

«УТВЕРЖДАЮ»
директор ИЦиГ СО РАН
академик РАН
Кочетов А.В.
« 4 » _____ 2025 г.



ОТЗЫВ

ведущей организации 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.

Современные технологии массового генотипирования сельскохозяйственных растений позволяют получить данные по однонуклеотидным заменам в масштабе полного генома для большого количества образцов растений. Это предоставляет возможность для анализа генетических вариаций по множеству локусов одновременно в рамках целой популяции растений. В результате становится возможным выделить субпопуляции, привязанных к географическим местам произрастания растений, проследить эволюционную историю их формирования, оценить степень смешения между географически разделенными субпопуляциями. Методы для решения подобных задач опираются на модели дрейфа генов Райта-Фишера, которая развивалась в течение десятков лет. Однако у этой модели есть ряд недостатков, которые могут повлиять на результаты. Эти недостатки связаны с упрощенным представлением распределений частот аллелей для мультиаллельных локусов виде нормального, предположением о том, что обмен генетическим материалом для всех локусов генома происходит однообразно, а также отсутствием моделей, описывающих сложными событиями смешения генов в популяции, допускающий участие более чем двух популяций-источников.

Все эти проблемы возникают при изучении полиморфизмов в популяции нута *Cicer arietinum*, сельскохозяйственной культуры, относящейся к бобовым и служащее незаменимым источником высококачественного белка во многих странах мира. Эта культура занимает второе место в мире среди бобовых по производству зерна. Однако изучение генетического разнообразия нута в настоящее время имеет ряд проблем: неполнотой выявления и описания генетико-географические групп и отсутствием надежных данных о взаимодействии этих групп в процессе эволюции нута.

Диссертационная работа Иголкиной А.А. посвящена решению указанных актуальных задач. В методической части работы представлено усовершенствование модели Райта-Фишера с учетом композиционной природы частот аллелей и разработаны несколько моделей для популяционного анализа (модель для оценки характеристических частот аллелей в популяции с учетом ковариационных связей между образцами, модель для реконструкции эволюционной истории популяций, учитывающую события смещения сложной структуры и неравномерность сигнала смещения вдоль хромосом). Эти модели были реализованы в виде программных пакетов, проведено их тестирование в вычислительных экспериментах.

Разработанные в диссертации методы были применены к анализу генетических вариаций для панели образцов нута из коллекции ВИР. В результате были получены важные результаты, касающиеся выделения генетико-географических групп в популяции нута, предложены модели их происхождения.

Новизна полученных результатов и выводов.

Работа Иголкиной А.А. представляет новые результаты по двум направлениям исследований. Во-первых, в работе представлены новые методы биоинформатического анализа полиморфизмов в популяции растений. В частности, впервые модифицирована модель для оценки частот аллелей в популяциях на основе использования теории анализа композиционных данных. В работе выведены аналитические формулы для оценки математического ожидания и дисперсии, на этой основе для описания дрейфа частот аллелей предложена многомерная аппроксимация модели Райта-Фишера в пространстве балансов. Кроме того, в работе предложена иерархическая Байесовская модель для оценки частот аллелей в популяции, которая учитывает их ковариацию между индивидуумами. Выведены аналитические выражения для оценки апостериорных вероятностей в этой модели, а также их производные по параметрам. Предложенный метод был реализован в пакете программ popdisp.

Впервые была разработана модель для реконструкции эволюционной истории популяций, которая включает описание таких событий, как множественные и вложенные смещения субпопуляций, учитывает композиционную природу частот аллелей. Отметим также еще один важный элемент методической новизны в работе Иголкиной А.А. – это возможность оценки популяционных сценариев для отдельных, достаточно больших, участков генома независимо. Это позволяет проследить генетическую изменчивость между субпопуляциями на уровне фрагментов хромосом, выявить участки генома наиболее подверженные вариациям в различных популяциях, вплоть до реконструкции сценариев обмена геномными блоками. Данные методы анализа были реализованы в программе migadmi, которая по сравнению с существующими методами, такими как TreeMix, MixMapper, admixtools, обладает целым рядом преимуществ: может оценивать количество смещений больше двух, число вложенных смещений может быть неограниченно, эволюционные сценарии оцениваются для участков хромосом, модель основана на композиционном анализе частот аллелей.

Методы, разработанные в диссертации Иголкиной А.А. были использованы для анализа популяции нута, что позволило получить ряд новых интересных биологических результатов. Идентифицированы новые генетико-географические популяции нута: Ливанская, Западно-Среднеазиатская, две Средиземноморские популяции. Эти результаты позволяют более точно описать эволюционную историю нута по сравнению с моделями, которые применялись ранее. Результаты анализа также позволили выдвинуть гипотезу о связи между путями распространения нута и генетической близости его субпопуляций.

В диссертационной работе также впервые определено происхождения двух смешанных популяций нута разновидности дези (Эфиопской и Среднеазиатской). Установлено географическое происхождение нута кабули из региона Турции, но не Средней Азии.

Апробация работы и публикации.

