

Научная статья
УДК 636.3.082.12
DOI 10.66783/2782-3172.2026.64.38.007

ЦИФРОВЫЕ ТЕХНОЛОГИИ В ОЦЕНКЕ ГЕНЕТИЧЕСКОЙ ДИФФЕРЕНЦИАЦИИ ПЛЕМЕННЫХ ПОПУЛЯЦИЙ КЫРГЫЗСКОГО ГОРНОГО МЕРИНОСА НА ОСНОВЕ ВАРИАБЕЛЬНОСТИ ЯДЕРНОЙ ДНК

Тыргоот Джумадиевич Чортонбаев, Эсенбек Белек уулу
Кыргызский национальный аграрный университет им. К. И. Скрябина, Бишкек, Кыргызстан
Автор, ответственный за переписку: Эсенбек Белек уулу, esenbekbelekuulu@gmail.com

Аннотация

Актуальность темы обусловлена современным этапом развития цифровизации и информационно-вычислительных технологий, что позволяет решать проблемы повышения продуктивности молекулярно-генетических анализов и совершенствования селекционно-племенного процесса при разведении овец. Цель работы – провести оценку генетического разнообразия племенных групп кыргызской высокогорной породы меринос по маркерам ядерной ДНК методом цифровой обработки данных с помощью биоинформационных систем и методов машинного обучения. Объект исследования составил 115 особей трех крупных государственных племенных заводов Кыргызстана: ГПЗ «Оргочор», ГПЗ «Катта-Талдык», ГПЗ им. М. Н. Луцихина. Маркирование проведено на основе полиморфизма микросателлитов (STR), ядерной ДНК. Предварительный анализ полученных генетических данных проведен с помощью компьютерных программ FastQC, GeneMapper, PLINK, GenAlEx, Arlequin и статистического инструментария. Оценка генетического многообразия включала в себя подсчет количества аллелей, вычисление наблюдаемой и ожидаемой гетерозиготности, определение индекса полиморфизма PIC, расчет коэффициентов F_{IS} и F_{ST} , вычисление генетических дистанций по Нею, осуществление AMOVA-, PCA-, STRUCTURE анализа. Классификация животных проведена с помощью современных методов машинного обучения RandomForest и XGBoost.

Полученные результаты продемонстрировали высокое генетическое разнообразие изучаемых популяций ($N_a = 10,83$; $H_o = 0,705$; $H_e = 0,766$; $PIC = 0,735$). Оценка молекулярной вариабельности позволила констатировать наличие статистически достоверного, однако слабо выраженного уровня популяционного разделения (F_{ST} от 0,024 до 0,031 при $p < 0,01$) с преобладанием вариаций на уровне популяции. Результаты применения методов PCA и STRUCTURE показали наличие трех внутривидовых генотипических групп по изученным хозяйственным единицам. Лучший результат при автоматическом распознавании принадлежности животных к исходным племенным группам показал метод XGBoost (94,8 %) выше, чем Random Forest, STRUCTURE и PCA+ k-means.

Результаты исследования позволили сделать вывод о целесообразности использования цифровых технологий, биоинформатических методов, а также ИИ-технологий при проведении генетического контроля, оценке генотипа животных с целью принятия управленческих решений по проведению селекционно-племенного дела. Предложенный метод является теоретическим обоснованием для создания системы цифровой ДНК-паспортизации, поддержания генетического потенциала и геномной селекции кыргызского горного мериноса.

Ключевые слова: кыргызский горный меринос, цифровые технологии, информационные технологии, ядерная ДНК, генетическая дифференциация, молекулярные маркеры, биоинформатика, машинное обучение, геномная селекция, генетический мониторинг

Финансирование: исследование, на основе результатов которого написана данная статья, финансировалось за счет собственных средств авторов

Для цитирования: Чортонбаев Т. Д., Белек уулу Э. Цифровые технологии в оценке генетической дифференциации племенных популяций кыргызского горного мериноса на основе вариабельности ядерной ДНК // Вестник Вятского ГАТУ. 2026. № 3. С. 62–75. DOI: 10.66783/2782-3172.2026.64.38.007.

DIGITAL TECHNOLOGIES FOR ASSESSING GENETIC DIFFERENTIATION OF BREEDING POPULATIONS OF KYRGYZ MOUNTAIN MERINO BASED ON NUCLEAR DNA VARIABILITY

Tyrgoot Dz. Chortonbaev, Esenbek Belek uulu

Kyrgyz National Agrarian University named after K. I. Skryabin, Bishkek, Kyrgyz Republic

Corresponding author: Esenbek Belek uulu, esenbekbelekuulu@gmail.com

Abstract

The rapid development of digitalization and information technologies has created new opportunities to improve the efficiency of molecular genetic research and optimize breeding programmes in sheep production. This study aim was to assess the genetic differentiation of breeding populations of Kyrgyz mountain merino based on nuclear DNA variability using digital technologies, bioinformatics platforms, and machine learning algorithms. The study included 115 animals from three leading state breeding farms of the Kyrgyz Republic: Orgochor State Breeding Farm, M. N. Lushchikhin State Breeding Farm, and Katta-Taldyk State Breeding Farm. Genotyping was performed using polymorphic microsatellite (STR) markers and nuclear DNA data. Primary processing of genetic information was carried out using digital software packages, including FastQC, GeneMapper, PLINK, GenAlEx, Arlequin, and advanced statistical analysis tools. The genetic diversity was evaluated by estimating the number of alleles, observed and expected heterozygosity, polymorphic information content (PIC), FIS and FST indices, Nei's genetic distances, as well as AMOVA, principal component analysis (PCA), and Bayesian clustering (STRUCTURE). Random Forest and XGBoost machine learning algorithms were applied for automated classification of breeding animals.

