработка метода оценки

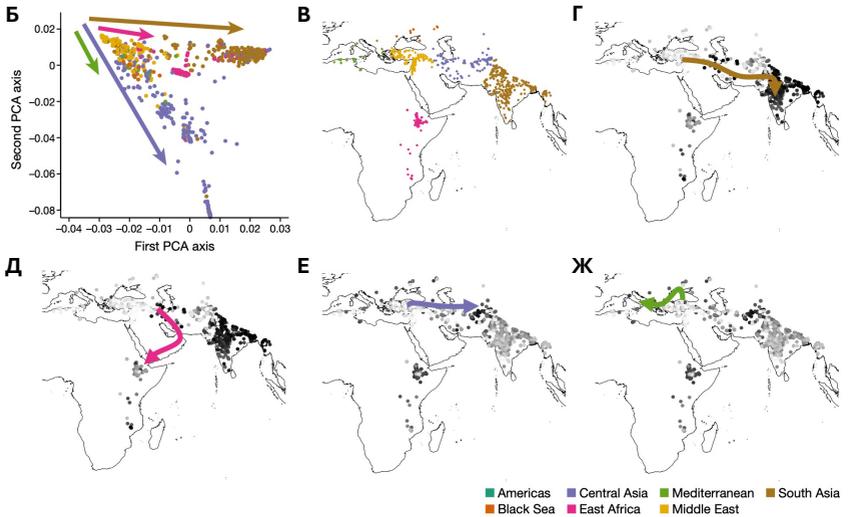


Рис. 1 — А. Карта центров происхождения культурных растений, предложенных Н.И. Вавиловым. Оригинал находится во Всероссийском институте растениеводства имени Н.И. Вавилова. Фото выполнено Иголкиной А.А. Б-Ж. Гипотетические пути распространения нута на основании анализа главных компонент. Рисунок заимствован из работы Varshney et al., 2021, Fig. 2. Б. Представление генетического разнообразия выборки в координатах двух первых главных компонент. В. Географические точки сбора образцов. Популяции раскрашены согласно географическим кластерам. Г. Путь распространения нута в Южную Азию. Д. Путь распространения нута в Восточную Африку. Е. Путь распространения

генетических характеристик групп образцов с учетом их географического распределения.

Для решения задачи реконструкции графа смещений для выделенных групп существует несколько методов. Различия между ними заключаются в подходе к моделированию событий смешения: асимметричная для популяций-источников модель [0], модель с ограниченным числом вложенных событий смешения [0] и модель смешения с не более чем двумя источниками [0]. Не существует единого метода, который бы использовал симметричную модель с более сложными событиями смешения, допускающий участие более чем двух популяций-источников. Таким образом, существует потребность в разработке нового усовершенствованного метода.

Существует еще несколько возможных усовершенствований, которые ранее не применялись в методах реконструкции эволюционной истории популяций. Во-первых, как уже упоминалось выше, целесообразно оценивать граф смещений в режиме скользящего окна. Во-вторых, использование композиционного подхода к аппроксимации модели Райта-Фишера вместо стандартной нормальной аппроксимации позволит не только более точно учитывать природу частот, но и перейти от моделирования биаллельных локусов к мультиаллельным.

**Цель** данной работы заключается в установлении происхождения популяций нута в результате разработки и применения новых методов для реконструкции эволюционной истории.

Для достижения поставленной цели необходимо было решить следующие задачи:

1. Предложить аппроксимацию модели Райта-Фишера с учетом композиционной природы частот аллелей.
2. Разработать модель для оценки характеристических частот аллелей в популяции с учетом ковариационных связей между образцами, которые определяются их географическим местоположением.
3. Разработать модель для реконструкции эволюционной истории популяций, учитывающую события смешения сложной структуры и неравномерность сигнала смешения вдоль хромосом.
4. Разработать программные пакеты на основе этих моделей и провести их тестирование на симулированных данных.
5. Исследовать генетическую структуру панели образцов нута из коллекции ВИР и сгруппировать образцы в отдельные популяции по географическому и генетическому сходству.
6. Применить разработанные модели к выделенным генетико-географическим группам разнообразия нута.
7. Установить происхождение Эфиопской и других популяций нута дези.
8. Установить регион происхождения нута кабули.

**Научная новизна**

1. Идентифицированы новые генетико-географические популяции нута: Ливанская, западно-Среднеазиатская, а также две отдельные Средиземноморские популяции – северобережная и южнобережная.
2. Впервые разработана модель – `popdisp` – для оценки частот аллелей в популяциях, использующая теорию анализа композиционных данных и позволяющая учитывать ковариационную связь образцов в популяции.
3. Впервые разработана модель – `migadmi` – для реконструкции эволюционной истории популяций, способная моделировать множественные и вложенные события смещения, а также учитывающая композиционную природу частот аллелей.
4. Впервые продемонстрирована связь между генетическим сходством образцов внутри популяций и географически оптимальным путем перемещения популяций по ареалу.
5. Впервые определено происхождение двух смешанных популяций нута дези (Эфиопской и западно-Среднеазиатской), а также установлено происхождение нута кабули. Кроме того, предложены гипотезы о торговых путях, которые могли способствовать формированию этих смешанных популяций.

**Теоретическую значимость работы** определяют:

- Аппроксимация модели Райта-Фишера в терминах композиционного анализа данных.
- Байесовская иерархическая модель для оценки характеристических частот аллелей в популяции, учитывающая ковариационную матрицу связей между образцами. Расширение модели на случай многоаллельных генетических маркеров.
- Формула преобразования матрицы расстояний между образцами в ковариационную матрицу с учетом репрезентативного центра популяции.
- Параметризация события смещения нескольких популяций, позволяющая смешанной популяции обладать собственной дисперсией после события смещения.
- Модели смещения популяций с использованием композиционного представления частот аллелей.

**Практическую значимость работы** определяют:

- Программный код модели `popdisp`, предоставленный в открытый доступ на платформе GitHub: [github.com/iganna/popdisp](https://github.com/iganna/popdisp).
- Программный код модели `migadmi`, предоставленный в открытый доступ на платформе GitHub: [github.com/iganna/migadmi](https://github.com/iganna/migadmi).
- Более детальное описание популяционной структуры коллекции нута ВИР, имеющее сельскохозяйственное значение для поддержания генетического разнообразия при разработке новых сортов нута.
- Характеристические частоты аллелей популяций нута, которые могут быть использованы для определения происхождения образцов в рамках селекционных программ.

