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. Лущихина. 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 внутрипородных 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

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