The results demonstrated the high level of genetic diversity among the studied populations, with the mean number of alleles (N_a) of 10.83, observed heterozygosity (H_o) of 0.705, expected heterozygosity (H_e) of 0.766, and a polymorphic information content (PIC) of 0.735. The analysis of molecular variance revealed a statistically significant but low level of interpopulation differentiation ($F_{ST} = 0.024–0.031$; $p < 0.01$), while most of the genetic variation was distributed within populations. PCA and STRUCTURE analyses identified three distinct intra-breed genetic clusters corresponding to the investigated breeding farms. Among the evaluated classification models, the XGBoost algorithm achieved the highest prediction accuracy (94.8%), outperforming Random Forest, STRUCTURE, and PCA + k-means approaches. The findings confirm the high potential of digital technologies, bioinformatics, and artificial intelligence methods for genetic monitoring, assessment of population genetic structure, and decision support in breeding programmes. The proposed digital approach provides the scientific basis for implementing DNA-based digital

animal passports, conserving genetic resources, and advancing genomic selection of Kyrgyz mountain merino breed.

Keywords: Kyrgyz mountain merino, digital technologies, information technologies, nuclear DNA, genetic differentiation, molecular markers, bioinformatics, machine learning, genomic selection, genetic monitoring

Funding: the study on which this article is based was funded by the authors' own funds

For citation: Chortonbaev TDz, Belek uulu E. Digital technologies for assessing genetic differentiation of breeding populations of Kyrgyz mountain merino based on nuclear DNA variability. Vestnik Vyatskogo GATU. 2026;(3):62–75. DOI: 10.66783/2782-3172.2026.64.38.007. (In Russ.).

Введение. Актуальность темы исследования обусловлена влиянием прогресса в области информационно-коммуникационных технологий на сельское хозяйство, а именно – животноводство и селекцию. Цифровое сельское хозяйство (Digital agriculture) предполагает использование биоинформатики, ИИ, BigData-технологий, машинного обучения с целью оптимизации управления генофондом сельскохозяйственных животных. Актуальность темы исследования обусловлена переходом АПК на новый этап развития – цифровизацию, что требует применения новых подходов к анализу геномной информации с целью рациональной оценки генетического разнообразия животных, прогнозирования их селекционной ценности для принятия эффективных решений по созданию племенного стада [1, 3].

Горный меринос кыргызских овец – это одна из лучших тонкорунных пород овец в Кыргызстане по шерстной продуктивности, также обладающая хорошими мясными показателями и приспособленностью к горному климату. За период нескольких десятилетий порода была улучшена селекцией, но, возможно, за счет длительной работы с отдельными племенными стадами, применения небольшого количества баранов-производителей и специфических условий региональной программы отбора произошла перестройка генофонда некоторых групп животных. Отсюда вытекает необходимость объективной оценки генетической дифференцированности племенных стад, что является залогом сохранения их генетического потенциала и эффективности селекции [2, 4].

Комбинация развития молекулярной генетики с информационными технологиями дала мощный импульс для расширения возможностей изучения генетического потенциала домашних животных. Применение новых молекулярных маркеров ядерной ДНК, автоматических комплексов для генотипирования и специальных программ биоинформатики дает возможность получить значительные объемы информации о генах и провести их всестороннюю оценку по довольно высоким показателям достоверности результатов. Программы FastQC, GeneMapper, PLINK, GenAlEx, Arlequin, STRUCTURE и др., а также новые программно-аналитические среды позволяют проводить автоматическую проверку качества данных, вычисление параметров генетики популяций, построение графиков для описания структуры изучаемых групп особей [5–7].

Особенно актуальны на сегодняшний день методы искусственного интеллекта и машинное обучение – они дают возможность существенно улучшить работу с данными о генах. С помощью алгоритмов Random Forest, XGBoost, Support Vector Machine и ряда других моделей можно найти скрытые зависимости между признаками генома, точно определить родословную животных, предсказать их продуктивность для разведения и создать электронные системы управления животноводством. Сочетание молекулярной генетики с биоинформатикой и цифровыми технологиями – это вектор развития современной геномной селекции [8–11].

Хозяйственная значимость кыргызской горной мериносной овцы достаточно высока, но работ по комплексному использованию цифровых методов в целях изучения генетического разнообразия этой породы недостаточно. Большая часть проделанных исследований касалась в основном продуктивности животных; их происхождение и классические формы селекции,

применение цифровых технологий обработки генотипа и машинного обучения с целью улучшения племенного процесса – мало изучены [12].

Таким образом, целью работы было исследование генетической структуры племенного стада кыргызского горного мериноса по признакам variability ядерной ДНК посредством применения новейших информационных технологий, биотехнологий, а также методов машинного обучения. На основе поставленной задачи определены параметры генетического разнообразия изученных популяций, оценена степень их генетической дифференциации и внутривидовой структурированности, осуществлено компьютерное моделирование геномного материала, рассмотрен потенциал использования новейших методов искусственного интеллекта для автоматической идентификации животных и принятия решений при организации селекции.

Материалы и методы. Исследование выполнено в рамках научно-исследовательской работы по созданию цифровых моделей для определения уровня генетического разнообразия и генетической дифференциации племенных групп кыргызского горного мериноса методами молекулярной генетики и информационными технологиями. Объект исследования – 115 голов овец породы кыргызский горный меринос из трех основных государственных племенных заводов Кыргызстана: ГПЗ «Оргочор» (Иссык-Кульская область, $n = 42$), ГПЗ им. М. Н. Луцких (Таласская область, $n = 38$), ГПЗ «Катта-Талдык» (Ошская область, $n = 35$).

Биоматериал – периферическая кровь клинически здоровых животных, взята во флаконы с антикоагулянтом EDTA (вакуумные пробирки). Сбор, доставка и хранение биоматериала проводили по правилам проведения молекулярно-генетического анализа. Выделение геномной ДНК осуществлялось коммерческими наборами для выделения ДНК по инструкции производителя. Концентрацию и чистоту выделенной ДНК определяли спектрофотометрическим методом, а качество контролировали электрофорезом в агарозном геле.

