

Отзыв официального оппонента доктора биологических наук, заведующего лабораторией эволюционной генетики развития, заместителя директора по научной работе Федерального государственного бюджетного учреждения науки Института биологии развития им. Н. К. Кольцова РАН, Куликова Алексея Михайловича на диссертацию Бакоева Сирождина Юсуфовича «Анализ генетической архитектуры селекционно-ценных признаков свиней с применением современных методов обработки и интерпретации геномных данных», представленную на соискание ученой степени доктора биологических наук по специальности 1.5.7. Генетика (биологические науки)

Актуальность темы исследования

Сто лет назад, в 1922-1925 гг. выходит серия статей Сьюэла Райта «Менделевский анализ чистых пород домашнего скота». Статьи были посвящены проблемам анализа результатов инбридинга, и, по сути, представляли первую систему применения генетических методов в селекционной работе. И вместе с тем эти работы ознаменовали смену парадигмы в представлении о роли генотипа в формировании количественных признаков, в развитии и эволюции. Сто лет развития генетики позволили увидеть «молекулярную кухню» процессов онтогенеза, роста и дифференцировки, поддержания гомеостаза клеток, тканей и всего организма. Геномная, точнее «ОМИКСная» эра предоставила исследователям и селекционерам огромный объем информации, по изменчивости генома и его экспрессии, от уровня клеточной организации до популяций и сообществ. Но осмыслить такой объем информации и представить в виде прогностической модели, позволяющей направленные воздействия на изучаемую систему, оказалось чрезвычайно сложной задачей. Успешные методологические подходы, позволяющие пройти весь путь от анализа геномной изменчивости до определения значимости отдельных аллелей и подтверждения действия на них отбора или демографических процессов, появились буквально в последние годы. Актуальность темы исследования не вызывает сомнения. Обсуждаемая работа Сирождина Юсуфовича Бакоева выполнена в полном соответствии с указанными вызовами, сочетая разработку современных методологических подходов к анализу генетической изменчивости и практическое применение биоинформационных методов для определения молекулярно-генетических маркеров пород и мишеней для селекции селекционно-ценных количественных признаков.

Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации, их достоверность и новизна.

Работа Бакоева Сирождина Юсуфовича представляет последовательный конвейер обработки массивов геномных данных и соответствующих им данных фенотипа, обеспечивающий все этапы генетического анализа, от оценки изменчивости генома до выявления маркерных полиморфизмов, связи конкретных аллелей с признаками и подтверждения действия отбора на соответствующий локус. Но это только методологическая часть работы. Автор представляет огромный объем полученных данных и реальные практические результаты, определяющие породные маркеры и локусы, связанные с хозяйственно-ценными признаками и патологиями. По крайней мере для части этих локусов показано действие отбора и определен период родословной, когда селекция по породным признакам привела к накоплению соответствующей изменчивости в геноме. Все исследования выполнены на репрезентативных выборках животных, на высоком уровне биоинформационного анализа и с учетом всех необходимых контрольных анализов.

Научная новизна работы. Автор провел широкомасштабные исследования нескольких пород свиней и их дикого предка - дикого кабана. Впервые изучено генетическое разнообразие и уровень аутозиготности свиней породы крупная белая, для этой и других исследованных пород выявлены локусы, подвергавшиеся отбору в ходе

селекции. Выявлены новые гены, отвечающие за селекционно-ценные признаки, являющиеся мишенями для дальнейшей селекции свиней. Разработаны и успешно применены новые методологические подходы для анализа огромных массивов молекулярно-генетических данных и определения их связи с разнообразными количественными признаками, в том числе хозяйственно-ценными.

