

ANNOTATION

For the dissertation for the Doctor of Philosophy (PhD) degree
on the specialty “6D070100 Biotechnology” of

AMANDYKOVA MAKPAL DUMANKYZY

on the theme

«Genetic diversity and genome variation study of populations of Kazakhstan camels»

General description of the dissertation. The dissertation is devoted to the study of the genetic diversity and genome of camel populations in Kazakhstan, assessment of their interspecific distance, and comparative analysis of their phylogenetic similarity.

Relevance of the research topic. Camel farming is one of the leading branches of agriculture throughout the world; the level of use of camel products is increasing every year. Camels are widely used for the production of high-quality milk, meat, and wool and are an important species in terms of environmental sustainability. In addition, due to their high adaptability to the conditions of desert regions, these animals make an irreplaceable contribution to the process of economic development. The genetic characteristics of individual camel species and breeds are considered the basis for their sustainable use. Thus, by characterizing camels at the level of phenotypic and genotypic variability, the foundations are laid for optimizing breeding programs for traits useful for agriculture.

Currently, the structure of the genome of one-humped (*Camelus dromedarius*) and two-humped (*Camelus bactrianus*) camels are being studied in depth all over the world (genetic variability, molecular markers, identification of individual genes, identification of their origin, whole-genome analysis, etc.). As a result of such studies, it was determined that the divergence of one-humped and two-humped camels as separate species occurred 4.4-8 million years ago. Studies of mitochondrial DNA and nuclear markers have established that wild camels (*Camelus ferus*) are a distinct species and that their genetic separation from the ancestors of domestic camels occurred approximately 0.6–1.8 million years ago. As a result of the development of next-generation sequencing technologies, it has become possible to perform whole-genome sequencing of the camel genome, and whole-genome sequencing data has been uploaded to the GenBank system.

Optimization of targeted breeding work by studying genes associated with camel productivity and identifying molecular markers will improve the economic efficiency of camel farming. For example, in recent years, several studies have been carried out to describe the genes responsible for milk and meat productivity and wool quality, as well as mutations occurring in these genes.

However, even though work on the genetics of camels is being intensively studied all over the world, the genetic structure of the Kazakh populations of one-humped and two-humped camels, as well as their hybrids, which in recent years have been widely used in the camel farming of our country, is poorly described.

Thus, studying the genetic structure of camels is important for their sustainable use in agriculture and improving the management system of this industry, as well as preserving their biodiversity and understanding the characteristics of their environmental adaptation and evolutionary history.

The purpose of the research. Study of the genetic diversity and genome of camel populations in Kazakhstan, comparative analysis of their phylogenetic similarity.

Research objectives:

1. Conducting PCR-RFLP analysis of genes encoding casein proteins responsible for the qualitative properties of camel milk and statistical processing of the results;
2. Carrying out whole-genome sequencing of the hybrid camels genome and comparative analysis of sequencing results with data of whole-genome sequencing of one-humped, two-humped, and wild camels using bioinformatics analysis methods;
3. Analysis of the genetic structure of the studied camels using the results of whole-genome sequencing.
4. Description of the phylogenetic similarity of the studied camels and comparative analysis.

The research objects. Blood samples and genomic DNA isolated from the peripheral blood of one-humped, two-humped, and hybrid camels.

The subject of the research. Characteristics of the genetic structure and genome of populations of one-humped, two-humped, and hybrid camels.

Research methods. A method of genomic DNA isolation from blood samples, genomic DNA quantitative and qualitative analysis, PCR-RFLP analysis, whole-genome sequencing, statistical processing of PCR-RFLP analysis results, bioinformatics analysis of the whole-genome sequencing data.

The scientific novelty of the research.

- For the first time, the polymorphism of Kazakh camel populations in genes encoding casein proteins (*CSN2*, *CSN3*, *CSN1S1*), affecting the quality properties of camel milk, was described, and animals with “useful” genotypes were identified based on the studied SNPs;

- For the first time, whole-genome sequencing of hybrid camels was carried out, and the resulting raw data was uploaded to the GenBank database (BioProject number: PRJNA985411);

- For the first time, the genetic structure of hybrid camels has been described, as well as their phylogenetic similarity with one-humped, two-humped, and wild camels.

Theoretical and practical significance of the research. As a result of the research carried out in the dissertation, the polymorphism of the camel populations of Kazakhstan in the *CSN3*, *CSN2*, and *CSN1S1* genes, encoding the casein proteins of camel milk, was described. Genotyping data can be used as comparative material in further work on the study of the genetics of milk productivity of camels. As a result of studies conducted based on analysis of whole-genome sequencing data of one-humped, two-humped, hybrid, and wild camels, their genetic structure was characterized, private and common SNPs were identified, and their phylogenetic similarity was described. Raw data obtained as a result of whole-genome sequencing of hybrid camels has been uploaded to the GenBank database and is available to the global scientific community, paving the way for a more in-depth study of the hybrid camel genome.