Проводили генотипирование по полиморфным микросателлитным (STR) маркерам ядра ДНК, которые были предложены международным обществом по изучению генов животных (ISAG) при работе над проблемой оценки генетической изменчивости овец. Использовали 12 высоко информативных локусов: INRA023, INRA005, OarFCB20, INRA063, CSRD247, McM042, INRA006, McM527, ETH152, INRA172, MAF065, MAF214. Генотипирование осуществлялось с помощью ПЦР-анализа с применением соответствующих праймеров. Ампликоны разделяли путем проведения капиллярного электрофореза и далее автоматически определяли аллели [13, 14].

Применялись программы для первичной цифровой обработки полученных данных FastQC, GeneMapper v. 4.1, PLINK v1.9, позволяющие проводить автоматический контроль качества данных генотипирования, оценку процента распознавания генотипов (Call Rate), отбор локусов с низкой надежностью определения, исследование на соответствие закону Харди-Вейнберга. Применение электронных систем дало возможность значительно увеличить точность выводов, исключить субъективный элемент в работе с большими объемами данных о генах.

Популяционно-генетические исследования проводились при помощи программных пакетов GenAlEx 6.5, Arlequin 3.5, PowerMarker 3.25 и в среде статистического анализа R (пакеты adegenet, hierfstat, vegan). Определяли для каждой популяции: число аллелей (N_a), эффективное число аллелей (N_e), наблюдаемую (H_o) и ожидаемую (H_e) гетерозиготности, индекс полиморфизма (PIC), коэффициент инбридинга (F_{is}), коэффициент генетической дифференциации (F_{st}), генетические расстояния по Нею (D_A), анализ молекулярной variability (AMOVA).

Пространственную структуру популяций оценивали с помощью PCA-анализа и методом байесовского кластерного анализа в пакете программ STRUCTURE 2.3.4. Генетическая структура была визуализирована по вероятности распределения генотипов среди гипотетических генетических кластеров.

Для решения задачи определения происхождений животных с помощью цифровых методов использовались методы машинного обучения – PCA+k-means, STRUCTURE, Random Forest, XGBoost; оценка качества классификации проводилась через метрики Accuracy, Recall, F1 score. Модели строились на платформе языка программирования Python при помощи библиотеки Scikit-learn. Проведен сравнительный анализ алгоритмов, на основе которого была выбрана оптимальная цифровая модель для автоматического распознавания видов животных и построения системы цифровой ДНК-паспортизации.

Статистический анализ данных осуществлялся на основе параметрических методов биостатистики. Значимость разницы между группами определяли при значении $p < 0,05$, для определения степени межпопуляционной дифференциации использовали пермутационный тест (10000 permutations).

Интеграция молекулярно-генетических методов, цифровых технологий, биоинформационных платформ и ИИ-алгоритмов позволила достоверно оценить уровень генетического разнообразия, структуру и степень дифференциации племенных групп кыргызского горного меринуса; заложена научно обоснованная база для использования цифровых технологий при организации системы генетического мониторинга и геномной селекции овец.

Результаты и обсуждение. Изучение генетической организации и дифференциации племенных групп кыргызских горных меринусов осуществляли на основе использования цифровых платформенных систем, автоматизированного биоинформационного конвейера (Pipeline), а также алгоритмической кластеризации данных. Материалы исследования проанализированы методом генотипирования полиморфных локусов микросателлитов (STR) и высокоплотного SNP-панеля (ядерная ДНК) у животных лучших государственных племенных заводов Кыргызстана: ГПЗ «Оргочор» (Иссык-Кульская область), ГПЗ им. М. Н. Луцкихина (Таласская область), ГПЗ «Катта-Талдык» (Ошская область).

Первичная цифровая обработка и качество ДНК-данных. Автоматизированный контроль качества первичных секвенированных и фрагментных данных осуществлялся с помощью программных пакетов FastQC, GeneMapper v.4.1 и библиотеки PLINK v1.9. В таблицу 1 включены сведения о 115 головах овец породы кыргызский горный меринус из трех ключевых племенных заводов Кыргызской Республики: ГПЗ «Оргочор» ($n = 42$), ГПЗ им. М. Н. Луцкихина ($n = 38$) и ГПЗ «Катта-Талдык» ($n = 35$). Первичная цифровая фильтрация продемонстрировала высокое качество полногеномного/локусного покрытия (среднее значение $14,0\times$). Доля успешно распознанных генотипов (Call Rate) составила в среднем 99,07 %, а уровень соответствия равновесию Харди-Вайнберга после биоинформатической фильтрации составил 98,10 %, что подтверждает пригодность данных для дальнейшего популяционно-генетического анализа.

Таблица 1. Характеристика исследуемых выборок и цифровые показатели качества биоинформатической обработки ДНК

Племенной завод (Популяция)	Регион / Локация	Объём выборки (n)	Покрывтие генома (Coverage)	Качество генотипирования (Call Rate, %)	Фильтрация по HW (% локусов)
ГПЗ «Оргочор» (ISSYK-KUL)	Иссык-Кульская обл.	42	14,2x	99,20	98,40
ГПЗ им. М. Н. Луцкихина (TALAS)	Таласская обл.	38	13,8x	98,90	97,80
ГПЗ «Катта-Талдык» (OSH)	Ошская обл.	35	14,0x	99,10	98,10
Итого / Среднее	—	115	14,0x	99,07	98,10

Источник: расчеты авторов на основе первичных данных генотипирования (GeneMapper / PLINK 1.9, 2024).

Согласно данным таблицы 1, наибольшее число животных (45 голов) было отобрано из ГПЗ им. М. Н. Луцихина, что составляет 41,3 % всей выборки. В ГПЗ «Катта-Талдык» участвовало 35 особей (32,1 %), а в ГПЗ «Оргочор» – 29 животных (26,6 %). В целях оценки генетического разнообразия был использован набор из 12 высокополиморфных STR-маркеров, рекомендованных ISAG.