- Установленное смешанное происхождение некоторых популяций дези, что указывает на существование двух древних, независимых торговых путей из Ближневосточного региона (включая Ливан, Турцию, Южно-Черноморский регион и Плодородный полумесяц) в Эфиопию. Этот результат позволяет выдвигать гипотезы, которые могут быть применены при анализе разнообразия и определении центров доместикиции других культурных растений.

### **Методология и методы исследования**

Программный код методов `popdisp` и `migadmi` выполнен на языке программирования Python. Для оптимизации использовались методы из существующих библиотек: метод `hmc`, осуществляющий Гамильтониан Монте Карло, из библиотеки `ruhmc` и метод `SLSQP`, последовательного квадратичного программирования, из библиотеки `scipy`. Анализ координат точек сбора образцов, а также построение матриц географических расстояний были проведены с помощью инструментов пакетов `leastcostpath`, `elevatr`, `movecost`, `sp` на языке программирования R. Анализ генетической структуры данных был выполнен с помощью программы `ADMIXTURE` (запуск из интерфейса командной строки). Подтверждение результатов `migadmi` было выполнено путем сравнения с результатами работы программ `TreeMix` (запуск из интерфейса командной строки), `MiaMapper` (язык программирования Matlab) и `ADMIXTOOLS` (язык программирования R). Программный код для визуализации был выполнен на языках программирования R и Matlab.

### **Основные результаты и положения, выносимые на защиту:**

1. Разработан и протестирован метод `popdisp` для оценки характеристических частот аллелей в популяциях, отличающийся тем, что использует знание о ковариационной структуре популяции, а также использует анализ композиционных данных для работы с частотами аллелей в популяции.
2. Разработан и протестирован метод `migadmi`, реконструирующий эволюционную историю популяция, отличающийся тем, что позволяет учитывать сложные события смешения (больше двух исходных популяций и множественные вложенные события), а также использует анализ композиционных данных для работы с частотами аллелей в популяции.
3. В результате реконструкции эволюционной истории популяций нута с помощью методов `popdisp` и `migadmi` было показано, происхождение Эфиопской популяции нута дези является смешанным, при этом исходными популяциями выступают Черноморская, Индийская и Ливанская популяции нута дези.
4. Происхождение западной-Среднеазиатской популяции нута дези является смешанным, при этом исходными популяциями выступают Ливанская и восточная-Среднеазиатская популяции нута дези.

5. Происхождение нута кабули относится к региону современной Турции, а не к азиатскому региону, как можно было бы предположить из-за сходства названий 'кабули' и 'Кабул'.
6. Генетические сходства между образцами нута внутри каждой популяции больше соответствует оптимальному перемещению по ареалу с учетом географических особенностей, чем прямолинейному пути распространения.

**Достоверность** полученных результатов для популяций нута подтверждается их согласованностью с ранее опубликованными археологическими находками и не противоречит, а скорее уточняет данные генетического анализа, проведенного на других панелях образцов нута. Несмотря на различия в результатах программ TreeMix, MixMapper и ADMIXTOOLS, программа migadm1 показала результаты, согласующиеся с большинством из них. Надежность работы программ popdisp и migadm1 обеспечивается детальным тестированием на симулированных данных.

**Личный вклад автора** заключался в участии в процессе формулирования цели исследования, а также в определении конкретных задач, необходимых для ее достижения. Автор лично предложил и аналитически построил композиционную аппроксимацию модели Райта-Фишера, разработал и программно реализовал новые методы popdisp и migadm1, разработал схему для симуляции данных и провел тестирование разработанных программных пакетов. Автор самостоятельно провел анализ генетических данных панели нута из коллекции ВИР и определил генетически однородные и географически связанные популяции нута. Автором единолично были составлены матрицы географических расстояний между координатами точек сбора образцов, применены как новые разработанные, так и существующие модели к данным по нуту, а также проанализированы полученные пути происхождения популяций нута. Все рисунки, не заимствованные из опубликованных источников, выполнены автором лично.

### **Публикации и апробация работы**

В рамках темы диссертации было опубликовано 8 статей в научных журналах, индексируемых в Web of Science и Scopus. Результаты работы были представлены на конференциях "Вероятностное моделирование в геномике" в 2024 году (Probabilistic Modeling in Genomics, ProbGen2024, Вена, Австрия) и "Московской международной конференции по вычислительной молекулярной биологии" в 2021 году (Moscow Conference on Computational Molecular Biology, MCCMB'21 – Москва, Россия).

## **Содержание работы**

**Глава 1** представляет обзор литературы по теме диссертации и включает пять разделов.

В **разделе 1.1** рассматривается задача оценки характеристических частот аллелей в популяции, а также обсуждаются проблемы, связанные с учетом географического местоположения образцов при данной оценке.

В **разделе 1.2** обсуждаются методы реконструкции эволюционной истории популяции, способы описания событий смешения в известных подходах, а также возможные направления их усовершенствования.

**Раздел 1.3** посвящен описанию модели Райта-Фишера и критическому анализу двух её аппроксимаций: с использованием бета-распределения и нормального распределения.

**Раздел 1.4** содержит краткое изложение теории анализа композиционных данных, необходимой для разработки новой аппроксимации модели Райта-Фишера, где частоты аллелей в популяции рассматриваются как композиционные данные.

В **разделе 1.5** представлено описание Байесовского подхода в моделировании и его преимущества при работе с большим количеством параметров.

В **Главе 2** представлено описание двух новых разработанных автором моделей - `popdisp` и `migadm1`. Обе модели программно реализованы в виде двух пакетов, которые доступны на платформе Github. Обе программы, `popdisp` и `migadm1`, представляют собой универсальные инструменты, которые могут быть использованы как совместно, так и отдельно. Они могут быть применены к широкому спектру задач, не ограничиваясь только географическим распространением популяций.

В **разделе 2.1** приведён вывод композиционной аппроксимации модели Райта-Фишера, которая используется в обеих моделях. Частоты аллелей в популяциях подчиняются определённым ограничениям: они всегда должны быть неотрицательными, а их сумма для одного генетического локуса в популяции должна равняться единице. Эти ограничения затрудняют использование стандартных методов статистического анализа, так как они часто предполагают нормальное распределение данных, что не верно частот аллелей. Композиционный анализ устраняет эти ограничения, переводя частоты аллелей в балансы, которые могут принимать любые значения на вещественной оси и рассматриваются независимо друг от друга. С учетом композиционной природы данных автором разработана композиционная аппроксимация изменения частот аллелей в популяции Райта-Фишера, основанная на балансах.

