ABSTRACT

of the PhD dissertation submitted for the degree of Doctor of Philosophy (PhD) in the educational program «8D05104 – Genetics»

TAKENOV NURLAN KAIRATULY

on the topic **«Study of genotypes of the tuberculosis pathogen *Mycobacterium tuberculosis* in Kazakhstan»**

General characteristics of the dissertation research. The dissertation is devoted to a comprehensive study of the genetic structure and phylogenetic diversity of the tuberculosis pathogen, *Mycobacterium tuberculosis*, in Kazakhstan. The research employed whole-genome sequencing (WGS) and population genetics methods to identify the molecular mechanisms of drug resistance and the patterns of infection spread.

Relevance of the research. Tuberculosis (TB) remains a serious global threat: every year the World Health Organization (WHO) registers more than 10 million new cases, of which approximately 390,000 are drug-resistant forms. A particular danger is posed by multidrug-resistant TB (MDR-TB), which significantly reduces the effectiveness of therapy. Kazakhstan is among the high-burden countries: in 2024, the proportion of MDR-TB was about 27% among new patients and 44% among previously treated patients. These indicators call for a precise determination of the leading mechanism of resistance: whether it results from primary transmission (transmitted resistance) or has developed during treatment (acquired resistance). Distinguishing between these pathways is critical for adjusting infection control measures and treatment regimens.

Reconstruction of transmission pathways is impossible without an in-depth analysis of the genetic structure of the pathogen. The epidemic process is largely determined by the dominance of specific phylogenetic lineages, in particular the Beijing genotype, which is characterised by high transmissibility and a propensity for resistance. Lineages L2 (East Asian) and L4 (Euro-American) predominate in Kazakhstan; however, the genetic diversity of their sublineages and their contribution to the circulation of MDR-TB remain insufficiently studied.

Traditional genotyping methods (spoligotyping, MIRU-VNTR) have low resolving power and do not allow recent transmission chains to be reliably identified. Whole-genome sequencing (WGS) is a reliable solution: analysis of single nucleotide polymorphisms (SNPs) provides high accuracy in reconstructing phylogeny and transmission clusters. Despite this, studies of complete *M. tuberculosis* genomes in Kazakhstan remain isolated.

Thus, conducting a comprehensive WGS study within the framework of the educational programme "Genetics" is a highly relevant scientific task. The work aims to detail the genetic structure of circulating TB isolates in Kazakhstan down to the sublineage level, to characterise resistance profiles, and to analyse the geographic spread of infection, in order to improve the national epidemiological surveillance system.

Objects of the study. The objects of the study were 272 clinical isolates of *M. tuberculosis* recovered from biological samples of patients with laboratory-confirmed TB. The collection and selection of materials were conducted between January 2023 and December 2024 at the National Scientific Center of Phthisiopulmonology of the Ministry of Health of the Republic of Kazakhstan.

Aim of the study. The aim of the study is a comprehensive investigation of the genetic structure, molecular characterisation of the phylogenetic diversity, determination of drug resistance profiles, and assessment of recent transmission pathways of clinical *M. tuberculosis* isolates circulating in the territory of the Republic of Kazakhstan, based on whole-genome sequencing.

Tasks of the study:

1. Assemble a collection of clinical *M. tuberculosis* isolates and systematise patients' clinical and epidemiological data for subsequent genetic analysis.
2. Establish the phylogenetic structure of circulating *M. tuberculosis* isolates using WGS and bioinformatics data processing.
3. Characterise the intra-lineage genetic diversity of the pathogen population, detailing it down to the sublineage level.
4. Determine the molecular profile of drug resistance of the isolates based on SNP detection in target genes according to the WHO catalogue, and compare it with membership of specific genetic sublineages.
5. Assess the intensity of recent infection transmission and the proportion of endogenous TB reactivation in the population based on SNP cluster analysis.
6. Identify spatial-epidemiological patterns and local foci of transmission of pathogen genotypes in the Almaty region and Almaty city using geospatial analysis.