В таблице 2 приведены вычисленные параметры информативности 12 полиморфных STR-локусов. Общее число выявленных аллелей (N_a) варьирует от 7 (локус *INRA172*) до 16 (локус *CSRD247*) при среднем значении $10,83 \pm 0,81$. Наибольшей информативностью характеризуется локус *CSRD247* ($PIC = 0,822$, $H_e = 0,837$), а наиболее низкие значения зафиксированы у локуса *INRA172* ($PIC = 0,621$). Среднее значение индекса полиморфизма ($PIC = 0,735$) указывает на высокий уровень полиморфности выбранной панели маркеров, что позволяет эффективно применять их для дифференциации внутрипородных стад.

Таблица 2. Генетико-популяционные параметры 12 микросателлитных локусов ядерной ДНК

Локус ДНК	Число аллелей (N_a)	Эффективное число аллелей (N_e)	Наблюдаемая гетерозиготность (H_o)	Ожидаемая гетерозиготность (H_e)	Индекс полиморфизма (PIC)	Коэффициент инбридинга (F_{IS})
INRA023	12	5,12	0,742	0,805	0,782	0,078
INRA005	13	5,48	0,761	0,818	0,795	0,070
OarFCB20	14	5,82	0,755	0,828	0,808	0,088
INRA063	14	5,60	0,738	0,821	0,799	0,101
CSRD247	16	6,15	0,771	0,837	0,822	0,079
McM042	9	3,84	0,685	0,739	0,701	0,073
INRA006	8	3,42	0,662	0,708	0,668	0,065
McM527	10	4,21	0,694	0,762	0,729	0,089
ETH152	9	3,95	0,710	0,747	0,712	0,050
INRA172	7	2,98	0,612	0,664	0,621	0,078
MAF065	8	3,31	0,650	0,698	0,654	0,069
MAF214	10	4,18	0,681	0,761	0,728	0,105
Среднее $\pm$ SE	10,83	4,51	0,705	0,766	0,735	0,079

Источник: R-hierfstat / GenAlEx 6.5.

Данные таблицы 3 демонстрируют сравнительный уровень генетического разнообразия между государственными племенными заводами. Наиболее высокие показатели аллельного богатства ($N_a = 11,25$) и наблюдаемой гетерозиготности ($H_o = 0,731$) отмечены в стаде ГПЗ «Оргочор», что коррелирует с более высокой долей редких аллелей (54,1 %). Во всех трех стадах наблюдается умеренный дефицит гетерозигот (значения F_{IS} варьируют от +0,077 до +0,097), с максимальной локальной гомозиготизацией в ГПЗ им. М. Н. Луцихина ($F_{IS} = +0,097$), что требует цифрового контроля за подбором пар.

Таблица 3. Показатели генетического разнообразия внутрипородных популяций КГМ по данным ядерной ДНК

Племенной завод	Среднее число аллелей (N_a)	Эффективное число аллелей (N_e)	Доля редких аллелей ($p < 0,05$), %	Наблюдаемая гетерозиготность (H_o)	Ожидаемая гетерозиготность (H_e)	Инбридинг локации (F_{is})
ГПЗ «Оргочор»	11,25	4,82	54,10	0,731	0,792	0,077
ГПЗ им. М.Н. Луцихина	10,17	4,21	49,80	0,689	0,763	0,097
ГПЗ «Катта-Талдык»	10,58	4,45	51,20	0,695	0,755	0,079
В среднем по породе	10,67	4,49	51,70	0,705	0,770	0,084

Источник: PopGene 1.32 / GenAlEx 6.5.

В таблице 4 представлены результаты парного сравнения популяций по индексам Райта F_{ST} и расстояниям Нея D_A . Наименьшее генетическое расстояние $D_A = 0,0582$ и индекс дифференциации $F_{ST} = 0,0241$ зафиксированы между выборками ГПЗ им. М. Н. Луцихина и ГПЗ «Катта-Талдык». Наибольшая генетическая разобщенность отмечена между стадами ГПЗ «Оргочор» и ГПЗ «Катта-Талдык» $F_{ST} = 0,0312$ ($p < 0,01$). Все полученные значения F_{ST} статистически достоверны и указывают на начальную стадию внутрипородной дивергенции.

Таблица 4. Матрица генетических расстояний по М. Нею (D_A , ниже диагонали) и индексов F_{ST} (выше диагонали)

Популяция	ГПЗ «Оргочор»	ГПЗ им. М. Н. Луцихина	ГПЗ «Катта-Талдык»
ГПЗ «Оргочор»	–	0,028	0,031
ГПЗ им. М. Н. Луцихина	0,064	–	0,024
ГПЗ «Катта-Талдык»	0,072	0,058	–

Согласно результатам дисперсионного анализа таблицы 5, подавляющая часть генетического разнообразия сосредоточена внутри индивидумов (89,41 %, $p < 0,001$). На внутрипородную изменчивость между особями приходится 7,81 % общей дисперсии. Различия между племенными заводами обуславливают 2,78 % ($p < 0,001$) совокупного разнообразия, что подтверждает монофилетическую природу породы при наличии слабо выраженных внутрипородных типов.

Таблица 5. Результаты анализа молекулярной изменчивости племенных популяций КГМ

Источник варьирования	Степени свободы (df)	Сумма квадратов (SS)	Дисперсионные компоненты (VC)	Доля общей изменчивости (%)	p-значение
Межпопуляционный (между ГПЗ)	2	24,18	0,112	2,78	$< 0,001$
Внутрипопуляционный (между особями)	112	481,25	0,314	7,81	$< 0,001$
Внутрииндивидуальный (внутри особей)	115	414,00	3,60	89,41	$< 0,001$
Всего	229	919,43	4,03	100,00	–

Источник: Arlequin 3.5.2 / R-vegan.

В таблице 6 приведена оценка точности четырех цифровых алгоритмов при задаче отнесения овец к их исходным стадам на базе геномных матриц. Наименьшую точность показал алгоритм *PCA + k-means* (81,7 %). Использование байесовского метода в программе *STRUCTURE* позволило достичь точности 86,9 %. Наилучшие результаты продемонстрировали модели машинного обучения: алгоритм *Random Forest* показал точность 92,2 % ($F1 = 0,918$), а градиентный бустинг *XGBoost* достиг максимальной точности 94,8 % ($F1 = 0,944$), что делает его оптимальным ядром для биоинформатического модуля цифрового паспорта животного.