Значимость выводов и рекомендаций, полученных в диссертации, для науки и практики. Предложенный подход применим к любым видам, в том числе к модельным, включая нематод, дрозофил, мышей, крыс, данио-рерио, человека и др. Практически для всех них разработаны ДНК-микрочипы, предназначенные для выявления SNP-полиморфизмов, что на порядок дешевле полногеномного секвенирования и делает возможным фундаментальные исследования в области эволюционной биологии. Практическое значение данной работы также неоспоримо. Автор показал возможности создания дешевых в использовании панелей маркеров породных признаков, молекулярных маркеров хозяйственно-ценных признаков и маркеров наследственных заболеваний, предложил эффективный метод оценки рисков инбридинга. Полученные результаты предполагают широкое применение в селекционной практике. Все данные являются оригинальными и новыми. Впервые изучено генетическое разнообразие свиней крупной белой породы определены целевые локусы и набор генов селекционно-значимых признаков. Предложены новые методы анализа генетической архитектуры породы.

Оценка содержания диссертации

Диссертация занимает 276 страниц основного текста и 46 страниц приложений, включающих 4 оригинальных скрипта для обработки данных и 6 патентов и авторских свидетельств, включает 385 литературных источников, проиллюстрирована 79 рисунками и 21 таблицей. Структура диссертации представлена следующими разделами: введение, обзор литературы, материалы и методы исследования, результаты и обсуждение (общим разделом), заключение, список использованных источников и приложения.

Раздел «Введение» занимает 12 страниц и содержит все необходимые подразделы, включая актуальность темы, обоснование выбора объекта и цели исследования, постановку цели и задач исследования, новизну и значимость, положения, выносимые на защиту, публикации, личный вклад автора и др. формальные пункты. Обозначив цель работы как анализ генетической структуры пород свиней крупная белая, ландрас и дюрок и выявление ключевых генетических вариантов, определяющих селекционно-ценные признаки, Автор определяет семь задач для достижения поставленной цели. Это (1) анализ генетической изменчивости пород в сравнении с изменчивостью общего предка пород – популяций дикого кабана; (2) сравнение генетической изменчивости и уровня аутозиготности свиней породы крупная белая от разных центров селекции; (3) определение подписи отбора в геномах этой породы и аннотация выявленных локусов; (4) полногеномный ассоциативный анализ (GWAS) признака «количество поросят в приплоде»; (5) на основе результатов GWAS и анализа подписей отбора определение статистически значимых предикторов признака «количество поросят при рождении»; (6) оценка влияния вариаций числа копий (CNV) на изменчивость селекционно-значимых признаков; (7) разработка алгоритма визуализации и классификации областей гомозиготности (ROH) на основе данных полногеномного генотипирования и определение их связи с полигенными признаками.

Обзор литературы занимает 40 страниц и состоит из восьми разделов. Состав разделов Обзора в целом отражает перечень поставленных Автором задач. Первый раздел (1.1) посвящен проблемам domestikации сельскохозяйственных животных. Сирождин Юсуфович рассматривает условия успешной domestikации и причины, препятствующие ей. Кратко изложена история domestikации свиней, в первую очередь европейских и отечественных пород.

Раздел «Подписи отбора» (1.2) посвящен методам оценки действия отбора на уровне последовательности ДНК. Автор описывает методы сканирования генома на наличие

районов (гаплотипов) с низкой генетической изменчивостью и высоким неравновесием по сцеплению. Как наиболее эффективные, Автор отмечает методы iHS (integrated haplotype score), nSI (number of segregating sites by length), iHH12 (integrated haplotype homozygosity pooled), XP-EHH (cross-population extended haplotype homozygosity). В разделе также приводятся наборы генов, отбор по которым был показан для разных пород свиней и при сравнении геномов свиней с нормальным фенотипом и с дефектами конечностей.

В разделах 1.3. и 1.4 Сирождин Юсуфович уделяет внимание исследованиям областей ROH (областей гомозиготности), позволяющих выявить степень инбридинга и определить демографическую историю популяции. Автор отмечает классификацию этих областей по происхождению (IBD и IBS), связь размеров идентичных ROH у родственных популяций со временем дивергенции от общего предка, описывает наиболее популярные методы определения ROH.

Раздел 1.5 посвящен анализу связи вариации числа копий (CNV – copy number variegation) с фенотипом. Показаны методы обнаружения CNV на основе данных чипа SNP и GWAS анализ связи CNV с признаками фенотипа. Собственно кросс-геномному поиску ассоциаций генетической изменчивости с количественными признаками (GWAS) посвящен раздел 1.6. Рассмотрены наиболее популярные методы редукции анализируемых генетических вариантов (SNP), включая регрессию Лассо, Ридж-регрессию.