В диссертации указано 8 статей в научных журналах, индексируемых в Web of Science и Scopus, в которых были опубликованы материалы по теме диссертации. 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).

Обоснованность научных положений и выводов, сформулированных в диссертации.

Научные положения и выводы, выносимые на защиту, являются обоснованными. Работоспособность новых методов по оценке частот аллелей в популяциях и методов реконструкции их эволюционной истории была оценена в модельных экспериментах. Эти эксперименты проводились на основе моделирования 100 наборов данных, в которых эволюция популяции происходила на основе заданной структуры матрицы ковариации, соответствующей структуре бинарного дерева. Исследовался сценарий, в котором в процессе эволюции популяций происходили события смещения субпопуляций. Результаты показали, что среднее абсолютное отклонение оценок параметров составило менее 0.001, что показало высокую степень идентифицируемости параметров в модели migadmi.

Что касается результатов, полученных программой migadmi в модели эволюции популяции нута, то их сравнение с результатами, полученными с помощью других подходов (TreeMix, MixMapper, admixtools) проводилось при решении задачи о происхождении разновидности нута кабули. Результаты показали в целом согласие на качественном уровне для модели migadmi и других моделей, с учетом того, что migadmi позволяет реконструировать события эволюции нута на более детальном уровне.

Таким образом, положения и выводы диссертации можно считать в достаточной степени обоснованными.

Соответствие содержания диссертации автореферату и указанной специальности.

Диссертация состоит из введения, 4 глав и заключения. Полный объем диссертации составляет 138 страниц, включая 32 рисунка и 4 таблицы. Список литературы содержит 98

наименований. Глава «Общая характеристика работы» предоставляет читателю краткую информацию об актуальности поставленной проблемы, существующих подходах, их достоинствах и недостатках: вначале идет раздел, посвященный биологической проблеме – изучению эволюции и распространению популяции нута. Затем обосновываются задачи по совершенствованию существующих математических моделей описания эволюции частот аллелей в популяции в связи с необходимостью описания сложных событий эволюционной истории. Это позволяет обосновать поставленные в работе цели и сформулировать задачи. Формальные разделы этой части включают описание научной новизны, теоретической и практической значимости, и др. разделы.

Глава «Обзор литературы» посвящена целиком биоинформатическим методам оценки эволюционно-географической истории популяций, разбираются их преимущества и недостатки. Глава 2 посвящена разработке математического аппарата диссертации – моделей для реконструкции эволюционной истории популяций. В ней подробно описаны методические, созданные для выполнения диссертационной работы, программы *popdisp* и *migadmi*. Глава 3 описывает результаты исследования эволюции популяции нута, которые были известны до работы соискателя. Глава 4 посвящена применению разработанных моделей эволюции к исследованию истории популяции нута. В то же время, эта глава включает и описание результатов численного моделирования для оценки работоспособности предложенных методов и разработанных программ. Общий итог исследования подводится в главе «Заключение». Работа включает разделы «Словарь терминов» и «Список сокращений и условных обозначений».

Изложение автореферата соответствует тексту диссертации.

Значимость результатов для науки и производства.

В рамках диссертационной работы Иголкиной А.А. было создано новое программное обеспечение для оценки частот аллелей в популяциях с использованием теории анализа композиционных данных (*popdisp*), также для реконструкции эволюционной истории популяций (*migadmi*). Работоспособность и применимость программных комплексов *popdisp* и *migadmi* была продемонстрирована как на модельных примерах, так и на задаче по определению происхождения двух смешанных популяций нута дези (Эфиопской и западно-Среднеазиатской). С помощью разработанных инструментов также установлено происхождение нута кабули. Описание разработанных методов моделирования и программных комплексов и результатов моделирования было опубликовано высокорейтинговыми научными журналами, что свидетельствует о высоком уровне проведенной работы. В связи с этим можно заключить, что представленные в диссертационной работе методы биоинформатического анализа и компьютерного моделирования могут найти широкое применение как в дальнейших научных исследованиях, так и в практике, в частности, в работах по изучению генетического разнообразия сельскохозяйственных культур.

Замечания.

Несмотря на общую высокую оценку диссертации Иголкиной А.А., необходимо отметить ряд следующих замечаний, которые в основном касаются структуры работы и представления материала.

1. В работе описание биологического объекта (нута и его популяционных исследований) приведено в виде отдельной главы 3, хотя этот материал более логично смотрелся бы в обзоре литературы, который посвящен только методической/биоинформатической части работы.

2. В тексте диссертации встречаются отсылки на другие исследования, в которых принимала участие автор, но результаты которых явно не представлены в тексте как часть диссертационной работы. Например, Глава 3 завершается сообщением о том, что *“Еще одним исследованием, которое было проведено с участием автора настоящей диссертации и применено к генетическим данным панели из 421 образца, была разработка метода детектирования вредных мутаций в белок-кодирующих областях [Kovalev2018]. Модель на основе случайного леса (Random Forest) показала, что, будучи обученной на данных о вредных мутациях Arabidopsis thaliana, она способна с высокой точностью предсказывать вредные мутации у других растений. В результате применения модели к несинонимичным мутациям в данных 421 образцов нута была получена карта вредных мутаций вдоль генома.”* Работа Ковалева упоминается в списке работ по теме диссертации (стр. 14), однако в списке литературы ее нет. Каково отношение этих результатов к теме работы, к её целям и задачам? Интуитивно понятно, что определение полиморфизмов (генетических данных) – необходимая часть предварительной работы для решения поставленных в диссертации задач. Однако формально эта работа не входит в задачи, поставленные автором диссертации.