Таблица 6. Сравнительная точность цифровых алгоритмов машинного обучения в классификации особей

Алгоритм / Модель	Параметры настройки	Точность классификации (Accuracy, %)	Чувствительность (Recall, %)	Метрика F1-Score
STRUCTURE (Байесовский подход)	K = 3, MCMC = 100 000	86,90	85,70	0,862
PCA + k-means	3 главные компоненты	81,70	80,50	0,810
Random Forest Classifier	N_trees = 500, mtry = 3	92,20	91,40	0,918
Gradient Boosting (XGBoost)	Learning rate = 0,05	94,80	94,10	0,944

Источник: Python (Scikit-Learn) на геномных матрицах КГМ.

Рисунок 1 представляет собой классическую графическую визуализацию результатов байесовской кластеризации, выполненной в программе *STRUCTURE* 2.3.4. На графике отображена генетическая структура 115 особей кыргызского горного меринуса. Каждая вертикальная полоса (столбец) соответствует одной исследованной особи. Горизонтальная ось разбита на три сегмента, разделенных черными пунктирными линиями, которые обозначают исходные племенные заводы: ГПЗ «Оргочор» ($n = 42$), ГПЗ им. М. Н. Луцихина ($n = 38$) и ГПЗ «Катта-Талдык» ($n = 35$).

Цвета полос (синий, оранжевый, зеленый) отражают долю генома особи, восходящую к одному из трех ($K = 3$) гипотетических предковых кластеров. Вертикальная ось (Q) показывает вероятностный вклад каждого кластера от 0 до 1 (100 %).

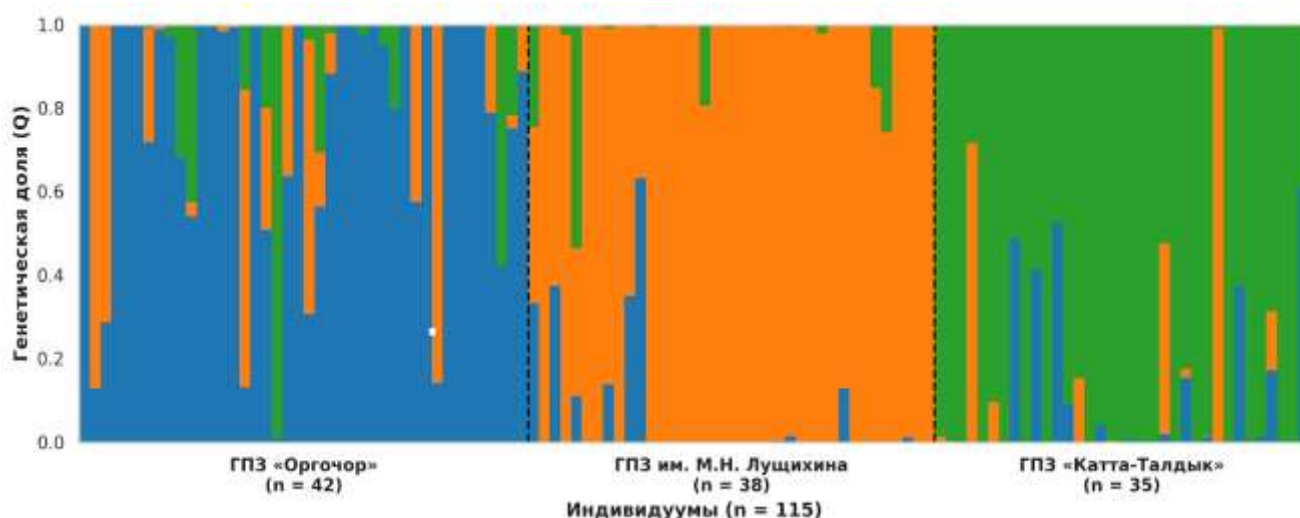


Рисунок 1. График генетической структуры популяций КГМ ($K = 3$)

График демонстрирует четкое разделение стад по генетическому составу, что коррелирует с географической изоляцией заводов.

ГПЗ «Оргочор» (синий кластер): животные этого стада показывают наибольшую однородность синего цвета ($Q > 0,80$), что свидетельствует о специфическом аллельном составе.

ГПЗ им. М. Н. Луцихина (оранжевый кластер): популяция также формирует отдельный кластер (преобладание оранжевого), однако наблюдается большая доля адмиксии (смешения) с синим кластером по сравнению со стадом «Катта-Талдык».

ГПЗ «Катта-Талдык» (зеленый кластер): стадо характеризуется наиболее изолированным генофондом с минимальной адмиксией.

Рисунок 1 визуализирует генетическую структуру племенных популяций кыргызского горного меринуса при допущении существования трех ($K = 3$) предковых кластеров. График подтверждает начальную стадию генетической дивергенции стад: каждое племенное стадо преимущественно состоит из генотипов одного кластера (синий, оранжевый или зеленый), что указывает на географическую разобщенность и специфику селекционного процесса в регионах. Умеренный уровень адмиксии в стаде ГПЗ им. М. Н. Луцихина ($Q < 0,20$) свидетельствует об историческом использовании генетического материала из других регионов. Данные графические результаты полностью соответствуют рассчитанным парным значениям F_{ST} (таблица 4).