Пусть  $f_j^i$  - частота одного аллеля в  $i$ -ом биаллельном локусе и  $j$ -ой популяции,  $f_A^i$  - характеристическая частота этого аллеля в  $i$ -ом локусе,  $V$  - ковариационная матрица, связывающая популяции. Баланс между частотами в каждой популяции представляется собой  $\text{ilr}$ -преобразование, которое для случая биаллельных локусов является просто  $\text{logit}$  функцией:

$$x_j^i = g(f_j^i) = \ln \left( \frac{1 - f_j^i}{f_j^i} \right).$$

В результате преобразований было получено, что для описания дрейфа в пространстве балансов можно использовать следующее распределение, что является композиционной аппроксимацией модели Райта-Фишера:

$$\vec{x}^i \sim mv\mathcal{N}\left(\mathbf{1}_M x_A^i, \frac{s^i V}{(f_A^i(1 - f_A^i))}\right), \quad (1)$$

где  $\vec{x}^i = (x_1^i, x_2^i, \dots, x_M^i)$ ,  $\mathbf{1}_M$  - единичный вектор длины  $M$ ,  $s_i$  - параметр масштаба,  $x_A^i = \log((1 - f_A^i)/f_A^i)$ .

**Раздел 2.2** посвящен модели `popdisp`, которая предназначена для оценки характеристических частот аллелей в популяции с учетом ковариационной матрицы между образцами. Модель `popdisp` представляет собой байесовскую иерархическую модель (Рисунок 2 А.), построенную на предположении, что в каждой точке сбора образцов существует определённая популяция, характеризующаяся частотами аллелей, а наблюдаемые образцы — это выборки из локальной популяции. Частоты аллелей в локальных популяциях предполагаются отклоняющимися от некоторой характеристической частоты аллелей и связанными с ней через заданную ковариационную матрицу. Если ковариационная матрица отражает некоторые пути перемещения по географии, то должна существовать и реальная географическая точка в ареале обитания (не обязательно точка сбора образцов), частоты аллелей в которой соответствуют характеристическим частотам аллелей в популяции. Такую точку можно назвать генетическим центроидом, однако понимание значения местоположения генетического центроида не является тривиальным и однозначным. Местоположение может отражать как центр доместикации предковой популяции, так и точку наибольшего обмена генами, что позволяет предположить его связь с древним торговым центром. Точная интерпретация положения генетического центроида в каждом отдельном случае требует отдельного анализа.

В настоящей работе предлагается использование метода Гамильтониан Монте-Карло для оптимизации параметров байесовской модели `popdisp`. Применение данного метода обусловлено его высокой эффективностью в задачах, требующих оценки сложных апостериорных распределений. Для применения Гамильтониан Монте-Карло необходимо получить аналитические выражения для градиентов апостериорного распределения по каждому из параметров модели. В работе проведены соответствующие математические преобразования.

Модель `popdisp` может быть применена для оценки характеристических частот аллелей в популяциях с учетом географического распределения образцов, т.е. координат точек сбора. На основе координат рассчитывается матрица географических расстояний, которая может представлять собой либо прямолинейные расстояния, либо пути наименьшей стоимости между всеми точками сбора и предполагаемым географическим центроидом популяции. В данной работе предложен метод преобразования такой матрицы расстояний в ковариационную матрицу (Рисунок 2 Б.). Возможность выполнения этого преобразования обусловлена допущением о существовании центроида.

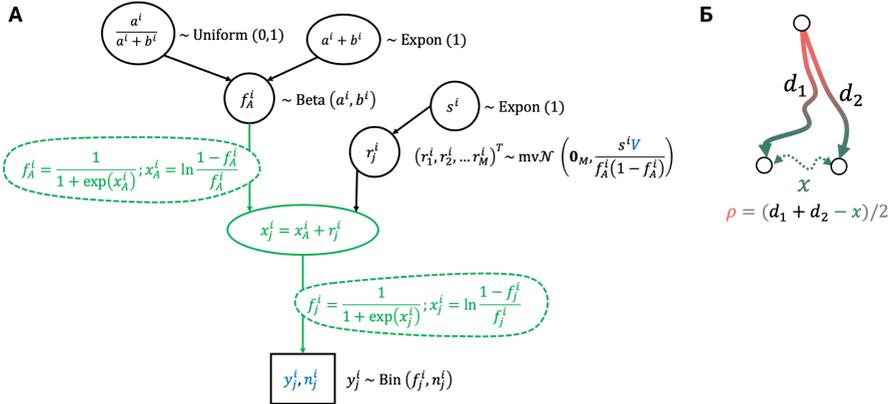


Рис. 2 — **А. Схема иерархической Байесовской модели popdisp для случая биаллельных генетических маркеров.** Наблюдаемые переменные выделены синим цветом. Те из них, которые подчиняются определенному распределению, представлены в прямоугольном блоке, а параметры модели отображены в овальных блоках. Стрелка от одного блока к другому показывает, что переменная в первом блоке является параметром распределения переменной во втором блоке. Зеленым выделены преобразования, связанные с обработкой частот аллелей в популяциях, как композиционных данных.  $n_j^i$  - общее количество аллелей в  $i$ -ом локусе и  $j$ -ой популяции.  $y_j^i$  - количество альтернативных аллелей в  $i$ -ом локусе и  $j$ -ой популяции.  $f_j^i$  - частота альтернативного аллеля в  $i$ -ом локусе и  $j$ -ой популяции.  $f_A^i$  - частота альтернативного аллеля в  $i$ -ом локусе в общей исходной популяции.  $V$  - ковариационная матрица, описывающая накопленный генетический дрейф. **Б. Аппроксимация ковариации.**  $d_1$  - длина пути от положения генетического центроида до первой точки сбора образцов.  $d_2$  - длина пути от положения генетического центроида до второй точки сбора.  $x$  - длина пути между двумя точками сбора.  $\rho$  - аппроксимация длины общего пути от центроида до двух точек сбора. Эта величина предполагается пропорциональной ковариации между двумя точками.

**Раздел 2.3** посвящён модели migadm1, которая принимает на вход частоты аллелей в популяциях (например, оценённые с помощью popdisp) и позволяет реконструировать граф смешений, отражающий миграции и события смешений между этими популяциями. В отличие от существующих подходов, модель migadm1 рассматривает более сложные события смешений, а именно:

1. Симметричная модель смешения, учитывающая наличие собственной дисперсии у всех популяций, участвующих в смешении.