3. В литературном обзоре, к сожалению, отсутствуют чёткие формулировки и определения многих важных понятий. Яркий пример – термин “популяция”. Некоторые понятия, смысл которых неоднозначен для читателя, вообще не имеют определений. Пример: «географический сигнал в данных» (стр. 21); на стр. 51 термин «эффективный размер популяции» встречается всего один раз и ему не дано определения. Вероятно, подразумевается, что читатель должен быть знаком с этими определениями, однако для строгого изложения результатов, введение таких определений необходимо.

4. В работе представлены два пакета программ анализа генетических и популяционных данных (popdisp и migadmi). Было бы полезно дать описание этих пакетов с указанием типов входных и выходных данных, их внутренней структурой в виде блок-схем.

5. Есть замечания к представлению формул и рисунков: не для всех формул даны расшифровки обозначений (например, первая формула на стр. 30); не для всех графиков (в частности на некоторых из панелей рисунков) подписаны оси (см. рис. 1.5, 4.6, 4.7, 4.8), на рис. 4.5 не расшифрованы цвета.

6. Описание результатов тестирования алгоритмов на симулированных данных (раздел “4.3.1 Тестирование на симуляциях”) дает лишь итоговые оценки точности

реконструкции сценариев. Хотелось бы иметь более детальное описание проведения экспериментов.

7. К сожалению, в диссертации никак не описана процедура “взвешенной оценки” географических путей, поэтому приведённые на стр. 90-91 рассуждения *“Однако, поскольку в Средиземноморских популяциях такие пути проходят частично по воде, получение взвешенной оценки таких путей является нетривиальной задачей, так как требует экспертного анализа относительной сложности перемещения по воде и по суше.”* выглядят не очень понятно и не вполне обоснованно.

8. В разделе 4.3.3 было бы полезно привести значения ВИС для исследованных моделей различных сценариев происхождения популяции нута кабули, чтобы можно было оценить, насколько действительно размер эффекта в различии между моделями позволяет говорить о том, что только одна из них способна адекватно описать имеющиеся данные.

9. В заключении указано: *“Генетические сходства между образцами в 9 из 11 популяций нута соответствуют длинам оптимальных путей перемещения по ландшафту больше, чем линейным расстояниям между образцами”* – вообще, это довольно очевидный вывод. Было бы скорее интересно попытаться рассмотреть, почему в 2 из 11 популяций это не так.

10. Последний пункт в списке практической значимости работы (стр. 12): *“Установленное смешанное происхождение некоторых популяций дези, что указывает на существование двух древних, независимых торговых путей из Ближневосточного региона...”* представляет собой слишком сильное утверждение. Строго говоря, методами биоинформатики (и биологии вообще) нельзя установить наличие или отсутствие торговых путей в древности. На этот вопрос, вероятно, точный ответ могут дать комплексные исследования совокупности данных (генетики, биоинформатики, археологии, истории и др.).

11. В работе допущены опечатки, неудачные формулировки, несогласованные предложения, англицизмы, некорректная калька с английского, перечисление которых заняло бы много места.

Однако, перечисленные замечания редакторского характера в целом не влияют на положительное впечатление от высокого уровня работы, интересных и новых результатов, полученных автором и общую оценку его квалификации как специалиста-математика и биоинформатика, способного решать сложные популяционно-генетические задачи.

Заключение.

Диссертационная работа А. А. Иголкиной 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 Иголкина Анна Андреевна заслуживает присуждения степени кандидата биологических наук по специальности 1.5.8 — "Математическая биология, биоинформатика".

Диссертация заслушана на семинаре Межлабораторного семинара Отделения молекулярной генетики, клеточной биологии и биоинформатики ФГБНУ ФИЦ ИЦиГ СО РАН 25.08.2025.

Сведения о ведущей организации:

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук»

Адрес организации: 630090, Новосибирск, пр-т Академика Лаврентьева, д.10

Телефон: +7 (383) 363-49-63

Электронная почта: icg-adm@bionet.nsc.ru

Сайт: <https://www.icgbio.ru/>

Главный научный сотрудник, и.о. зав. сектором
биоинформатики и информационных технологий в генетике
ИЦиГ СО РАН,
д.б.н. (1.5.8 – математическая биология, биоинформатика),
e-mail lashin@bionet.nsc.ru


С.А. Лашин

Председатель межлабораторного семинара по молекулярной
генетике, клеточной биологии и биоинформатике
ИЦиГ СО РАН,
д.б.н., доцент, e-mail ada@bionet.nsc.ru


Д.А. Афонников

Ученый секретарь ИЦиГ СО РАН, к.б.н.





Г.В. Орлова