Алгоритмы машинного обучения, такие как Случайный лес (Random Forest), Машина градиентного спуска (Gradient Boosting Machine (GBM) и другие, рассмотрены в разделе 1.7. Подробно обсуждается наиболее эффективная для задач прогнозирования связи генотип–фенотип модель экстремального градиентного усиления (XGBoost). Приведены примеры успешного использования алгоритмов обсуждаемых моделей для идентификации SNP, связанных с хозяйственно-ценными признаками крупного рогатого скота. Последний раздел 1.8 посвящен искусственным нейронным сетям. Обсуждая модели глубокого обучения (DL), Автор особое внимание уделяет сверточным нейронным сетям (CNN), позволяющим осуществлять анализ и классификацию сложных изображений, от различных типов клеток и клеточных органелл до фенотипа целого организма. Преобразование генетических данных в формальное искусственное изображение позволяет выявлять связи между элементами генома, упускаемые традиционными статистическими методами. В целом приведенный «Обзор литературы» хорошо описывает современный уровень работ, посвященных анализу генетических основ селекции сельскохозяйственных животных, определяет наиболее эффективные алгоритмы обработки массивов генетической информации.

Глава «Материалы и методы» изложена на 10 страницах и состоит из 3-х основных разделов – **Материалы**, **Подготовка данных**, **Методы**. В первом разделе (2.1) Автор приводит характеристики выборок свиней, используемых для решения поставленных задач, а также приводит характеристики биочипов GeneSeek® GGP PorcineHD Genomic Profiler v1 (Illumina Inc., США). Здесь-же кратко дается информация по использованию программ на этапе подготовки и анализа полученных данных. Второй раздел (2.2) целиком посвящен подготовке данных, в том числе фильтрации уникальных данных, полученных при сравнении геномов разных популяций, и задаче перекодировки новых массивов данных, полученных на предыдущем этапе анализа, в форматы, пригодные для очередного этапа.

Последний раздел «Методы» разделен на пять подразделов: 2.3.1 Анализ структуры, 2.3.2 Анализ генетического разнообразия, 2.3.3 Анализ выбора, 2.3.4 Демографический анализ, 2.3.5 Определение отбора с помощью статистики гаплотипа, и 2.3.6 Вызов CNV. Все статистические оценки выполнены в среде R. В первом разделе приведены методы шкалирования образцов, соответствующие им метрики дистанций, а также стандартные показатели степени родства. Второй раздел посвящен показателям генетического разнообразия исследуемых групп, в том числе оценкам неравновесия по сцеплению, равновесия Харди-Вайнбега, спектра частот сайтов (SFS), гетерозиготности и уровня генетической дивергенции между популяциями (D Таджики). В третьем разделе указан

метод GWDS анализа выбросов, использующий точные тесты Фишера для выявления случаев недавнего отбора по SNP. В четвертом разделе Сирождин Юсуфович перечисляет методы оценки текущего и исторических эффективных размеров исследуемых популяций и скорости миграции между популяциями. В пятом разделе описаны метод фазирования генотипов и статистические показатели структуры и разнообразия гаплотипов, находящихся под действием отбора. Значения аллель-специфической расширенной гаплотипической гомозиготности (EHH), интегрированной гомозиготности гаплотипа (iHS), кросс-популяционной расширенной гаплотипической гомозиготности (XP-EHH) и числа сегрегирующих сайтов по длине фрагмента (nSL) определяли в программе Selscan. В шестом разделе описаны методы идентификации повторяющихся мотивов CNV, объединения CNV в регионы и сопоставления CNV между образцами.

