

**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 significantly expanded due to the improvement of methods of working with ancient biomaterials, development of new methods of DNA analysis, development of modern technologies of bioinformatic analysis, and accumulation of data on ancient DNA of humans, animals, and lower organisms.

Studies of human and animal bone remains from ancient settlements showed that closer contact with domesticated animals increased the frequency of transmission of zoonoses, and higher population density increased the potential for pathogens to spread within and between groups. 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 in different time periods allow us to apply phylogenetic analysis approaches, perform geographic and molecular dating, and obtain information about the spread of infections. By studying the history of mutational changes in ancient genomes of bacteria and viruses, it is possible not only to understand the nature of epidemics, but also to prevent the emergence of new pandemics and epidemics. However, it is worth noting that, due to the complexity of ancient DNA analysis, the number of paleo DNA laboratories worldwide is still very limited, with very few laboratories working with ancient pathogens. Analysis of ancient pathogens requires making sure that the agent identified in ancient bone material is not the result of contamination. For this purpose additional investigations are necessary. Verification of a positive result requires confirmation by several methods of analysis.

The present work is especially relevant due to the fact that archaeological material from the Central Eurasian space, including the territory of modern Kazakhstan, has not been practically studied in the context of ancient pathogens and migration history of especially 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**

The main tasks of the research to accomplish purpose are as following:

1. The analysis of archaeological finds from the early periods of the history of the Central Eurasian region and the collection of materials for the study, representing the ancient bone remains of people.
2. Paleo-DNA isolation and analysis. Preparation of DNA-libraries and high throughput sequencing.
3. Bioinformatics screening of NGS data for known human pathogens including plague, tuberculosis, inflammatory and gastrointestinal pathogens.
4. Verification of bioinformatics screening results for significant pathogens.
5. Genome reconstruction and phylogenetic analysis of the most significant ancient strains of human pathogens in the context of world data on the genomes of ancient and modern pathogens.
6. Analysis of links between the migration history of human populations and detected 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 paleo-DNA 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 paleo-DNA 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 was carried out on archaeological material representing human bone remains from burials from territory of modern Kazakhstan and adjoining areas of Kyrgyzstan and Russia. Such big data set (360 ancient

individuals) from territory of the Central Eurasian Steppe has been studied for the first time. The novelty of the study also lies in establishment of ancient strains of *Yersinia pestis*, Hepatitis B, *Salmonella enterica* and other pathogenic bacteria and viruses circulating among human populations of Central Eurasian region since the Bronze Age till the Middle Ages.

Special attention is paid to ancient strains of plague and hepatitis B virus. All recent advances in DNA analysis of ancient and modern strains of *Yersinia pestis* and HBV in the evolutionary, medical and ecological context were summarized for these ancient pathogens of dangerous infections.

In collaboration with a large group of scientists from all over the world the evolutionary history of HBV was established on the basis of studying all the data on HBV strains identified in ancient humans (137 objects) from ~10,500 to ~400 years ago known to date. The results show pathways and shifts in viral diversity that reflect known historical human migrations and demographic events, as well as links to co-temporality. Moreover, Kazakhstan ancient HBV strains (1 sample of Pazyryk culture and 2 samples from Hunno-Ksienbi crypt Berel (East Kazakhstan region); 1 sample from Tasmolinsk culture of Akbeit necropolis (Karaganda region)) were found to belong to genotype D, which determines their genetic relationship with modern HBV strains (A and D genotypes) circulating in Eurasia.

The discovery of an ancient strain of *Yersinia pestis* of the Iron Age in a sample from the Kyzyl necropolis (Kara-Gandinskaya Oblast, Kazakhstan) and medieval *Yersinia pestis* strains in 3 samples from the Nestorian Kara-Jigach cemetery (Chui Oblast, Kyrgyzstan), which are genetic precursors of the second plague epidemic ("Black Death") in Europe, also represents a novelty of the study.

**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.

**The main provisions for the defence are as follows:**

1. A collection of bone remains, paleo-DNA libraries, full-genome sequencing data, representing 360 ancient individuals and 110 archaeological sites dated from the 7th century BC to the 14th century AD from the territory of modern Kazakhstan (81 sites, 205 samples), Kyrgyzstan (22 sites, 120 samples) and Russia (7 sites, 35 samples) suitable for genome analysis of ancient pathogens and ancient humans was created.

2. The most widespread pathogens of the ancient world of the Central Eurasian steppe of the Bronze and Iron age 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). Parasitic diseases transmitted by animals, such as helminthiasis, salmonellosis, typhoid fever were observed in individuals representing different necropolises. Cases of venereal diseases were frequent. Such dangerous diseases as hepatitis B (HBV), plague (*Yersinia pestis*), salmonellosis (*Salmonella enterica*) were registered only in single individuals.

3. Ancient strains of HBV from territory of Kazakhstan (3 individuals 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 *Yersinia 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 *Yersinia 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, paleoDNA 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 paleo-DNA 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 118 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 26 figures, 4 tables, and 9 appendices.