The results of studies conducted on the polymorphism of genes encoding casein proteins of camel milk and the identification of animals with “useful” genotypes based on the studied SNPs can be used in breeding programs aimed at optimizing the quality properties of camel milk.

The main provisions for the defense:

1. It was established that point mutation g.2126A>G of the *CSN2* gene encoding the β -casein protein, point mutation g.1029T>C of the *CSN3* gene encoding the κ -casein protein, and point mutation g.942G>T of the *CSN1S1* gene encoding the protein α s1-casein, which affects the quality properties of camel milk, is distributed unevenly among camel populations bred in Kazakhstan.

2. As a result of determining SNP polymorphisms in the *CSN2* and *CSN3* genes, animals with “useful” genotypes for dairy camel breeding were identified, and the possibility of using them in breeding programs aimed at optimizing the quality properties of camel milk was considered.

3. Based on raw data of whole-genome sequencing, the genetic and population structure, as well as phylogenetic similarities of one-humped, two-humped, wild, and, for the first time, hybrid camels were characterized.

4. Private and common SNPs for one-humped, two-humped, wild, and hybrid camels were identified and compared. Based on them, it was determined that hybrid camels are closer in their genetic structure to one-humped camels than to two-humped camels.

Main research results and conclusions:

1. As a result of studying the polymorphism of genes encoding casein proteins responsible for the quality properties of camel milk, for the first time in the study, animals with “beneficial” genotypes for both SNPs studied in the *CSN2* and *CSN3* genes and which belong to the “core” of populations were identified. Thus, 10 % of the 100 camels studied can be attributed to the “core” populations; the majority of these animals were dromedary camels (70 %).

2. Analysis of whole genome sequencing data of one-humped, two-humped, hybrid, and wild camels revealed a total of 43,552,164 SNPs in all camels. The highest number of SNPs was identified in hybrid camels, thus the distribution of SNPs among the camel groups studied was as follows: hybrid camels (15,904,987 SNPs) > dromedary camels (10,872,754 SNPs) > Bactrian camels (9,310,730 SNPs) > wild camels (7,463,693 SNPs).

3. Based on the study of the genetic and population structure of one-humped, two-humped, wild, and, for the first time, hybrid camels, it was found that the genetic structure of hybrid camels is closer to one-humped camels than to two-humped camels. The number of common SNPs between hybrid and one-humped camels (4,617,974 SNPs) was much higher than the number of common SNPs between hybrids and two-humped camels (776,978 SNPs).

4. According to the results of phylogenetic analysis, hybrid camels were in the same branch of the phylogenetic tree as one-humped camels and showed a close relationship to them. The fact that *Camelus ferus* wild camels were located on a distant branch of the phylogenetic tree from groups of domesticated camels suggests that they represent a genetically isolated population.

Personal contribution of the author. The development and organization of laboratory work and research on the dissertation topic were carried out with the personal participation of the author.

Connection with the plan of basic scientific work. The dissertation work was carried out within the framework of the projects SC MSHE RK OR11465435 «Development and application of new genomic technologies for protecting organisms from mutagenic effects, increasing the productivity of natural resources and improving the quality of life of the population» (2021-2022) and AR14870678 «Study of genetic diversity and population genetic structure of *Camelus dromedarius* and *Camelus bactrianus* in Kazakhstan» (2022-2024).

Approbation of works. The research results of this dissertation were presented and discussed at the following international scientific conferences:

- VI international scientific conference of young scientists and students, Shymkent, Republic of Kazakhstan, December 7-8, 2018;
- International scientific conference of students and young scientists “Farabi Alemi”, Almaty, Republic of Kazakhstan, April 6-9, 2020;
- International scientific conference of students and young scientists “Farabi Alemi”, Almaty, Republic of Kazakhstan, April 6-8, 2021.

Publications. The main results of the dissertation were published in 7 printed works, including 2 articles in peer-reviewed scientific journals included in the Web of Science and Scopus databases with a Q2 index (*SJR* – 0.641, *percentile* – 71; *SJR* – 0.924, *percentile* - 51); 2 articles in domestic periodical journals recommended by the Committee for Quality Assurance in the Field of Science and Higher Education of the MSHE of the Republic of Kazakhstan; 3 theses were published in the proceedings of the international conferences. Based on the results of the dissertation work, 2 implementation certificates were received.

Scope and structure of the dissertation. The dissertation is presented on 86 pages of computer text, designed in compliance with the required standards, and includes normative references, notations and abbreviations, introduction, literature review, materials and research methods, research results and their discussion, and conclusion. The dissertation contains 22 figures, 13 tables, and 3 appendices. The list of used literature contains 171 sources.