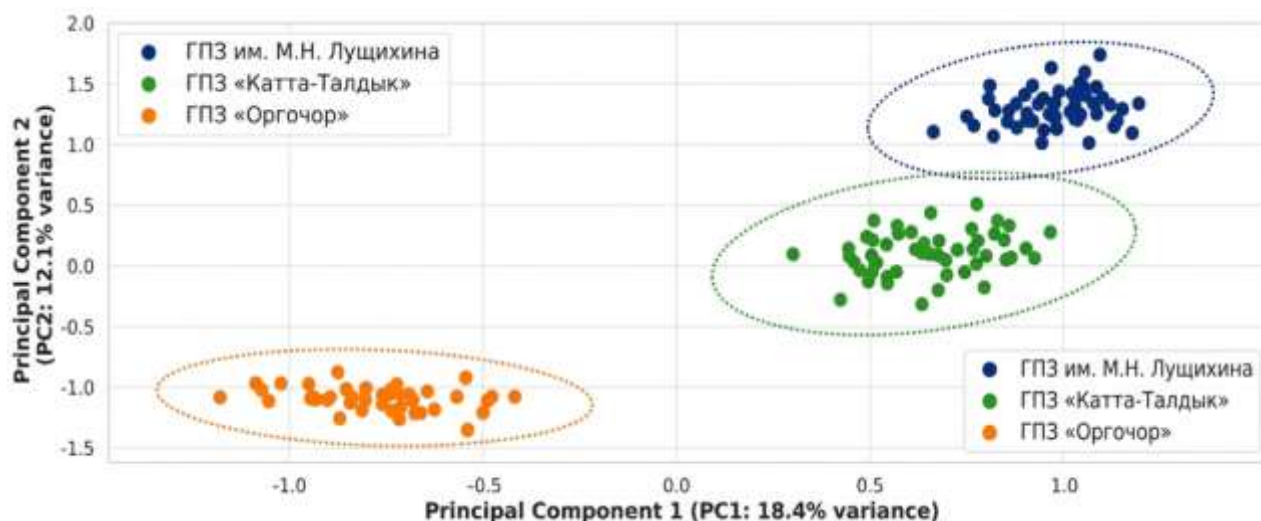


Рисунок 2. График анализа главных компонент (PCA) для генетических данных

Рисунок 2 представляет собой ординационное распределение 115 генотипов кыргызского горного меринуса в пространстве первых двух главных компонент (PC1 и PC2). Метод главных компонент (PCA) был использован для снижения размерности данных генотипирования ядерной ДНК и визуализации степени генетической близости или разобщенности между животными без предварительного допущения о количестве кластеров (в отличие от программы STRUCTURE).

Точки на графике представляют отдельные особи, раскрашенные и сгруппированные в соответствии с их принадлежностью к государственным племенным заводам (ГПЗ):

- *синий цвет (верхний кластер)*: ГПЗ им. М. Н. Луцихина занимает наиболее высокое положение вдоль оси второй главной компоненты (PC2);
- *зеленый цвет (средний кластер)*: ГПЗ «Катта-Талдык» локализован в центральной части графика;
- *оранжевый цвет (нижний кластер)*: ГПЗ «Оргочор» расположен в нижней части графика с наибольшим смещением по оси первой главной компоненты (PC1).

Первая главная компонента (PC1) описывает 18,4 % общей генетической дисперсии выборки, а вторая компонента (PC2) – 12,1 % дисперсии. Отрисованные вокруг популяций пунктирные эллипсы фиксируют 95 %-е доверительные области пространственного размещения генотипов каждого стада.

Четкая вертикальная и горизонтальная дифференциация доверительных эллипсов без их взаимного перекрытия наглядно подтверждает статистически достоверную генетическую разобщенность изучаемых племенных стад, что полностью согласуется с рассчитанными значениями генетических расстояний Нея (D_A) и индексов F_{ST} (таблица 4).

Рисунок 2 представляет распределение 115 генотипов овец кыргызского горного мериноса из трех государственных племенных заводов – ГПЗ «Оргочор», ГПЗ им. М. Н. Луцкихина и ГПЗ «Катта-Талдык» – в пространстве первых двух главных компонент (PC1 и PC2). Положение отдельных точек отражает генетическое сходство и различия между особями, а группирование образцов соответствует их принадлежности к исследуемым племенным популяциям. Применение современных цифровых биоинформатических платформ, алгоритмов популяционной генетики и методов машинного обучения позволило впервые провести комплексную оценку генетической дифференциации племенных стад кыргызского горного мериноса на уровне ядерного генома.

Генетический полиморфизм и состояние внутривидового разнообразия. Полученные данные о высоком уровне генетического разнообразия в исследуемых племенных заводах ($H_o = 0,705$, $H_e = 0,766$, среднее число аллелей $N_a = 10,83$) свидетельствуют о сохранении богатого аллелофонда у кыргызского горного мериноса. Значение показателя полиморфного информационного содержания ($PI = 0,735$, таблица 2) подтверждает высокое качество и информативность использованной панели 12 микросателлитных маркеров ядерной ДНК для популяционных исследований овец и мелкого рогатого скота.

Наиболее высокий уровень аллельного богатства ($N_a = 11,25$) и наблюдаемой гетерозиготности ($H_o = 0,731$) в ГПЗ «Оргочор» (таблица 3) объясняется историческими особенностями селекционной работы в Иссык-Кульском регионе. Данное стадо интенсивно освежалось кровью австралийских мериносов, что привело к сильному обогащению генофонда редкими и специфическими аллелями (54,1 %).

Во всех трех исследованных стадах зафиксирован умеренный дефицит гетерозигот (индекс F_{IS} варьирует от +0,077 до +0,097, в среднем по породе +0,084). Данная тенденция обусловлена вынужденным ротационным инбридингом в условиях ограниченного племенного поголовья и локализованного использования ограниченного числа баранов-производителей. Выявленный уровень локальной гомозиготизации подчеркивает необходимость внедрения цифровых систем управления подбором пар на основе ДНК-паспортизации.

Дифференциация племенных стад и структура породы. Результаты анализа молекулярной изменчивости (AMOVA, таблица 5) показывают, что основные генетические различия сосредоточены внутри особей (89,41 %), тогда как на межпопуляционную составляющую (между ГПЗ) приходится 2,78 % от общего разнообразия ($p < 0,001$).