Результаты и обсуждение представлены общей главой. В разделе 3.1 представлены оценки генетического разнообразия трех изученных пород и диких кабанов. Показаны хорошая дифференцируемость пород по молекулярным маркерам и наличие генетических примесей во всех группах. Относительный объем примесей зависит от количества учитываемых кластеров гаплотипической изменчивости. D-статистика Паттерсона показала, что наибольшее генетическое сходство с европейскими кабанов имеют свиньи крупной белой породы. Все три породы имеют относительно низкие уровни генетической изменчивости, по сравнению с популяциями дикого кабана. Сирождин Юсуфович также отмечает более высокий уровень гетерозиготности, характерный для популяций дикого кабана, хотя на рисунке 3, иллюстрирующим этот вывод, такая зависимость не очевидна, облако значений образцов для дикого кабана должно быть выше. Анализ паттернов распределения SNP неслучаен у всех четырех изученных групп. У коммерческих пород наблюдается уменьшение количества регионов с низкой плотностью SNP и увеличение – с высокой плотностью.

Скорость распада неравновесия по сцеплению принимает наибольшие значения у дикого кабана, что свидетельствует о значительно более высоком генетическом разнообразии в природных популяциях и возможной сложной структуре популяций, предполагающей малую численность субпопуляций, значительный поток генов при миграции и генетический дрейф в субпопуляциях, по сравнению с популяциями коммерческих пород. Показатели сниженной взаимной информации (RMI) со значениями выше 0.45 выявили набор неслучайно ассоциированных локусов в геномах коммерческих пород, предполагая их участие в отборе по хозяйственно-ценным признакам.

Анализ коэффициентов инбридинга показал сходство средних показателей для всех 4-х групп, и максимальный разброс значений коэффициентов для дикого кабана, превышающий дисперсию этих показателей у коммерческих пород в два раза. Оценка распределения коэффициентов инбридинга позволила сделать вывод о необходимости корректировки стратегии внутривидовой селекции для сохранения необходимого уровня изменчивости и избегания инбридинга для пород Ландрас и Дюрок. Для дикого кабана наличие коэффициентов инбридинга 0.4-0.5 говорит о длительной изоляции некоторых групп животных и существенном инбридинге в таких группах. Проведен анализ распределения длин сегментов со статусом «идентичности по происхождению» (HBD), определяющих степень инбридинга и принадлежащих классам k (число поколений от общего предка) с математическим ожиданием длины $1/R_k$ морганид, где $R \approx 2$. Параметр R_k описывает скорость распада HBD сегментов данного класса. Для коммерческих пород показана низкая аутозиготность для современных классов $k=1-2$ (поколений) и исторических классов $k>64$, и высокая аутозиготность для классов $k=4-64$, что в последнем случае соответствует периоду интенсивной селекции породы. Для дикого кабана высокая аутозиготность наблюдается уже с класса $k=2$ и почти постоянно, включая исторические классы $k=128-256$, что свидетельствует о существенной изоляции отдельных групп и фрагментации ареала. Анализируя кумулятивную аутозиготность, Сирождин Юсуфович заключает, что в популяциях коммерческих пород высокий суммарный вклад в

поддерживаемый инбридинг вносят классы с параметрами R_k 8-128 (с некоторыми вариациями для разных пород), то есть сохраняющие сегменты HBD от общих предков 4-64 поколения назад. При этом индивидуальные тренды кумулятивной аутозиготности в целом хорошо следуют усредненным значениям. Дикий кабан демонстрирует более-менее монотонное увеличение средней аутозиготности с класса k_2 до класса k_{32} ($R_k=64$), очень слабое влияние вплоть до k_{128} ($R_k=256$) и высокое влияние аутозиготности от общих предков k_{256} ($R_k=512$). Кроме этого, индивидуальные тренды заметно разделяются на две группы – выше и ниже средней линии. Вместе эти данные предполагают, что часть образцов собрана в изолированных популяциях, имеющих высокий уровень инбридинга уже как минимум 250 лет.