2. Возможность моделирования вложенных событий смешения, когда популяция, являющаяся смешанной в одном событии, выступает популяцией-источником в другом.
3. Допускается участие более чем двух популяций-источников в одном событии смешения.
4. Оценка параметров смешения производится в режиме скользящего окна вдоль хромосом.
5. Оценка вкладов популяций-источников проводится на основании оценок долей дисперсий, а не параметров смешения.

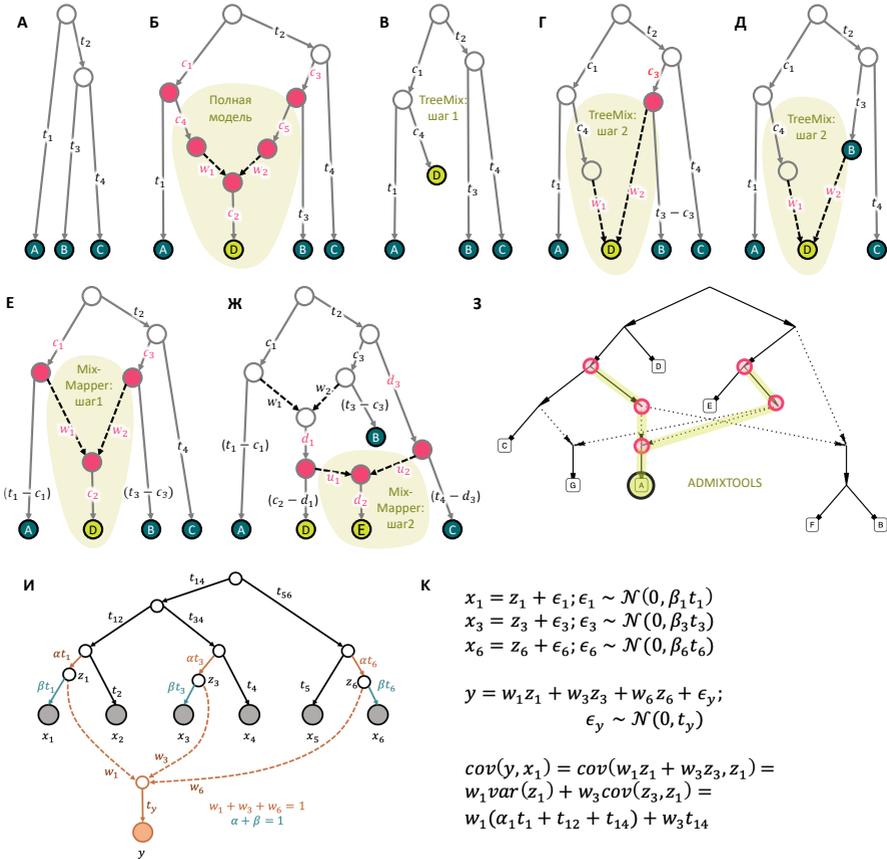
Один из режимов работы программы *migadm1* - жадное построение графа смешений. В этом случае на вход подается набор несмешанных популяций и набор популяций, которые предполагаются смешанными. *Migadm1* перебирает все возможные события смешения по очереди добавляет в модель события по принципу наименьшего значения критерия ВИС. Для параметризации событий смешения *migadm1* использует оригинальный алгоритм, основанный на символьных вычислениях, так что метод оптимизации параметров получает функцию максимального правдоподобия для оптимизации в аналитическом виде.

**Глава 3** посвящена обзору популяционных исследований по нуту и описание исходных данных.

В **разделе 3.1** приведены общие сведения о нуте или бараньем горохе (*Cicer arietinum* L.): описание генпулов, генетические характеристики и особенности опыления. Важно отметить, что у нута *C. arietinum* L. существует две группы сортов: дези и кабули. Эти группы различаются фенотипически: у кабули светлые лепестки и крупные светлые семена, а у дези сиреневые лепестки и меньшие тёмные семена. Нут дези считается первым domesticiрованным нутом, так как напоминает дикий предшественник *C. reticulatum*. Кабули вероятно был отобран из местных дези, а затем распространился, однако точное место domestикации кабули неизвестно.

**Раздел 3.2** содержит изложение крупных популяционных исследований по нуту, которые выявили дополнительные неопределённости в происхождении некоторых его популяций (Varshney et al., 2019, 2021). Исходным материалом для исследований послужили панели из 429 и 3366 образцов соответственно, из коллекции генетических ресурсов ICRISAT (International Crops Research Institute for the Semi-Arid Tropics), которая включала не только местные сорта, но и дикорастущие формы, а также элитные культивары.

На основании анализа  $F_{ST}$  авторы подтверждают регион Плодородного полумесяца как центр первичной domestикации нута, а также выделяют Южную Азию и Эфиопию в качестве вторичных центров. Популяционная структура была определена на основе географического расположения точек сбора образцов, а не генетических данных. Далее, с помощью анализа главных компонент на основе генетических данных, были выдвинуты несколько гипотез о происхождении и миграции популяций (Рисунок 1Б-Ж).



**Рис. 3 — Параметризация событий смешения в различных программах.** **А.** Бинарное дерево связей между популяциями. Синие кружки — наблюдаемые популяции, пустые кружки — предковые. Стрелки показывают эволюционные связи. **Б.** Модель смешения для популяции D: популяции A и B не напрямую вносят вклад в D. **В.** Шаг 1 программы TreeMix: построение бинарного дерева для всех популяций, включая смешанную. **Г.** Шаг 2 программы TreeMix: добавление в бинарное дерево события смешения. **Д.** Результат TreeMix: наблюдаемая популяция вносит прямой вклад в смешанную без учета популяций-предшественников. **Е.** Модель одного смешения в MixMapper. **Ж.** Модель двух событий смешения в MixMapper. **З.** Модель смешения в ADMIXTOOLS, полученная с помощью команды `rand_graph = random_admixturegraph(7, numadmix = 3)`. **И.** Параметризация события смешения, которая используется в модели migadmi. **К.** Пример вывода ковариации между смешанной популяцией и несмешанной популяцией.

