ANNOTATION
of the dissertation work for a degree of Doctor of Philosophy (Ph.D)
in specialty "6D010700 – Biotechnology"

MUSRALINA LYAZZAT ZENURAINOVNA

Paleogenetic analysis of pathogenic microorganisms in archaeological finds from the territory of Kazakhstan

General description of the research. This Ph.D. dissertation work belongs to the field of paleogenetics. It is devoted to the analysis of ancient pathogenic organisms in the human bone material, representing archaeological findings from the territory of the Central Eurasian steppe. Special attention is paid to ancient strains of plague and hepatitis B virus. With respect to these ancient pathogens of dangerous infections, all recent advances in DNA analysis of ancient and modern strains of Yersinia pestis and HBV in the evolutionary, medical, and ecological context have been summarized.

Significance of the research. In recent years, the possibilities of paleogenetics have expanded significantly due to the development of biotechnologies, improvement of methods of working with ancient biomaterials, development of modern bioinformatic analysis technologies, and accumulation of data on ancient DNA sequences of humans, animals, and lower organisms. The development of biotechnological methods of analysis, allowing not only to detect differences between ancient and modern DNA based on the degradation and chemical modification of ancient DNA, but also to successfully reconstruct ancient genomes, became the basis for the creation of specific software packages, without which it is now almost impossible to analyze ancient genomes.

In addition to the historical aspect, research on ancient pathogens is also relevant from the perspective of modern medicine. Sequences of human strains of dangerous viruses and bacteria, isolated from different geographic locations and at different time periods, allow to apply phylogenetic analysis approaches, perform geographic and molecular dating, and gain information about the distribution of infections.

The present work is of particular relevance due to the fact that archaeological material from Central Eurasia, including the territory of modern Kazakhstan, has not been practically studied in the context of ancient pathogens and migration history of particularly dangerous infections.

The purpose of the research. The purpose of this work was to perform paleogenetic DNA analysis of human remains from burial sites in the Central Eurasian region for the presence of ancient strains of pathogenic microorganisms and viruses.

Research objectives

1. Analysis of archaeological finds from the early periods of Central Eurasian history and the collection of research materials representing ancient human bone remains.

2. Extraction and analysis of aDNA samples. Preparation of DNA-libraries and NGS-sequencing.

3. Bioinformatics screening of NGS data for known human pathogens.

4. Verification of bioinformatics screening results for important pathogens.

5. Genome reconstruction and phylogenetic analysis of the most important ancient strains of human pathogens.

6. Analysis of relationships between the migration history of human populations and discovered pathogens.

The research objects and materials. Objects of research were samples of bone tissues and teeth of 360 ancient individuals dated from VII century BC to XIV century AD, 205 of them were from the territory of modern Kazakhstan, 120 - from Kyrgyzstan, 35 - from Russia.

Research methods. In the dissertation work, we used a set of modern methods for paleogenetic analysis of human bone remains, focusing on the genomes of ancient pathogens. All manipulations related to the preprocessing and separation of ancient DNA from the archaeological material were performed in a specialized laboratory for paleogenetic works with the use of all possible measures to protect against the contamination of ancient DNA with modern DNA. NGS sequencing of aDNA libraries was performed on Illumina plates (HiSeq 4000 / MiSeq). Bioinformatic processing of primary NGS data and screening for known pathogens was performed using the software packages EAGER and HOPS specialized for ancient DNA analysis. The presence of ancient strains of plague pathogens in human aDNA samples was established both by the real-time PCR method using specific primers to reference *Yersinia pestis* DNA and by methods of bioinformatic analysis of NGS data. Verification of the results of bioinformatic screening for significant pathogens was performed by Capture of reference DNA using global pathogen databases. Automatic reconstruction of ancient pathogen genomes and phylogenetic analysis of significant pathogens were performed taking into account all known global paleogenetic and modern data using MEGAN, VCF, BEAST, and R-Studio software packages.

The scientific novelty of the research. In this work firstly in the world the analysis of ancient causative agents on archaeological material, representing bone remains of the person from burials from territory of modern Kazakhstan and adjoining areas of Kyrgyzstan and Russia is carried out. Such big data set (360 ancient individuals) from territory of the Central-Eurasian Steppe was studied for the first time. With respect to these ancient pathogens of dangerous infections, all recent advances in DNA analysis of ancient and modern *Y. pestis* and HBV strains in evolutionary, medical, and ecological context were summarized.

The discovery of an ancient strain of *Y. pestis* of the Iron Age in a sample from the Kyzyl necropolis and medieval *Y. pestis* strains in 3 samples from the Nestorian Kara Djigach cemetery, which are genetic precursors of the second plague epidemic ("Black Death") in Europe.

The theoretical and practical significance of the research. The theoretical significance of the study lies in the fact that to date there has been no work related to the study of ancient infectious diseases circulating in Kazakhstan.

The theoretical significance of the work lies in the use of international experience of modern paleogenetic methods of analysis, in broad international collaboration for obtaining detailed information and comprehensive analysis of genomes of ancient pathogenic microorganisms and viruses, their phylogeny and relationship with historical events. Combining the results of genetic studies of ancient pathogens with data from other disciplines, such as archaeology, history and paleopathology, as well as genetics of human populations made it possible to build a more complete picture of the relationship between humans and pathogens and modern infectious diseases, trace ancient migration routes and fill gaps in the aspects of human history.

Practical significance of the dissertation work. In the course of the research the most effective methods of analysis of ancient pathogens have been established, which are of methodological importance. The interdisciplinary results of the research are of interest for specialists of adjacent sciences (physicians, anthropologists, archaeologists, demographers and historians). The results of works have wide educational value and will be used in educational process at preparation of courses of lectures and seminars for students of biological, medical, historical specialities, for development of refresher courses, and also will be used in popularization of science.