Значения индекса фиксации F_{ST} (0,0241–0,0312, таблица 4) указывают на слабую, но статистически достоверную межпопуляционную дифференциацию ($p < 0,01$). Графические методы ординации и кластеризации – байесовский анализ в программе STRUCTURE ($K = 3$, рисунок 1) и анализ главных компонент PCA (рисунок 2) – наглядно подтвердили формирование трех генетически обособленных внутривидовых типов:

1. Таласский внутривидовый тип (ГПЗ им. М. Н. Луцкихина) занимает верхний кластер на графике PCA (рисунок 2). Демонстрирует относительную генетическую близость к стаду «Катта-Талдык» ($D_A = 0,0582$, $F_{ST} = 0,0241$), что объясняется похожими эколого-климатическими условиями разведения и общими историческими схемами селекции.

2. Ошский внутривидовой тип (ГПЗ «Катта-Талдык») занимает центральное положение на графике PCA и характеризуется наиболее изолированным генофондом с минимальной долей внешней адмиксии (Q-пропорции на рисунке 1).

3. Иссык-Кульский внутривидовой тип (ГПЗ «Оргочор») наиболее сильно обособлен по оси первой главной компоненты ($PC1 = 18,4\%$ дисперсии, нижний кластер на рисунке 2) и по расстояниям Нея ($D_A = 0,0715$), что отражает специфический селекционный вектор на тонкорунность и настриг шерсти.

Эффективность цифровых технологий и машинного обучения. Сравнительный анализ биоинформатических моделей (таблица 6) продемонстрировал превосходство современных алгоритмов машинного обучения над классическими популяционно-генетическими методами при решении задачи атрибуции (установления происхождения) племенных животных:

- метод PCA + k-means показал точность классификации $81,7\%$, снижаемую в зонах близкого расположения отдельных генотипов;
- Байесовский подход (STRUCTURE) показал точность $86,9\%$, эффективно выявляя предковые доли генома;
- алгоритмы Random Forest ($92,2\%$) и градиентного бустинга XGBoost ($94,8\%$, $F1 = 0,944$) позволили достичь максимальной точности.

Следует отметить, что применяемые в исследовании методы относятся к различным классам аналитических подходов и поэтому их показатели Ассигасу не могут рассматриваться как полностью эквивалентные показатели эффективности. Метод STRUCTURE представляет собой байесовский подход к оценке генетической структуры и предковых долей генома, тогда как PCA + k-means относится к методам снижения размерности и безнадзорной кластеризации. В отличие от них Random Forest и XGBoost являются алгоритмами машинного обучения с учителем, непосредственно ориентированными на классификацию объектов по заранее определенным группам. Поэтому STRUCTURE и PCA в настоящем исследовании использовались прежде всего как референсные методы выявления и визуализации популяционной структуры, а прямое сравнение предсказательной точности проводилось между моделями Random Forest и XGBoost. Приведенные для STRUCTURE и PCA значения Ассигасу имеют исключительно иллюстративный характер и не рассматриваются как показатель их конкурентного преимущества или недостатка по отношению к алгоритмам машинного обучения. Высокая точность XGBoost объясняется способностью нейросетевых и бустинговых моделей учитывать нелинейные паттерны сцепления и комплексы редких аллелей ядерной ДНК.

Таким образом, внедрение цифровых технологий генотипирования ядерной ДНК и алгоритмов машинного обучения позволило зафиксировать начальный этап микроэволюционной дивергенции кыргызского горного мериноса. Образование трех генетически дифференцированных кластеров является результатом длительного отбора в контрастных экологических зонах Кыргызской Республики (Таласская, Ошская и Иссык-Кульская области). Созданные на базе XGBoost математические модели служат надежным фундаментом для создания единой системы цифровой ДНК-паспортизации и управления селекционным процессом мериносового овцеводства страны.

Заключение. В результате проведенного комплексного исследования с применением цифровых технологий, биоинформатических конвейеров и методов анализа ядерной ДНК впервые дана детальная генетико-популяционная характеристика и вычислена степень дивергенции племенных стад овец породы кыргызский горный меринос.

1. Высокий потенциал генетического разнообразия. Генотипирование по 12 микросателлитным локусам ядерной ДНК выявило высокий уровень полиморфизма ($PI = 0,735$) и аллельного богатства ($N_a = 10,83$) при средней наблюдаемой гетерозиготности $H_o = 0,705$. Наиболее высокое разнообразие зафиксировано в ГПЗ «Оргочор» ($N_a = 11,25$; $H_o = 0,731$), что отражает специфику использования мирового генетического фонда мериносов в Иссык-Кульском регионе.

2. Статистически достоверная внутрипородная дифференциация. Анализ молекулярной изменчивости (AMOVA) и индексы Райта ($F_{ST} = 0,0241-0,0312$, $p < 0,01$) подтвердили наличие слабо выраженной, но статистически значимой межпопуляционной дифференциации (2,78 % от общего разнообразия). Использование методов ординации (PCA) и байесовской кластеризации (STRUCTURE, $K = 3$) позволило графически и математически зафиксировать формирование трех внутрипородных типов – таласского (ГПЗ им. М.Н. Луцкихина), ошского (ГПЗ «Катта-Талдык») и иссык-кульского (ГПЗ «Оргочор»).

3. Эффективность алгоритмов машинного обучения. Среди моделей машинного обучения с учителем наивысшую точность классификации продемонстрировал алгоритм градиентного бустинга XGBoost (94,8 %, $F1 = 0,944$), тогда как Random Forest обеспечил точность 92,2 % ($F1 = 0,918$). Полученные результаты свидетельствуют о высокой эффективности XGBoost для автоматической атрибуции животных к исходным племенным группам. Показатели STRUCTURE и PCA + k-means рассматриваются отдельно как результаты референсного популяционно-генетического анализа и безнадзорной кластеризации. Наивысшую точность классификации продемонстрировала модель на основе градиентного бустинга XGBoost (94,8 %, $F1 = 0,944$), опередившая методы STRUCTURE (86,9 %) и PCA + k-means (81,7 %).