В частности были показаны два возможных независимых пути распространения нута из Плодородного полумесяца: (i) из Ближневосточного региона в Южную Азию и (ii) из Ближневосточного региона в Восточную Африку и Средиземноморский регион. Важно отметить, что результаты данного исследования поднимают новые вопросы о том, как нут распространялся из Ближнего Востока в Восточную Африку и Средиземноморский регион. Кроме того, результаты не раскрывают в полной мере эволюционные пути нута, в частности, не исследуют процессы смешения популяций и происхождение нута кабули.

**Раздел 3.3** приводит описание исходных данных для настоящей диссертации. В России существует обширная и независимая коллекция семян, известная как коллекция ВИР имени Н.И. Вавилова. Коллекция насчитывает более 300 000 образцов растений, в том числе 3380 образцов нута, почти половина из которых представляет староместные сорта с известными географическими координатами точек сбора. Эти сорта можно рассматривать как окончательный результат диверсификации культурных линий нута, локально адаптированных к выращиванию в конкретных географических зонах. Из имеющихся образцов нута в коллекции ВИР лабораторией М.Г. Самсоновой была отобрана панель из 421 образца, охватывающая все центры, предположенные Вавиловым, фенотипирована и генотипирована по 2759 однонуклеотидным локусам (Sokolkova et al., 2020).

Отобранная панель достаточно полно описывают разнообразие коллекции и послужила основой для успешного проведения ряда исследований при участии автора настоящей диссертации, таких как: Sokolkova et al., 2020; Igolkina et al., 2020a, 2020b; Kovalev, 2018. Проведенная серия исследований укрепила понимание того, что эти данные являются достаточно надежной основой для проведения серьезного эволюционного анализа.

В **Главе 4** представлены результаты анализа эволюционной истории нута с использованием двух разработанных моделей — popdisr и migadmi.

В **разделе 4.1** детально представлен анализ популяционной структуры нута, проведенный на основе 421 образца коллекции ВИР, который позволил выделить четко выраженные генетически однородные и географически связанные популяции (так называемые генетико-географические группы). На первом этапе с помощью программы ADMIXTURE был проведен анализ генетических кластеров, которые оказались географически компактно локализованы. Образцы без доминирования сигнала одного из кластеров или с несоответствием между географическим расположением и генетической кластеризацией были исключены из дальнейшего анализа. В результате после фильтрации осталось 294 образца, принадлежащие девяти генетико-географическим группам (Таблица 1, Рисунок 4А,Б). Получившиеся группы расширяют понимание разнообразия популяций нута. Так, в Средиземноморском регионе были определены такие популяции, как северная-Средиземноморская и южная-Средиземноморская. Образцы в Ближневосточном регионе были подразделены на Турецкую, Ливанскую и Черноморскую группы. А Среднеазиатская популяция показала две группы: восточную и западную.

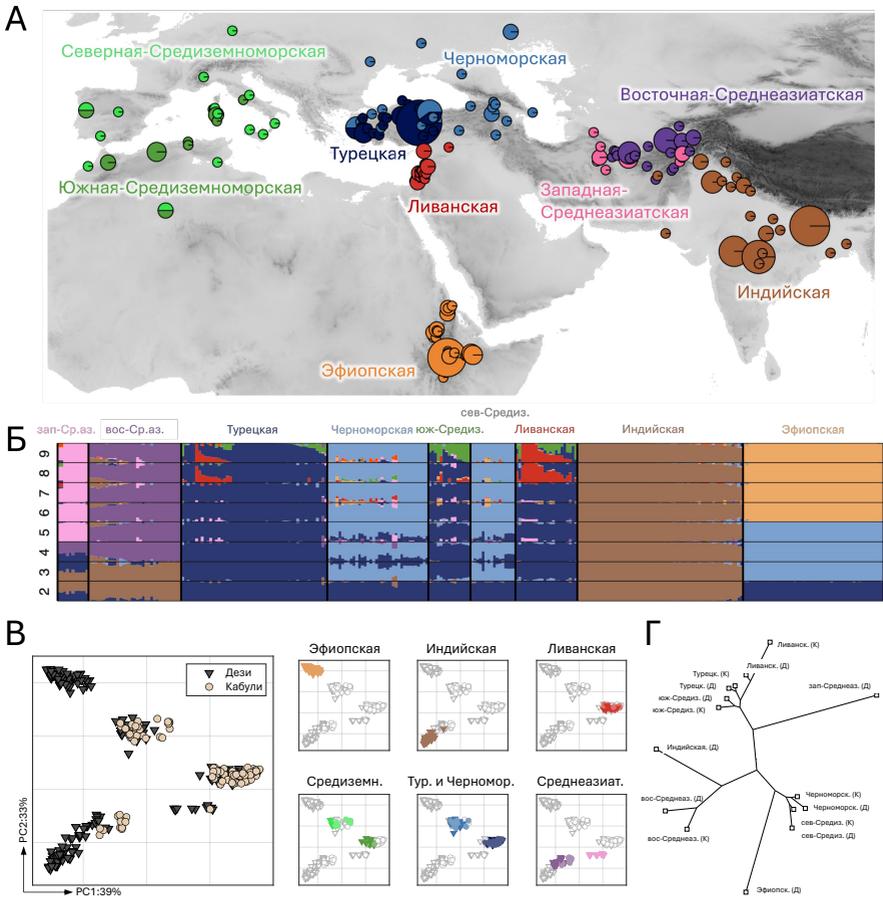


Рис. 4 — Характеристика образцов в выделенных популяциях (генетико-географических группах) после фильтрации. (А) Географическое разделение популяций, каждая из которых обозначена своим цветом. Размер круговой диаграммы показывает относительное количество образцов в точке сбора. (Б) Результаты программы ADMIXTURE для параметра  $K$  от 2 до 9. Результаты демонстрируют генетическую чистоту популяций. (В) Представление образцов в первых двух главных компонентах. (Г) Иерархическая кластеризация популяций на основании значений статистики  $F_{ST}$ .

Таблица 1 — Количество образцов в популяциях до и после фильтрации.

Зеленым цветом подсвечены образцы, которые были взяты для анализа.

Регион	До фильтрации			После фильтрации		
	дези	кабули	неопр.	дези	кабули	остальное
Эфиопия	61	3	0	50	0	14
Индия	64	5	1	57	0	32
Ливан	18	16	0	10	12	12
Турция	40	73	4	13	39	29
Черноморский				18	18	
западно-Среднеаз.	70	25	2	10	0	54
восточно-Среднеаз.				20	13	
северно-Средиземн.	19	19	1	5	11	8
южно-Средиземн.				9	6	

В каждой популяции присутствовали образцы обеих групп сортов нута (дези и кабули), однако первая главная компонента в каждом генетико-географическом кластере показывает различие между этими группами (Рисунок 4В), что подтверждает отсутствие репродуктивной изоляции и раздельное культивирование этих групп. В соответствии с группами полученные генетико-географические популяции были дополнительно разделены (Таблица 1).

