ANNOTATION

for the dissertation work of Kairzhanova Alma Duisenbaykyzy on the topic «Genetic diversity of Francisella tularensis strains circulating on the territory of Kazakhstan», for the degree of Doctor of Philosophy Ph.D. on the educational program 8D09101 - "Veterinary welfare of animals".

Tularemia is a zooanthroponotic infection caused by the gamma-proteobacterium *Francisella tularensis*. This acute infectious disease has a pronounced natural foci with periodic epizootics. *F. tularensis* is a highly virulent, gram-negative intracellular coccobacillus, non-spore-forming, aerobic or microaerophilic bacterium. There are currently four recognized subspecies of *F. tularensis*: tularensis, holarctica, mediasiatica and novicida.

In Kazakhstan, natural foci of tularemia have been identified in 12 out of 14 oblasts, the total area of foci is about a quarter of the country's territory (552 thousand km2). Tularemia is characterized by a wide range of hosts and a variety of transmission routes. The main source and reservoir of *F. tularensis* are rodents, while blood-sucking insects are involved in maintaining the circulation of the pathogen in natural foci. Differences in susceptibility and infectious sensitivity determine the role of mammals as sources of infection - further spreaders of the pathogen. Sanitary well-being can be achieved only by controlling the incidence of disease among animals, including domestic animals, as there is a possibility of frequent contact with infected wild mammals and blood-sucking insects.

Knowledge of the genotypes of circulating strains is important for epidemiologic and epizootic monitoring at the local and global levels. At the local level, genotyping allows tracing the source and pathways of infection. At the global level, genotyping allows differentiating a natural outbreak from an artificially created outbreak as a result of malicious intent, and allows tracing evolutionary changes.

Goal and objectives of the research

The goal of the dissertation work is to study the genetic diversity of *Francisella tularensis* strains circulating in Kazakhstan using highly discriminatory methods and to map the distribution of genotypes to improve the control of tularemia.

In accordance with the goal the following objectives were set:

1. Creation of collections of DNA samples of *Francisella tularensis* strains suitable for genotyping by MLVA and PCR.

2. Development of *Francisella tularensis* genotyping protocol by multilocus analysis of VNTR repeats.

3. MLVA typing of *Francisella tularensis* strains by hypervariable VNTR markers. Obtaining MLVA profiles.

4. Determination of genetic diversity of *Francisella tularensis* strains, cluster analysis and construction of minimal island trees. Determination of geographical distribution of genotypes in Kazakhstan.

5. Whole genome sequencing of *Francisella tularensis* strains. Analysis of the obtained results. Construction of minimal island trees based on SNP data.

Objects of research: In this dissertation work, DNA samples isolated from collection strains of *Francisella tularensis* RGP on PCV "National Scientific Center of especially dangerous infections named after M. Aikimbayev" (NSCEI) serve as research material.

Subject of the research: Genetic diversity of *Francisella tularensis* strains circulating on the territory of Kazakhstan.

Methods of research. Microbiological, genetic and bioinformatic methods of research are used in the scientific work.

Scientific novelty of the performed work consists in the following:

MLVA genotyping of 148 strains of *Francisella tularensis* circulating in Kazakhstan was carried out for the first time, whole genomic data of 39 strains of *Francisella tularensis* subsp. *holarctica* isolated in Kazakhstan from natural water bodies, ticks, rodents, predators and from one migratory bird were obtained for the first time. Based on MLVA typing and whole genomic data, maps of genotype distribution in our country were drawn up.

Practical significance lies in the development of MLVA genotyping scheme of *Francisella tularensis* circulating in Kazakhstan. The proposed scheme has a higher discriminatory ability and improves genotyping schemes.

Theoretical significance consists in obtaining 148 MLVA profiles and 39 whole genomic data of *F.tularensis* strains circulating in Kazakhstan. The obtained results were uploaded to publicly available international databases and can be used by other scientists when studying the genetic diversity of *F.tularensis* strains circulating in the world and in Kazakhstan. Maps of genotypes distribution on the territory of our country were made, which visually shows the distribution of *Francisella tularensis* strains.

Main points for defense:

1. MLVA genotyping of *Francisella tularensis* strains by hypervariable VNTR markers.

2. Determination of geographical distribution of *Francisella tularensis* strains genotypes in Kazakhstan.

3. Whole genome sequencing of *Francisella tularensis* strains with construction of minimal residual trees based on SNP data.

Results of the research.

1. A collection of DNA samples of 148 *Francisella tularensis* strains from 8 regions of our republic suitable for genotyping by MLVA, PCR and whole genomic sequencing was created. At the same time, the largest number of strains were taken from West Kazakhstan region, which amounted to 41.89% in percentage ratio. A high degree of DNA purification was established in all samples, as indicated by the ratio of wavelengths 260/280 in the average range of 1.8. The DNA concentration ranged from 2 to 78 ng/μL. The nucleotide sequence of 16S rRNA gene was analyzed to exclude contamination and species identification of DNA samples. As a result of the analysis, the nucleotide sequences obtained had a maximum identity of 99-100% with the nucleotide sequences of *F. tularensis* species. Intraspecific identification of the studied strains of *F. tularensis* was carried out by PCR using primers specific to the RD1 gene, which revealed two subspecies in our DNA collection: *F. tularensis* subsp. *holarctica* and *F. tularensis* subsp *mediasiatica*.