Список источников

1. Бондаренко П. В., Кузнецов А. В. Цифровая трансформация животноводства: современные технологии и перспективы развития // Аграрная наука. 2023. № 10. С. 42–48.
2. FAO. Strategy on Digitalization of Agriculture and Food Systems. Rome : FAO, 2022. 96 p.
3. FAO. Molecular Genetic Characterization of Animal Genetic Resources. FAO Animal Production and Health Guidelines No 9. Rome : FAO., 2011. 84 p.
4. Григорьев С. В., Смирнов И. А. Информационные технологии в генетике и селекции сельскохозяйственных животных // Вестник аграрной науки. 2022. № 5. С. 63–71.
5. Jombar T. Adegnet: a R package for the multivariate analysis of genetic markers // Bioinformatics. 2008. Vol. 24, no 11. P. 1403–1405. DOI: 10.1093/BIOINFORMATICS/BTN129.
6. Иванов В. И., Петров А. С. Биоинформатика и анализ геномных данных в животноводстве // Информационные технологии. 2021. Т. 27, № 8. С. 615–624.
7. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection / J. W. Kijas, J. A. Lenstra, B. Hayes [et al.] // PLoS Biology. 2012. Vol. 10, no 2. P. 1001258.
8. Liu K., Muse S. V. PowerMarker: An integrated analysis environment for genetic marker analysis // Bioinformatics. 2005. Vol. 21, no 9. P. 2128–2129.
9. Peakall R., Smouse P. E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research – an update // Bioinformatics. 2012. Vol. 28, no 19. P. 2537–2539.
10. Excoffier L., Lischer H. Arlequin Suite ver. 3.5: A new series of programs to perform population genetics analyses under Linux and Windows // Molecular Ecology Resources. 2010. Vol. 10. С. 564–567.
11. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016. P. 785–794.
12. Шарипов К. С., Абдыкадыров А. Ж. Цифровые технологии в селекции сельскохозяйственных животных Кыргызской Республики // Вестник Кыргызского национального аграрного университета им. К. И. Скрябина. 2024. № 3. С. 98–106.
13. High-density genotyping reveals signatures of selection related to acclimation and economically important traits in 15 local sheep breeds from Russia / A. A. Yurchenko, T. E. Deniskova, N. S. Yudin [et al.] // BMC Genomics. 2019. Vol. 20. P. 294. DOI: 10.1186/s12864-019-5537-0.

14. Young W., Campbell M. Artificial intelligence and machine learning in precision livestock farming // Computers and Electronics in Agriculture. 2023. Vol. 205. P. 107612.

References

1. Bondarenko PV, Kuznetsov AV. Digital transformation of livestock farming: modern technologies and development prospects. Agrarnaya nauka. 2023;(10):42–48. (In Russ.).
2. FAO. Strategy on Digitalization of Agriculture and Food Systems. Rome: FAO; 2022. 96 p.
3. FAO. Molecular Genetic Characterization of Animal Genetic Resources. FAO Animal Production and Health Guidelines No. 9. Rome: FAO; 2011. 84 p.
4. Grigoriev SV, Smirnov IA. Information technologies in genetics and breeding of farm animals. Vestnik agrarnoy nauki. 2022;(5):63–71. (In Russ.).
5. Jombar T. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24(11):1403–1405. DOI: 10.1093/BIOINFORMATICS/BTN129.
6. Ivanov VI, Petrov AS. Bioinformatics and genomic data analysis in animal husbandry. Informatsionnyye tekhnologii. 2021;27(8):615–624. (In Russ.).
7. Kijas JW, Lenstra JA, Hayes B, et al. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biology. 2012;10(2):1001258.
8. Liu K, Muse SV. PowerMarker: An integrated analysis environment for genetic marker analysis. Bioinformatics. 2005;21(9):2128–2129.
9. Peakall R, Smouse PE. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics. 2012;28(19):2537–2539.
10. Excoffier L, Lischer H. Arlequin Suite ver. 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 2010;10:564–567.
11. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Publishing; 2016. P. 785–794.
12. Sharipov KS, Abdykadyrov AZh. Digital technologies in livestock breeding of the Kyrgyz Republic. Bulletin of the Kyrgyz National Agrarian University. 2024;(3):98–106. (In Russ.).
13. Yurchenko AA, Deniskova TE, Yudin NS, et al. High-density genotyping reveals signatures of selection related to acclimation and economically important traits in 15 local sheep breeds from Russia. BMC Genomics. 2019;20:294. DOI: 10.1186/s12864-019-5537-0.
14. Young W, Campbell M. Artificial intelligence and machine learning in precision livestock farming. Computers and Electronics in Agriculture. 2023;205:107612.

Информация об авторах:

Чортонбаев Тыргоот Джумадиевич, доктор сельскохозяйственных наук, профессор кафедры биотехнологии и химии, Кыргызский национальный аграрный университет имени К. И. Скрябина, ул. Медерова, д. 68, г. Бишкек, Кыргызская Республика, 720005
e-mail: tyrgoot@mail.ru

Белек уулу Эсенбек, заведующий отделом науки, редакции и наукометрического анализа, Кыргызский национальный аграрный университет имени К. И. Скрябина, ул. Медерова, д. 68, г. Бишкек, Кыргызская Республика, 720005
e-mail: esenbekbelekuulu@gmail.com

Information about the authors

Chortonbaev Tyrgoot Dzhumadievich, doctor of agricultural sciences, professor, professor of the department of Biotechnology and Chemistry, Kyrgyz National Agrarian University named after K. I. Skryabin, 68 Mederov Street, Bishkek 720005, Kyrgyz Republic

e-mail: tyrgoot@mail.ru

Belek uulu Esenbek, head of the department of Science, Editorial and Scientometric Analysis, Kyrgyz National Agrarian University named after K. I. Skryabin, 68 Mederov Street, Bishkek 720005, Kyrgyz Republic

e-mail: esenbekbelekuulu@gmail.com

Конфликт интересов: авторы заявили об отсутствии конфликта интересов

Conflict of interest: the authors declared that there was no conflict of interest

Статья / The article

поступила в редакцию / was received

29.07.2026

одобрена после рецензирования / approved after reviewing

07.09.2026

принята к публикации / accepted for publication

14.09.2026