Research methods. Microbiological, molecular-genetic, and bioinformatics research methods were applied in this work. WGS of clinical *M. tuberculosis* isolates, followed by SNP analysis, reconstruction of phylogenetic trees and assessment of evolutionary dynamics, was performed using the TB-Profiler, MTBseq and IQ-TREE 2 tools. Statistical processing was carried out in the R environment (v4.3.1) using descriptive statistics and non-parametric tests (Mann–Whitney U test, Pearson's χ^2 test, Fisher's exact test). The diagnostic performance of molecular methods was assessed by calculating sensitivity and specificity, and spatial-genetic analysis was performed using the Mantel test and mapping methods.

Scientific novelty of the research. The dissertation research has produced comprehensive data characterising the population-genetic structure, phylogenetic diversity, drug resistance profiles and transmission pathways of *Mycobacterium tuberculosis* in the Republic of Kazakhstan on the basis of high-resolution WGS.

In a sample of clinical isolates, the phylogenetic structure of the pathogen in the region was characterised down to the sublineage level: the dominance of Lineage 2 Beijing (72.4%) and Lineage 4 Euro-American (26.8%) was established, and within the Beijing genotype the predominance of the Central Asia and Central Asia Outbreak (CAO) sublineages was demonstrated.

For the first time, pronounced heterogeneity of the drug resistance profile within the Beijing genotype was established depending on membership of a particular sublineage: a strong association of the CAO and Europe/Russian W148 Outbreak

sublineages with MDR-TB was shown, whereas the Central Asia sublineage predominantly remains drug-susceptible.

For the first time, at the level of whole-genome data for the region under study, the molecular resistance profile was characterised, confirming that the main determinants of resistance to rifampicin and isoniazid are the high-confidence mutations *rpoB* S450L (91.1%) and *katG* S315T (99.2%).

For the first time in Kazakhstan, an integrated assessment of TB transmission was applied (combining SNP cluster analysis with a threshold of ≤12 SNPs and epidemiological data), which made it possible to quantitatively distinguish the mechanisms of disease development: it was demonstrated that 92.65% of cases are associated with active transmission of infection and 7.35% with endogenous reactivation.

For the first time, using the Mantel test, the spatial-genetic structure of the pathogen population in the Almaty agglomeration was assessed, revealing statistically significant local foci of clonal transmission over short distances (up to 8 km) against a background of an overall diffuse distribution of genotypes.

Based on the results obtained, an integrated database of clinical-epidemiological and genomic data of 272 clinical *M. tuberculosis* isolates was created, and their complete genome sequences were successfully deposited in the international NCBI (GenBank) database under project number PRJNA1277026, which expands opportunities for further molecular-epidemiological research.

Practical significance. The effectiveness of introducing WGS and SNP analysis into the practice of the phthisiology service of the Republic of Kazakhstan has been substantiated, as a means of increasing the accuracy of molecular-epidemiological surveillance. The data obtained on the predominance of the transmission mechanism of disease spread (92.65%) and on local spatial clustering (within a radius of up to 8 km) can be used in planning infection control measures and in determining priority areas for epidemiological contact investigation. The identified concentration of severe forms of resistance in the CAO and Europe/Russian W148 Outbreak sublineages allows molecular typing to be applied as an additional tool for stratifying the epidemiological risk of TB foci. The confirmed dominance of the *rpoB* S450L and *katG* S315T mutations creates an experimental basis for the development and implementation of targeted molecular test systems for rapid diagnosis of drug resistance. The genomic database created and deposited in NCBI (GenBank, project PRJNA1277026) serves as a reference resource that can be used for subsequent molecular-epidemiological studies in Kazakhstan and Central Asia.

