

ОТЗЫВ

Официального оппонента на диссертационную работу Иголкиной Анны Андреевны «Реконструкция эволюционной истории нута с применением моделирования сложных событий смещений и композиционного анализа данных» представленную на соискание ученой степени кандидата биологических наук по специальности 1.5.8 «Математическая биология, биоинформатика»

Актуальность темы

Диссертация Анны Иголкиной посвящена изучению эволюционной истории и доместикации нута на основе данных, полученных при помощи генотипирования секвенированием. При этом существенная часть работы посвящена разработке новых математических подходов, численных методов и программного обеспечения для реконструкции эволюционной истории, которые могут быть использованы для изучения других эукариотических организмов. Таким образом актуальность работы обусловлена двумя факторами. С точки зрения предметной области, нут является одной из наиболее важных сельскохозяйственных культур в мире. Исследование его генетического разнообразия и эволюционной истории важно для понимания формирования механизмов его адаптации и сохранения видового биоразнообразия. С методологической точки зрения, в диссертации предложены новые математические и вычислительные подходы на основе анализа композиционных данных для изучения сложных популяционных сценариев с множественными и вложенными примешиваниями. Также в работе показано, что генетическое сходство между образцами нута внутри каждой популяции соответствуют их географической структуре, учитывающей оптимальные пути перемещения с учетом географических преград.

Структура и содержание диссертации

Диссертация состоит из введения, 4 глав и заключения. Текст содержит 144 страницы, 32 рисунка и 4 таблицы. Список литературы содержит 96 наименований.

В главе 1 дан исчерпывающий обзор методов популяционной генетики, результатов в области анализа композиционных данных, а также представлены сведения по Байесовскому анализу данных, в частности использованию методов МСМС для оценки параметров модели.

В главе 2 построена композиционная аппроксимация модели Райта-Фишера. При помощи ilr -преобразования частоты аллелей преобразуются в балансы. Выведена многомерная нормальная аппроксимация дрейфа в пространстве балансов. Затем представлена иерархическая Байесовская модель `popdisp` для оценки частот аллелей в популяции на

основе построенной ранее композиционной аппроксимации модели Райта-Фишера, что позволяет применять метод для мультиаллельных локусов. Также представлена модификация метода `popdisp` на случай, учитывающий неравновесное сцепление. Также модель `popdisp` может быть информирована о географической структуре с помощью ковариационной матрицы для частот аллелей. Далее представлен метод `migadmi` для восстановления графов перемешиваний (*admixture graphs*) на основе композиционной аппроксимации модели Райта-Фишера, что является принципиальным отличием от другого метода оценки графов перемешиваний `admixtools – qpgraph`. Кроме того, уникальность метода `migadmi` заключается в возможности оценивать паттерны смещений вдоль хромосом.

Глава 3 фактически является обзором литературы к следующей части диссертации – исследованию популяционной истории нута. Глава начинается с представления общих сведений о нуте, затем представлен обзор популяционных исследований нута. Также в главе описан использованный в диссертации датасет, а также полученные ранее на нем результаты.

Глава 4 посвящена изучению популяционной структуры нута, включая применение представленных в главе 2 методов. Также методы `popdisp` и `migadmi` протестированы при помощи компьютерного моделирования. В результате реконструирована популяционная история нута, определены исходные и гибридные популяции, а также показана связь генетических различий с географическими особенностями ареала его произрастания.

Новизна полученных результатов и выводов

В диссертации А. Иголкиной представлена оригинальная аппроксимация модели Райта-Фишера в пространстве балансов (композиционный анализ данных) при помощи многомерного нормального распределения. На основе разработанной математической теории разработаны два метода `popdisp` (для оценки частот аллелей) и `migadmi` (для предсказания графов перемешиваний). Реконструирована популяционная история нута, определены исходные и гибридные популяции, проинтерпретирована связь с географической структурой ареала.