Анализ распределения сегментов HDB по хромосомам показал существенные различия во всех 4-х группах. Тем не менее, общие черты распределения для коммерческих пород заключаются в наибольшей представленности таких сегментов класса k_1 на X-хромосоме (у породы Ландрас значительно обогащены также 2-я, 9-я и 13 хромосомы), класса k_2 на 1-й (X), 2-й, 14-й и 15-й хромосомах, и присутствие на всех хромосомах начиная с класса k_4 ($R_k=8$). У дикого кабана сегменты HDB класса k_1 присутствуют на 13 из 18 хромосомах, и уже в классе k_2 – на всех хромосомах. Проведен анализ распределения SNP по сегментам HDB. Сирождин Юсуфович ранжировал HBD по длине и количеству SNP и определил топ-сегменты для трех пород и дикого кабана. По списку генов, входящих в состав ключевых сегментов, проведен анализ обогащения локусами количественных признаков. QTL определяющие признаки мяса и туши во всех случаях являются наиболее представительными. У породы дюрок они занимают более 80% от всех выявленных QTL, тогда как у других пород и дикого кабана они занимают 60%. У породы крупная белая второе место занимают признаки, определяющие состояние здоровья, у породы ландрас – продуктивности, у дикого кабана – продуктивности и здоровья. Показанные закономерности связаны с направлениями селекции соответствующих пород и процессом отбора у диких животных.

В разделе 3.2. Сирождин Юсуфович приводит анализ внутривидовой изменчивости свиней породы крупная белая. Сравнение генетической изменчивости трех независимых групп селекции методами главных компонент (PCA) и структуры изменчивости (admixture analysis) этих групп показало хорошую дифференцируемость по молекулярным маркерам и наличие возможной примеси в составе генома у двух групп данной породы. Распределение коэффициентов инбридинга в выборках показало сходство для групп 1 и 3 в диапазоне значений 0.2-0.3, и более высокие значения коэффициента у группы 2, в диапазоне 0.25-0.35. Представляя распределение HBD сегментов по классам, Автор показывает некоторые отличия между группами. Так, исторические классы сохранившейся предковой изменчивости важны для всех групп, но для группы 1 это класс k_{128} ($R_k=256$), а для групп 2 и 3 – k_{64} ($R_k=128$). В группе 3 медианное значение сегментов HDB класса k_{32} ($R_k=64$) принимает низкие значения, хотя оценки по верхним квантилям также на уровне оценок по 3-му квартилю групп 1 и 2. Для участков гомозиготности ROH, принадлежащих современным и историческим классам HDB, были определены значения неравновесия по сцеплению LD. Учитывая значительные изменения современного и исторического эффективных размеров популяции, Автор показывает резкое снижение LD для более поздних сегментов ROH и различия в интенсивности снижения для разных селекционных линий. Вклад предковых поколений в уровень аутозиготности сходен для всех трех линий и становится незначимым с 128 поколения ($R_k=256$). Интересно, что как в случае межпородных различий, независимая селекция в отдельных линиях крупной белой привела к заметной генетической дифференциации линий. Доля QTL, определяющая признаки мяса и туши, меняется в линиях от 54,5% до 62,5%, но существенно меняются вклады QTL, определяющие экспрессионную активность генов, контролирующих признаки мяса и туши, и QTL, ответственных за здоровье и репродукцию. Сирождин Юсуфович приводит данные по функциональной активности генов, локализованных в топ-HBD сегментах и заключает,

что у свиней из различных селекционных центров генетические отличия связаны с различными целями и направлением отбора в этих центрах. Выявленные участки генома перспективны для поиска генов, контролирующих хозяйственно-ценные признаки.