2. A protocol for *Francisella tularensis* genotyping by multilocus analysis of 11VNTR repeats was developed, which includes the composition of the reaction mixture for each five primer mixtures, amplification mode and fragment analysis. Based on the developed MLVA typing protocol, methodological guidelines have been issued, which can be recommended for specialists of medical and veterinary laboratories.

3. MLVA typing of 148 *Francisella tularensis* strains by hypervariable 11VNTR markers was carried out. MLVA profiles for all analyzed *Francisella tularensis* strains were obtained. Allelic variants and Hunter-Gaston discriminant index values for each analyzed locus were identified. The Hunter-Gaston Diversity Index (HGDI) of MLVA-11 for 148 *Francisella tularensis* strains was 0.9295. Conducting MLVA at 11 loci allowed the identification of 30 genotypes among the 148 strains analyzed, of which 6 genotypes were represented by single strains. As a result of MLVA typing at 11 VNTR repeats, 40 *Francisella tularensis* strains were selected for whole genome sequencing.

4. Genetic diversity of *Francisella tularensis* strains was determined, cluster analysis was carried out and minimum-axis trees were constructed. With the help of electronic system QGIS the map of genotype distribution of *Francisella tularensis* strains on the territory of the Republic of Kazakhstan was developed. According to our studies for genotyping of *F.tularensis* subsp. *holartica*, circulating in Kazakhstan, it is reasonable to use a simplified genotyping scheme, which includes five of 25 classical VNTR loci, Ft-M3, Ft-M4, Ft-M6, Ft-M20A, Ft-M22 and two additional selected loci in silico-FT-4 and in silico-FT-8. All seven loci can be amplified in a single PCR reaction with different fluorescent dyes

and analyzed on a genetic analyzer in a single run. Based on the seven variable loci, 39 strains are grouped into 19 genotypes.

5. Whole genome sequencing of 40 strains of *F. tularensis* was performed for the first time in Kazakhstan: 39 strains of *F. tularensis* subsp. *holarctica* and 1 strain of *F. tularensis* subsp. *mediasiatica*, with the number of reads ranging from 347926 to 1219022 reads per sample. According to the evaluation of raw read data, the sequencing accuracy was at least 99.9%. Whole-genome sequencing generated 10953 single-nucleotide polymorphisms, tree size 11027 (homoplasy 0.67%). The analyzed strain *F. tularensis* subsp. *mediasiatica* 240 clustered with strains of MI subtypes. Phylogenetic analysis of the wgSNP sequence of 39 strains of *F. tularensis* subsp. *holarctica* was performed together with representative publicly available WGS data. As a result, 2022 SNPs were identified among 219 strains including 39 strains from Kazakhstan and 180 publicly available datasets belonging to phylogenetic groups B.4, B.6 and B.12. Phylogenetic group B.6 was not detected in the present DNA collection. Minimal-axis trees were constructed based on SNP data.

Approbation of the work. The main provisions of the dissertation were published in the materials of scientific-practical conferences: International scientific-theoretical conference "Seifullin's readings - 17: "Modern agrarian science: digital transformation", dedicated to the 30th anniversary of independence of the Republic of Kazakhstan on the topic: "Genetic identification of *Francisella tularensis*" (Nur-Sultan, 2021); International scientific-practical conference "Actual problems and trends in the development of modern agrarian science and veterinary science", dedicated to the memory of Doctor of Veterinary Sciences, Professor Valentin Ivanovich Piontkovsky on the topic: "Genetic diversity of *Francisella tularensis* strains circulating in Kazakhstan" (Kostanay, 2021).

Publication of research results.

On the materials of the dissertation work published three scientific papers, including in the editions recommended by the Committee for Control in Education and Science of the Ministry of Education and Science of the Republic of Kazakhstan: Scientific and Practical Journal "Science and Education" of the West Kazakhstan Agrarian and Technical University named after Zhangir Khan (Uralsk, 2020), published methodological recommendations "MLVA typing of Francisella tularensis strains" ISBN 978-601-332-968-0 (Nur-Sultan, 2020), 1 publication in the journal "Microbiology Resource Announcement" included in Web of Science Core Collection and Scopus (Q4, 2020), 1 publication in the journal "PLOS Neglected Tropical Diseases" included in Web of Science Core Collection and Scopus. (Q1, 2021).

Connection of the dissertation with the state programs. The research was conducted under the grant funding program of the Committee of Science of the Ministry of Education and Science of the Republic of Kazakhstan under the project: "Comparative analysis of molecular-genetic features of genomes in anthrax and tularemia pathogens in Kazakhstan", AR05131460, 2018-2020.

Scope and structure of the dissertation. The dissertation work is outlined on 123 pages of computer text and includes a literature review, materials and methods of research, research results, conclusion, list of used sources, 3 appendices. The list of used sources consists of 204 titles of domestic and foreign authors. The work contains 12 tables, 21 figures.