The results of the research are included into the curriculum of the discipline "IMKB 5206" Engineering Molecular Cell Biology on the specialty "7M05115-Biomedical Engineering" at Al-Farabi Kazakh National University. In addition, methods of collecting ancient bone material for paleogenetic analysis have been tested and implemented at the Toraygyrov Archaeological Research Centre of the University of Pavlodar and the RSE on PVC "A.Kh. Margulan Institute of Archaeology" in Almaty.

The main provisions for the defence are as follows:

1. A collection of bone remains, aDNA libraries, full-genome sequencing data, representing 360 ancient individuals dated from the 7th century BC to the 14th century AD from the territory of modern Kazakhstan (205 samples), Kyrgyzstan (120 samples) and Russia (35 samples) suitable for genome analysis of ancient pathogens and ancient humans was created.

2. The most common pathogens of the ancient world of the Central Eurasian Steppe of the Bronze and Iron Ages were microorganisms causing caries, diseases of gums and other soft tissues of the mouth (*Treponema denticola, Streptococcus mutans, Streptococcus oralis, Streptococcus gordonii, Tannerella forsythia, Veillonella parvula* and others). Cases of venereal diseases were frequent. Diseases such as hepatitis B (HBV), plague (*Yersinia pestis*), salmonellosis (*Salmonella enterica*) were reported only in a few individuals.

3. Ancient strains of HBV from territory of Kazakhstan (3 individes of Pazyryk and Hunno-Ksyanbi cultures from Berel necropolis and 1 sample of Tasmolinsk culture from Akbeit necropolis) are characterized by genotype D, that determines kinship with modern HBV strains (A and D genotypes) circulating in Eurasia.

4. The spread and evolution of *Y. pestis* in Central Eurasia and Europe during the Late Neolithic to Early Iron Age (LNBA) epoch followed parallel pathways.

This is evidenced by the common origin of LNBA lines and the presence of an ancestral, less infectious *pla* gene variant. Some ancient lines, including the specimen from Kyzyl necropolis (Central Kazakhstan, Tasmolinskaya culture, Early Neolithic Age), as a result of extended deletions and pseudogenization, lacked genes responsible for virulence (adhesion gene *yapC*), evasion of the host immune system (flagellin genes *flgB* and *fliZ*) and efficient transmission of plague bacilli from fleas (genes *ymt*). The most extensive deletion (~83 kb), including the loss of genes of the type VI secretion system (T6SS), in particular T6SS-G, characterizes the strain *Y. pestis* strain from Kyzyl necropolis, which is associated with weakening of its virulence.

5. Medieval strains of *Y. pestis* from the Nestorian cemeteries of Kara Djigach and Burana (Kyrgyzstan) are genetic precursors of the second plague epidemic ("Black Death") in Europe. The phylogeny identifies these strains as the ancestral form for the published genomes of the Second Plague Pandemic, separating it with one SNP from the isolate of the beginning of the pandemic from the Volga region, and with two SNPs from isolates associated with the Black Death from Western Europe.

6. Carriage of multiple and significant pathogens in representatives of different cultures and different periods from East Kazakhstan, as well as analysis of principal components of human populations genome, indicates a special importance of the Altai Mountains as a crossroads of migration routes, determining the richness of time-determined migrations of the ancient population from East and West to the Central Eurasian region and socio-economic relations.

Relationship of the research with the scientific project. This Ph.D. dissertation work is partially related to the project "Analysis of Sign Phenomena Representing the Early History of the Great Steppe to Solve Questions of Kazakh Ethnogenesis" (2018-2020) and the 2020-2022 project "Paleogenetic Analysis of Pathogenic Microorganisms in Archaeological Human Remains Representing the Central Eurasian Region" and the grant "Analysis of *Salmonella enterica* Ancient Genomes from Human Remains from Central Asia".

The contribution of the author for the results described in this dissertation.

Working with literature data on the topic of the dissertation; collecting information on archaeological materials; working with the catalog from the museum of the M.V. Lomonosov Moscow State University Research Institute. Lomonosov to select the necessary samples; communication with archaeologists; collection of archaeological samples; all laboratory part: sample preparation, bone chip preparation, preparation of solutions for DNA isolation, aDNA isolation, inhibition test, screening for the presence of the *Y. pestis pla* gene by real-time PCR, preparation of libraries for full genome sequencing, calculation of concentrations and preparation of pools for full genome sequencing on Illumina equipment, preparation for amplification to confirm the ancient pathogen. Bioinformatic screening of pathogens in obtained aDNA samples. Searching for and refining archaeological characteristics from collected samples. Introduction of information about the samples into the PANDORA database. Preparation of an act of implementation, abstracts, articles, and thesis materials.

Research approbation. The main results and observations are presented and confirmed by publications in prestigious international scientific journals included in the database Scopus, Web of Science, journals recommended by the CCES. The results were discussed at the Institute scientific seminars, and reported at international and republican scientific conferences.

Publications. The main results of the research on the dissertation topic have been published in 9 scientific papers, including 4 articles in international peer-reviewed journals included in the *Thomson Reuters* and *Scopus* databases with impact factors (IF = 14.136; IF = 41.846; IF =11.205, IF = 49.962) and 3 articles in the journals recommended by the Education and Science Control Committee of the RK, 2 abstracts in national and international conferences held in Kazakhstan and abroad.

Dissertation structure. The thesis consists of 115 pages, an introduction, a literature review, materials and methods of the research, results and their discussion, a conclusion and a list of references including 184 references, of which 176 in English. The work contains 25 figures, 5 tables, and 6 appendices.