Раздел 3.3 посвящен поиску подписей отбора в геномах свиней крупной белой породы. Сирождин Юсуфович приводит результаты анализа геномов трех селекционных линий свиней породы крупная белая, выявляющего по значениям показателей iHS , nSI и $iH12$ локусы, подвергшиеся отбору. Полученные наборы локусов проанализированы на обогащение QTL, связанных с хозяйственно-ценными признаками. Распределение категорий QTL, определяющих хозяйственно-ценные признаки, обладает общими чертами для всех трех линий селекции. Доля QTL признаков мяса и туши всегда занимает порядка 60% от всех выявленных, на втором месте представлены QTL признаков здоровья, на третьем – признаков продуктивности. Интересно, что при этом тщательный анализ состава QTL показывает существенные различия между линиями. Для линии 1 это QTL признаков жиротложения и набора веса тела, качества мяса и показателей крови, расположенных на хромосомах 2, 8 и 14. Для линии 2 это QTL признаков иммунного ответа, качества мяса, метаболизма жиров и влагоудерживающей способности мяса, расположенные на хромосомах 1, 3, 4, 7 и 14. Для линии 3 это QTL признаков показателей крови, устойчивости к заболеваниям и жиротложения, расположенные на хромосомах 2, 4 и 12. Сопоставив общие сигналы отбора в трех линиях, Сирождин Юсуфович перечисляет 10 локусов в составе хромосом 4, 7, 8 и 14. Четыре из представленных локусов находятся в составе длинных некодирующих РНК и 2 – в составе проаннотированных генов. Автор дает подробную функциональную характеристику обоим генам, подтверждая их роль в формировании селекционно-ценных признаков. Проанализировав 53 варианта подписи отбора у свиней импортной селекции, Автор отмечает 31 вариант, локализованный в 26 генах в составе хромосом 1, 3, 4, 5, 7, 8, 9, 13, 14, 17 и приводит подробные функциональные характеристики 24-х из них. Автор отмечает, что отбор по выявленным генам связан с адаптацией к условиям кормления, содержания и требованиям к продуктивным показателям, а сами гены являются маркерами селекционно-ценных признаков.

Раздел 3.4 посвящен анализу предикторов высокой плодовитости свиноматок. Сирождин Юсуфович обсуждает применение метода GWAS для анализа ассоциаций количественных признаков и генетической изменчивости, способы фильтрации данных, разделения данных на обучающие и тестовые выборки, выбор и оценку аналитических моделей. Используя информационный критерий Акаике, Автор определяет распределение количественных показателей как распределение Вейбулла, и использует синтезатор *CornuGAN* для генерации выборки синтетических данных размером 5000 образцов, включающих данные генотипа и фенотипа, на основе экспериментальной выборки. Поиск генетических вариантов, ассоциированных с признаком плодовитости, проведен четырьмя алгоритмами машинного обучения (*Deep Learning*, *RF*, *GBM*, *XGBoost*) и двумя моделями с регуляризаторами (балансом между недообучением и переобучением) на основе гребневой и Лассо-регрессий. На основе информационных критериев *MAE*, *MSE*, *RMSE*, *R2* и *RMSLE*, Сирождин Юсуфович определяет *Deep Learning* и гребневую регрессию как модели, обладающие максимальной способностью идентифицировать целевые SNP. Генетические варианты, превышающие квантиль 0.99, были проанализированы на обогащение сигнальными и метаболическими путями. Наиболее значимыми при использовании обеих моделей выявлены пути деградации лизина, сигнальные пути *TGF-beta* и регуляции плюрипотентности стволовых клеток. Алгоритм гребневой регрессии дополнительно выявил высокую значимость сигнальных путей взаимодействия цитокинов и цитокиновых рецепторов. Отмечая высокое совпадение результатов обеих моделей, и также совпадение с результатами оценки подписей отбора, выявленных для данной популяции, Автор выделяет 18 генов, имеющих критически важное значение для признака плодовитости. Функциональный анализ этих генов убедительно подтверждает их

значимость в поддержании функциональной активности яичников и контроле процессов имплантации эмбриона, формирования плаценты и стадий эмбриогенеза.

В разделе 3.5 приведен анализ повторяющихся последовательностей CNV как предикторов, ассоциированных с селекционно-значимыми признаками свиней. Использованы три алгоритма определения CNV, и в анализ включены варианты, подтвержденные как минимум двумя алгоритмами и имеющие размер более 50 kb. По результатам анализа построена полногеномная карта CNV, сопоставлены локализация CNV и QTL количественных селекционно-ценных признаков и проведен полногеномный ассоциативный анализ CNV-GWAS с признаками скороспелость, толщина шпика, площадь постного мяса и с признаком дефекта конечностей свиней. По результатам анализа выявлено 29 CNV, ассоциированных с дефектами конечностей, без учета поправки FDR, и только 3 CNV, с учетом этой поправки. Признак процент мясности показал значимую связь с 10 CNV без учета поправки FDR, и с одним CNV, с учетом этой поправки, признак толщины шпика – с 32 CNV без учета поправки FDR, применение поправки не оставило значимых CNV, и признак скороспелости - 11 CNV и одним CNV соответственно. Проанализировав функциональную значимость генов, охваченных выявленными областями CNV, Автор заключает, что идентифицированные CNV могут выступать в качестве потенциальных предикторов фенотипов признаков продуктивности и дефектов конечностей свиней.