Theoretical significance. The obtained results supplement and expand contemporary understanding of the population structure and microevolution of *M. tuberculosis* under conditions of high prevalence of drug-resistant TB in Kazakhstan. The established intra-lineage heterogeneity of Beijing sublineages in terms of resistance profile – from an extremely unfavourable one in CAO and Europe/Russian W148 Outbreak to preserved susceptibility in Central Asia – contributes to understanding the mechanisms of differential epidemiological success of genetic sublineages of the pathogen within a single phylogenetic lineage. The obtained estimate, according to which 92.65% of the studied TB cases are associated with recent

transmission rather than with reactivation of endogenous infection, expands understanding of the role of active transmission of drug-resistant strains in sustaining the MDR-TB epidemic in high-burden regions. The identification of a spatial structure of genetic relatedness of strains that is statistically significant but pronounced only over short distances contributes to the development of the methodology of spatial-genetic analysis in the molecular epidemiology of tuberculosis.

Main provisions submitted for defence:

1. The *Mycobacterium tuberculosis* population in Kazakhstan is characterised by a pronounced dominance of Lineage 2 Beijing (72.4%), alongside the presence of Lineage 4 Euro-American (26.8%), which forms the basis of the genetic structure of circulating isolates.

2. Within the Beijing genotype, the Central Asia (49.2%) and Central Asia Outbreak (CAO) (44.2%) sublineages predominate, with CAO forming clonal epidemiological clusters with low SNP distance, reflecting the active circulation of the TB pathogen.

3. The genetic profile of drug resistance is characterised by a high frequency of MDR-TB (29.0%) and the predominance of the *rpoB* S450L (91.1%) and *katG* S315T (99.2%) mutations among isolates resistant to rifampicin and isoniazid, respectively – the key first-line drugs.

4. An integrated analysis of SNP clustering (threshold of ≤12 SNPs) and the clinical and epidemiological status of patients showed that 252 isolates (92.65%) are associated with recent transmission, whereas 20 isolates (7.35%) are associated with reactivation of a previously experienced infection, which confirms the predominantly transmission-driven spread of tuberculosis and its drug-resistant genotypes in the population.

5. The genetic-geographic structure of the *M. tuberculosis* population is characterised by the absence of global spatial association ($r = 0.06173$; $p = 0.1628$), alongside the persistence of local foci of clonal transmission of infection within a radius of up to 8 km ($r = 0.0623$; $p = 0.048$).

Author's personal contribution. The author independently formulated the research objectives, assembled the collection of clinical isolates, prepared samples for sequencing, and completed the full cycle of bioinformatics analysis. The author also performed statistical data processing, data interpretation, prepared scientific publications, and drafted the dissertation.

Connection with the scientific work plan. The dissertation work was carried out at the National Scientific Center of Phthisiopulmonology (Almaty) as part of the scientific research plan. The study utilised the material and technical base of the center. Reagents and consumables were provided with the support of the Global Fund to Fight AIDS, Tuberculosis and Malaria, as well as through the scientific center's own funds. The research topic aligns with the priority direction of scientific development in the Republic of Kazakhstan: "Life and Health Science".

Approbation of the work. The main results of the study were presented and discussed at the following scientific events:

- International Conference "The Union World Conference on Lung Health", Paris, France, 2023;

- International Conference "The Union World Conference on Lung Health", Bali, Indonesia, 2024;
- International Conference "The Union World Conference on Lung Health", Copenhagen, Denmark, 2025;
- International Scientific and Practical Conference "Genetics and Physiology: Past, Present, Future", Almaty, Kazakhstan, 2025.

Publications. 12 scientific works have been published on the topic of the dissertation, including 3 articles in journals indexed in the international databases Scopus and Web of Science, 2 articles in publications recommended by the Committee for Quality Assurance in Science and Higher Education of the Ministry of Science and Higher Education of the Republic of Kazakhstan, and 7 abstracts in the proceedings of international scientific and practical conferences.

Structure of the dissertation. The dissertation consists of an introduction, literature review, materials and methods, results of original research, conclusion, list of references, and appendices. The dissertation is presented on 136 pages of computer text, complying with the necessary standards, and includes 16 tables and 25 figures. The list of references contains 282 sources.