Анализ иерархической кластеризации статистики  $F_{ST}$  подтвердил выделенные популяции (Рисунок 1 Г), а две популяции - Эфиопская и западно-Среднеазиатская дези популяции - были идентифицированы как потенциально смешанные.

Ввиду малочисленности средиземноморских популяций и сложности их анализа методом `popdisp`, который требует нетривиальной и неоднозначной оценки путей перемещения по воде и суше, было решено исключить их из дальнейшего исследования. Тем не менее, относительно их происхождения можно сделать следующие выводы. Поскольку две популяции Средиземноморского региона принадлежат к разным группам смешения, где северная популяция объединена с Черноморской, а южная — с Турецкой и Ливанской, можно предположить, что они возникли в результате двух независимых путей распространения нута из Ближнего Востока на запад вдоль северного и южного побережий Средиземного моря соответственно.

**Раздел 4.2** посвящен описанию модели `popdisp`, её тестированию на симулированных данных, и анализу результатов, полученных в ходе её применения к данным о нуте. Программа `popdisp` оценивает характеристические частоты аллелей в популяции, используя генетическую информацию образцов популяции и матрицу ковариаций между точками сбора образцов, рассчитанную, например, на основе географической близости. Модель `popdisp` была программно реализована в одноименном программном пакете на языке Python и выложена в открытый доступ на платформе Github.

Для тестирования программы `popdisp` была разработана схема симуляции данных, моделирующая распространение по неоднородному ландшафту. По

результатам работы на симулированных данных программа `popdisp` продемонстрировала хорошую устойчивость к выбросам, зашумлению и искажениям в данных. Такая устойчивость `popdisp` достигается за счет связности всех образцов в единую структуру с использованием матрицы ковариаций. Важно подчеркнуть, что матрица ковариаций должна в действительности отражать распределение аллелей в популяции, а не быть случайной. Для проверки чувствительности программы `popdisp` к матрицам ковариации для каждой симуляции проводились запуски с разными матрицами, чтобы определить, какая из них дает наиболее достоверные оценки. Оказалось, что лучшие результаты достигаются при использовании исходной матрицы ковариации, использованной в симуляциях.

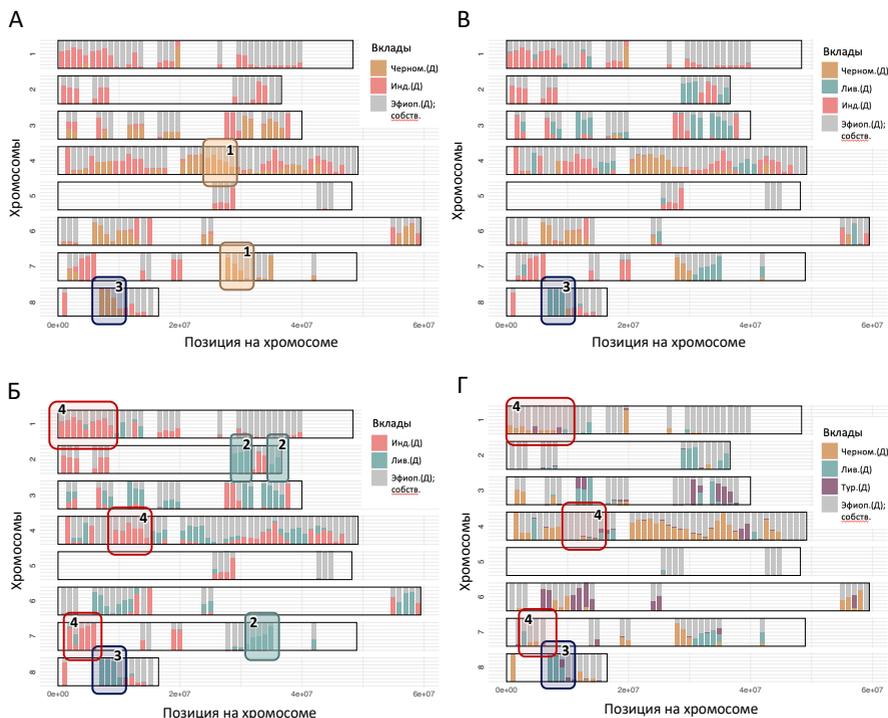
Способность `popdisp` различать матрицы ковариации побудила провести следующий эксперимент на данных по нуту. Для каждой популяции нута были построены две ковариационные матрицы: одна построена на основе линейных расстояний между точками сбора образцов в регионе, а другая — на основе длин оптимальных путей перемещения по ландшафту (функция Герцог). Оказалось, что матрицы расстояний, построенные на оптимальных путях, соответствуют более высокому значению функции правдоподобия модели `popdisp`. Таким образом, генетическое родство образцов в регионах связано не с прямолинейными путями, а с реальной сложностью перемещения по ландшафту, что, вероятно, отражает то, что нут распространялся вместе с человеком. В способности `popdisp` сравнивать географические пути перемещения видится еще одно применение пакета `popdisp`, кроме оценки характеристических частот аллелей.

Результаты работы программы `popdisp` на данных по нуту позволили получить характеристические частоты аллелей в популяциях, которые затем были использованы для реконструкции эволюционной истории этих популяций.

**Раздел 4.3** посвящен модели `migadm1` и включает её тестирование на симулированных данных и анализ эволюционной истории нута, проведённый с использованием этой модели. Программа `migadm1` реконструирует эволюционную историю популяций с возможностью учета сложных событий смешения. Тестирование на симулированных данных, что точность метода `migadm1` ограничена точностью метода оптимизации `SLSQP`, который используется для оптимизации функции цели.