Раздел 3.6 посвящен описанию алгоритма визуализации и классификации областей гомозиготности (ROH). Эта задача связана с поиском эффективного способа определения общей генетической архитектуры изменчивости у видов, пород и селекционных линий. Такая архитектура выявляется на основании анализа выборки индивидуальных карт ROH, и эффективным способом анализа считается применение сверточных нейронных сетей CNN, применяемых для анализа изображений, в данном случае стандартизованных карт ROH. Для выявления ROH Сирождин Юсуфович использовал программу plink, с учетом фильтров настройки по количеству и плотности SNP, длине фрагмента и дистанции между фрагментами, а также количеству допустимых нарушений гомозиготности при выявлении ROH и по всей длине фрагмента. Анализ проведен в четырех вариантах – со строгим или мягким фильтром по допустимому количеству гетерозигот при сканировании последовательности и в выявленных сегментах, и одновременно с выполнением условия наличия или удаления неравновесных по сцеплению SNP. На основании полученных данных Автор приходит к выводу о необходимости анализа полных данных, без удаления LD SNP, но с применением строгих фильтров по гетерозиготам в составе выявляемых ROH. На полученных таким образом картах ROH приведены выявленные сегменты в соответствии с их длиной и позицией на каждой хромосоме, и с цветовой дифференцировкой, соответствующей размеру сегмента.

Обсуждая CNN модель межпородной дифференцировки, Сирождин Юсуфович описывает компоненты модели, метрики модели, статистическую точность модели. По результатам анализа индивидуальных карт ROH свиней крупной белой породы и ландрас Автор делает вывод о безошибочной работе модели и 100%-ному определению породы свиней по геномному ландшафту сегментов ROH.

Анализ прогностической модели свиней с наследуемой патологией «шишки на ногах» показал значительно более сложную зависимость генетической архитектуры и фенотипа. Учитывая, что патологический фенотип может быть получен множеством сочетаний разнообразных метаболических, иммунных и воспалительных процессов, модель CNN на основе сырых данных не выявила надежных маркеров генетической архитектуры у больных животных. Эффективным оказалось применение метода топологического анализа данных, более устойчивого для многомерных нелинейно организованных данных и к шуму, представленному случайными и малозначимыми объектами. Сирождин Юсуфович использовал метод персистентной гомологии для выявления устойчивых компонент связности (программный пакет `ripserr`). Персистентные

диаграммы были вновь обработаны в модели CNN и были получены результаты с точностью 78.6%, что превышает уровень случайного угадывания. Более того, модель устойчиво предсказывала животных без дефектов с вероятностью 89%. Анализ усредненных областей важности на картах ROH показал возможность выявления разницы в областях гомозиготности у больных и здоровых животных и на основании этого выявление генов, ответственных за заболевание.

Выводы диссертационной работы соответствуют представленным задачам. В разделе Приложения представлены скрипты для для расчета и визуализации CNV, для расчета ROH и визуализации для модели, для расчета HBD и для графиков, а также два патента на изобретения «Способ диагностики предрасположенности свиней породы дюрок к бурситам на основе вариантов SNP rs81327419 (SSC13) в гене ROBO2» и «Способ оценки плодовитости свиней пород ландрас и крупная белая», и три свидетельства о регистрации программ для ЭВМ «База данных аутосомных ДНК-маркеров свиней», «Оценка информативности ДНК-маркеров. Метод Шенона», «Оценка моделей прогнозирования племенной ценности свиней».