Практическая и теоретическая значимость

Представленные в диссертации результаты значимы как с методологической, так и с предметной точки зрения. Методологически, теория композиционного анализа данных перенесена в область популяционной генетики и использована при разработке двух новых методов популяционного анализа. Эти методы могут быть использованы не только для анализа нута, но также в будущем найдут применение при анализе других эукариотических популяций. С точки зрения предметной области, восстановление популяционной истории может быть важен для дальнейшего изучения механизмов его адаптации к различным

условиям внешней среды, а также для сохранения его генетического разнообразия для минимизации рисков программ селекции.

Обоснованность научных положений и выводов, сформулированных в диссертации

Все результаты обоснованы и достоверны, работы выполнены на высоком научном уровне. В работе дан подробный вывод математической теории композиционного анализа данных, байесовских иерархических моделей, проведена валидация разработанных методов при помощи компьютерного моделирования. Результаты опубликованы в ведущих международных журналах, а также представлены на двух международных конференциях. Программный код доступен в открытом доступе в двух репозиториях GitHub, обеспечивая воспроизводимость результатов.

Замечания и комментарии

- 1) Разделы 3.1 и 3.2 относятся к обзору литературы, стоило бы их разместить в главе 1. В текущем виде обзор литературы описывает методологическое состояние области, но не даёт полного представления о предметной области.
- 2) Не хватает подробной документации представленных в диссертации методов в тексте диссертации или в соответствующих репозиториях GitHub.
- 3) На Рисунках 4.10 и 4.12 хотелось бы иметь точные значения критериев отбора моделей, а не просто указаний, какая модель лучшая, а какая вторая по критериям отбора моделей.
- 4) Раздел 1.4: непонятно, как выбирается ортонормированный базис для Симплекса S^D , чтобы потом можно было интерпретировать результаты анализа в пространстве R^{D-1} (формула 1.13). Как будут меняться результаты анализа при выборе разных базисов?
- 5) Рисунки 2.1, 2.2 – в чем отличие пунктирных линий от сплошных, а также блоков с границей нерегулярной формы? На рисунке 2.3 две стрелки идут «насквозь» блоков с зеленой пунктирной границей – является ли это артефактом или несет смысловую нагрузку?
- 6) В работе присутствует умеренное количество опечаток и неточностей: например, «Байесовский» пишет то с большой, то с маленькой буквы, в конце раздела 2.2.5 написано «что не требуется при оптимизации независимой оптимизации»

Заключение

Диссертационная работа А. А. Иголкиной «Реконструкция эволюционной истории нута с применением моделирования сложных событий смещений и композиционного анализа

данных» выполнена на высоком научном и методическом уровне и соответствует требованиям по актуальности, степени обоснованности научных положений и выводов, их достоверности, практической значимости и новизне, установленным пунктами 9-14 "Положения о присуждении ученых степеней" (утверждено Постановлением Правительства РФ от 24.09.2013 г. № 842 с изменениями Постановлений Правительства РФ от: 21.04.2016 г. № 335; 02.08.2016 г. № 748; от 29.05.2017 г. № 650; 20.03.2021 г. № 426; 11.09.2021 № 1539; 26.09.2022 г. № 1690), предъявляемым к диссертациям на соискание ученой степени кандидата наук. Учитывая наличие достаточного числа публикаций в рецензируемых научных изданиях первого квартиля и научную значимость полученных результатов, автор диссертационной работы заслуживает присуждения степени кандидата биологических наук по специальности 1.5.8 — "Математическая биология, биоинформатика".

Официальный оппонент:



Щур Владимир Львович,

доктор физико-математических наук

(специальность – 1.2.2 – Математическое моделирование, численные методы и комплексы программ),

Заведующий международной лабораторией статистической и вычислительной геномики,
Институт искусственного интеллекта и цифровых наук, Факультет компьютерных наук,
НИУ ВШЭ

Профессор департамента больших данных и информационного поиска, Факультет
компьютерных наук, НИУ ВШЭ

11.09.2025

Контактные данные:

Email vshchur@hse.ru

Телефон 8-916-570-75-96

Почтовый адрес

109028, Покровский бульвар, д. 11, НИУ ВШЭ, Москва, Россия