По результатам применения к реальным данным по нуту, программа `migadm1` раскрыла происхождение двух дези популяций (Эфиопской и западно-Среднеазиатской), тем самым указывая на потенциальные исторические торговые пути. Эфиопская дези популяция рассматривалась как смесь двух или трех популяций, и самый правдоподобный вариант соответствовал рассмотрению Эфиопской популяции, как смешанной из Черноморской и Индийской дези популяций. Распределение вкладов популяций-источников вдоль хромосом подтвердило, что Ливанская дези-популяция также вносит свой вклад в Эфиопскую популяцию (Рисунок 5). В результате можно заключить, что существовали три основных торговых пути в Эфиопию. Два из них были морскими: один шел из Индии, а другой, по-видимому, проходил из Черноморского региона, вероятно

через Аравийский полуостров, минуя Ливан. Из Ливана, вероятно, существовал отдельный сухопутный торговый путь в Эфиопию.



**Рис. 5 — Распределение вдоль хромосом вкладов различных популяций в Эфиопскую.** Цифрами обозначены локусы для особого внимания. Эфиопская дези популяция рассматривается как смесь, в которой участвуют следующие дези популяции: (А) Индийская и Черноморская; (Б) Индийская и Ливанская; (В) Индийская, Черноморская и Ливанская; (Г) все Ближневосточные.

Другой пример популяции со смешанным происхождением — это дези популяция из западно-Среднеазиатского региона. По результатам, полученным с помощью программы *trigadm1*, выяснилось, что эта популяция образовалась в результате смешения Ливанской дези популяции и восточно-Среднеазиатской дези популяции. Это смешение можно описать как результат встречного движения в одних и тех же широтах. Стоит отметить, что формирование западно-Среднеазиатской популяции из восточно-Среднеазиатской популяции является как бы движением в обратную сторону относительно первоначального направления при формировании западно-Среднеазиатской популяции.

Особым вопросом в формировании разнообразия нута является происхождение нута кабули. С лингвистической точки зрения, можно предположить, что кабули происходит от названия города Кабул, что говорит о возможном происхождении из восточно-Среднеазиатской популяции, однако археологические доказательства этого не были обнаружены. С помощью программы migadmi были протестированы четыре альтернативные гипотезы происхождения кабули (Рисунок 6), и казалось, что самая правдоподобная из них: происхождение кабули из Турецкого региона, а самая неправдоподобная - из Среднеазиатского.

Для подтверждения этих результатов был проведен анализ главных компонент характеристических частот аллелей всех рассматриваемых популяций. Результаты показали смещение всех популяций кабули вдоль оси первой главной компоненты в сторону Турецких дези, а не в сторону восточно-Среднеазиатских дези (Рисунок 6 Д). Дополнительным подкреплением происхождения кабули популяци из Турецкого региона является анализ хромосомных профилей смешений в популяциях Ливанских и Черноморских кабули в конкурирующих гипотезах. Результаты показали, что хромосомный локус связанный с контрастом дезикабули (фенотип: окраска семян) унаследован от Турецких кабули, а не от восточно-Среднеазиатских кабули (Рисунок 6 Е-Ж).

Таким образом, по результатам полученным с помощью программы migadmi, доместикация нута кабули из местных дези произошла в Турецком регионе, а затем он распространился на другие регионы, включая восточно-Среднеазиатский регион.

**В разделе 4.4** обсуждается сравнение реконструированной истории популяций нута с помощью модели migadmi с работой других программ: TreeMix, MixMapper и Admixtools.

Программа migadmi показала доминирующую согласованность с этими программами: результаты, представленные программой migadmi, согласуются с результатами большинства других программ по всем поставленным в исследовании вопросам: происхождение Эфиопской дези популяции, западно-Среднеазиатской дези популяции и нута кабули. С одной стороны, это свидетельствует о правильности полученных результатов, с другой стороны, это, скорее всего, указывает на более высокую точность параметров смешений.

В отличие от программ TreeMix и MixMapper, модель migadmi напрямую учитывает собственную дисперсию смешанной популяции после события смешения. Преимуществом программы migadmi по сравнению с admixtool является возможно моделировать события смешения с более чем тремя популяциями-источниками.

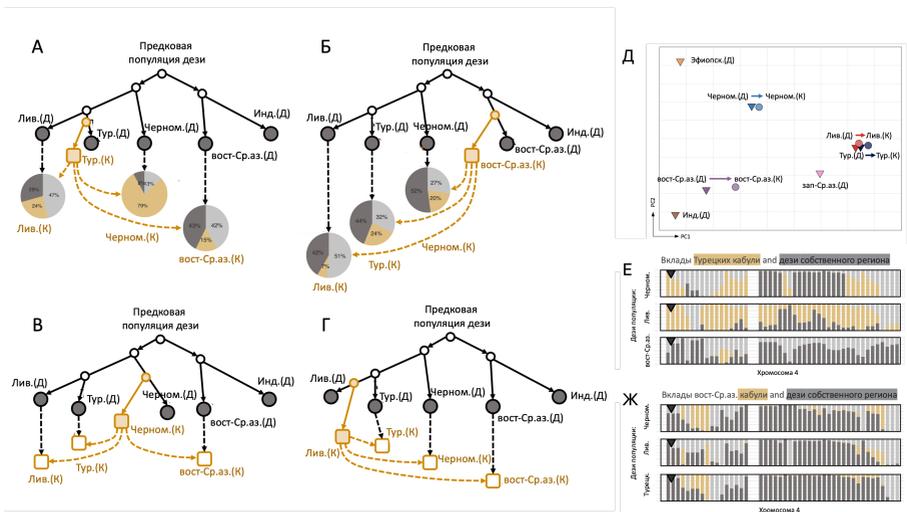


Рис. 6 — Происхождение нута кабули. **А-Б** Четыре альтернативные гипотезы происхождения популяций кабули. Серыми закрашенными кружками обозначены наблюдаемые популяции, серыми незакрашенными — предковые состояния. Желтым закрашенным квадратом показана кабули популяция, впервые возникшая при доместикации. Желтыми незакрашенными квадратами — популяции кабули, образовавшиеся в результате смешения местных дези и первых возникших кабули. Модели предполагают, что доместикация кабули произошла от: **(А)** Турецких дези. **(Б)** Восточно-Среднеазиатских дези. **(В)** Черноморских дези. **(Г)** Ливанских дези. **(Д)** Представление характеристических частот аллелей в осях первых двух главных компонент. Стрелками обозначены направления от популяции дези к популяции кабули в каждом регионе. **(Е)** Распределение вдоль хромосом вкладов популяций Турецких кабули и дези популяций из соответствующего региона. Треугольником обозначено положение маркера, статистически значимо ассоциированного с признаком дези/кабули. **(Ж)** Распределение вдоль хромосом вкладов популяций восточно-Среднеазиатских кабули и дези популяций из соответствующего региона.