Принципиальных замечаний нет, есть технические замечания. (1) В разделе «Материалы и методы» представлен ряд методов, например, скорость миграции, вероятности байесовского распределения популяций и др., используемые в составе пайплайнов при вычислении финальных показателей, таких как доля примесей в геномах пород или демографические показатели в ряду предшествующих поколений. Следовало указать, что самостоятельные результаты использования этих алгоритмов не представлены, а результат их действия используется для оценки соответствующих показателей. (2) Не все сокращения и аббревиатуры расшифрованы в соответствующем разделе в начале диссертации. В том числе R_k – скорость распада HBD-сегментов, N_e – эффективный размер популяции, показатели или подписи отбора по гаплотипам (EHH, XP-EHH, iHS, pSI и iH12), хотя в последнем случае в тексте в разделе «Материалы и методы» приводится их расшифровка. (3) На рисунках 20-23 формально не указана размерность шкалы оси ординат (в мегабазах). (4) На рисунке 34 несогласованность количества кластеров на самом графике и на подрисуночной подписи. (5) Рисунок 41 «Изменение эффективного размера популяции в группах» требует более подробного пояснения в тексте – в чем заключается более выраженная нестабильность динамики изменения N_e в группе КБЗ? (6) Неудачное название раздела 2.3.6 – «Вызов CNV» - калька с выражения «CNV calling», лучше «Выявление и интерпретация CNV». (7) К таблице 1 в подписи расшифровать значения ABBA-BAVA статистик. (8) В таблице 3 названия столбцов соответствуют классам R_k , но форма, в которой приведены классы (R_4 , $R_8...R_{512}$) путает, предполагая, что цифры соответствуют поколениям. Замечания, указанные выше, имеют редакторский или дискуссионный характер. Можно отметить также наличие опечаток и недостатков редакторской правки – в некоторых местах пропущены слова или предлоги.

Представленная работа выполнена на солидном фактическом материале, строго обоснована, убедительно аргументирована и не вызывает серьезных возражений по существу. Отмеченные погрешности ни в коей мере не умаляют общего значения данной работы.

Автореферат полностью отражает содержание диссертации, тщательно оформлен и прекрасно проиллюстрирован.

С.Ю. Бакоевым выполнено новое фундаментальное исследование в области генетики и геномики сельскохозяйственных животных, направленное на выявление молекулярно-генетических маркеров породы, выявление генетических основ селекционно-ценных признаков и на создание прогностических моделей для оценки эффективности селекции и разведения животных и рисков формирования наследственных патологий. Результаты работы широко апробированы на 14 Российских и международных конференциях и форумах, опубликованы в 31 статье, из которых 25 индексируемых в

международных базах Scopus и Web of Science. По теме диссертации зарегистрировано 2 патента и 3 авторских свидетельства о государственной регистрации.

Диссертационная работа Бакоева Сирождина Юсуфовича «Анализ генетической архитектуры селекционно-ценных признаков свиней с применением современных методов обработки и интерпретации геномных данных» соответствует требованиям, предъявляемым к докторским диссертациям Положения о порядке присуждения ученых степеней постановления правительства Российской Федерации № 842 от 24 сентября 2013 г., изложенным в пунктах 9 – 14, 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), а сам диссертант несомненно заслуживает присвоения искомой степени доктора биологических наук по специальности 1.5.7. - Генетика.

Официальный оппонент

Заведующий лабораторией эволюционной генетики развития
Заместитель директора по научной работе
Федерального государственного бюджетного учреждения науки
Института биологии развития им. Н.К. Кольцова РАН,

Доктор биологических наук
28.08.2025

Куликов Алексей Михайлович

Федеральное государственное бюджетное учреждение науки
Институт биологии развития им. Н.К. Кольцова РАН
Адрес: 119334, г. Москва, ул. Вавилова 26
Телефон: +7(499) 135-87-81, e-mail: amkulikov@gmail.com

«Подпись А.М. Куликова заверяю»:
Ученый секретарь ИБР РАН
Кандидат биологических наук
28.08.2025



Хабарова Марина Юрьевна