## Заключение

Реконструкция эволюционной истории популяций на основе генетических данных представляет собой задачу, богатую нюансами, которые еще не в полной мере учтены в существующих решениях. В настоящей диссертации был предложен двухступенчатый подход к решению этой задачи, учитывающий целый ряд важных факторов. Первый этап заключается в применении новой разработанной программы *popdisp*, которая учитывает географический сигнал при оценке характеристических частот аллелей популяций и способна сравнивать различные гипотезы о перемещениях популяции в ареале. Второй этап заключается в применении новой разработанной программы *migadmi*, которая учитывает события смещения сложной структуры, предлагая новый подход к оценке вкладов популяций-источников в смешанную популяцию, а также демонстрируя распределение этого сигнала вдоль хромосом. В качестве основы для моделирования случайного дрейфа в обеих программах была разработана и использована композиционная аппроксимация модели Райта-Фишера.

Предложенный подход был применен для реконструкции истории популяций нута и позволил ответить на интригующие вопросы о происхождении некоторых из них. В частности, было показано смешанное происхождение Эфиопской и восточно-Среднеазиатской популяций нута дези. Эти результаты позволяют проследить древние торговые пути, в частности, морской путь из Индии в Эфиопию через Аравийское море. Кроме того, настоящая работа разрешает вопрос о происхождении нута кабули. Вопреки лингвистически обоснованной гипотезе, кабули произошли не из Кабула, а из Турецкого региона, что подтверждается не только результатами программы *migadmi*, но и полногеномным анализом ассоциаций.

## Выводы

Выводы по разработке вычислительных средств:

1. Предложенная композиционная аппроксимация модели Райта-Фишера может быть эффективно использована в моделях популяционной генетики.
2. Разработанный метод *popdisp* может быть использован для устойчивой оценки характеристических частот аллелей в популяциях с заданной ковариационной матрицей между образцами.
3. Разработанный метод *migadmi* может быть использован для детальной реконструкции сложных эволюционных истории популяций с множественными и вложенные события смещения.

Выводы, полученные по применению разработанных вычислительных средств:

1. Генетические сходства между образцами в 9 из 11 популяций нута соответствуют длинам оптимальных путей перемещения по ландшафту больше, чем линейным расстояниям между образцами.
2. Северная и южная Средиземноморские популяции нута сформировались в результате продвижения на запад по северному и южному берегам Средиземного моря Черноморской и Турецкой популяций соответственно.
3. Эфиопская популяция нута дези сформировалась в результате смешения Черноморской, Индийской и Ливанской популяций.
4. Западная-Среднеазиатская популяция нута дези возникла в результате смешения Ливанской и восточной-Среднеазиатской популяций.
5. Нут кабули произошел из региона современной Турции, а не из азиатского региона.

### **Список публикаций по теме диссертации**

В рамках темы диссертации было опубликовано 8 статей в научных журналах, индексируемых в Web of Science и Scopus:

1. Igolkina A. A., Noujdina N. V., Vishnyakova M., Longcore T., von Wettberg E., Nuzhdin S. V., Samsonova M. G. Historical routes for diversification of domesticated chickpea inferred from landrace genomics // *Molecular Biology and Evolution*. 2023. Vol. 40, No 6. P. msad110.
2. Igolkina A. A., Meshcheryakov G., Gretsova M. V., Nuzhdin S. V., Samsonova M. G. Multi-trait multi-locus SEM model discriminates SNPs of different effects // *BMC Genomics*. 2020. Vol. 21, No S8. P. 490.
3. Igolkina A. A., Meshcheryakov G. semopy: A Python package for structural equation modeling // *Structural Equation Modeling: A Multidisciplinary Journal*. 2020. Vol. 27, No 6. P. 952–963.
4. Korvigo I., Igolkina A. A., Kichko A. A., Aksenova T., Andronov E. E. Be aware of the allele-specific bias and compositional effects in multi-template PCR // *PeerJ*. 2022. Vol. 10. P. e13888. doi:10.7717/peerj.13888.
5. Sokolkova A., Bulyntsev S. V., Chang P. L., Carrasquilla-Garcia N., Igolkina A. A., Noujdina N. V., von Wettberg E., Vishnyakova M. A., Cook D. R., Nuzhdin S. V., Samsonova M. G. Genomic analysis of Vavilov's historic chickpea landraces reveals footprints of environmental and human selection // *International Journal of Molecular Sciences*. 2020. Vol. 21, No 11. P. 3952.
6. Kovalev M. S., Igolkina A. A., Samsonova M. G., Nuzhdin S. V. A pipeline for classifying deleterious coding mutations in agricultural plants // *Frontiers in Plant Science*. 2018. Vol. 9. P. 1734.
7. Igolkina A. A., Bazykin G. A., Chizhevskaya E. P., Provorov N. A., Andronov E. E. Matching population diversity of rhizobial nodA and legume NFR5 genes in plant-microbe symbiosis // *Ecology and Evolution*. 2019. Vol. 9, No 18. P. 10377-10386. doi:10.1002/ece3.5556.

8. Igolkina A. A., Porozov Yu. B., Chizhevskaya E. P., Andronov E. E. Structural insight into the role of mutual polymorphism and conservatism in the contact zone of the NFR5–K1 heterodimer with the Nod factor // *Frontiers in Plant Science*. 2018. Vol. 9. doi:10.3389/fpls.2018.00344.

Кроме того, результаты работы опубликованы в сборниках тезисов международных и российских конференций:

1. Igolkina A., Samsonova M., Nuzhdin S. Compositional approximation of the Wright-Fisher model in estimating genetic centroids and admixture graphs // *ProbGen2024*, Вена, Австрия, 10-12 апреля 2024.
2. Igolkina A., Noujdina N., Longcore T., von Wettberg E., Samsonova M., Nuzhdin S. Models for crop species diversification within and between centers of domestication // *МССМВ2021*, Москва, Россия, 30 июля - 2 августа 2021.

*Иголкина Анна Андреевна*

Реконструкция эволюционной истории нута с применением моделирования сложных событий смещений и композиционного анализа данных

Автореф. дис. на соискание ученой степени канд. биол. наук

Подписано в печать \_\_\_\_\_.\_\_\_\_.\_\_\_\_\_. Заказ № \_\_\_\_\_

Формат 60×90/16. Усл. печ. л. 1. Тираж 100 экз.

Типография \_\_\_\_\_

