

## WHOLE GENOME SEQUENCING OF MYCOBACTERIUM TUBERCULOSIS IN KAZAKHSTAN: FIRST SEQUENCE RESULTS OF TWO CLINICAL ISOLATES

A. Molkenov<sup>1</sup>, U. Kairov<sup>1</sup>, U. Kozhamkulov<sup>1</sup>, S. Rakhimova<sup>1</sup>, A. Askapuli<sup>1</sup>, M. Zhabagin<sup>1</sup>, Zh. Abilova<sup>1</sup>, A. Akhmetova<sup>1</sup>, D. Yerezhepov<sup>1</sup>, A. Abilmazhinova<sup>1</sup>, V. Bismilda<sup>2</sup>, L. Chingisova<sup>2</sup>, Zh. Zhumadilov<sup>1</sup>, A. Akilzhanova<sup>1</sup>

1) Department of Genomic and Personalized Medicine, Center for Life Sciences, Nazarbayev University, Astana, Kazakhstan; \*askhat.molkenov@nu.edu.kz; 2) National Center of Tuberculosis Problems of the Republic of Kazakhstan, Almaty, Kazakhstan

**Objective.** The project is aimed to create the prerequisites for a personalized approach to the diagnosis and treatment of tuberculosis (TB) by identifying and comparing the whole genome sequences of *M.tuberculosis* strains isolated in Kazakhstan. Analysis for whole genome sequences obtained using the next generation sequencing technology will clarify the factors cause of the formation of highly virulent strains of *M.tuberculosis*, the evolution of local strains, and genetic markers of drug resistance.

**Methods.** Material collection from 50 patients, sputum extraction and determination of drug sensitivity was performed in the reference-laboratory "National Center of Tuberculosis Problems", Almaty, Kazakhstan. DNA libraries for whole genome sequencing were prepared from DNA extracted from the isolates. The whole genome sequencing was performed on Roche 454 GS FLX+ next-generation sequencing platform at the Center for Life Sciences, Nazarbayev University, Astana, Kazakhstan. The sequencing reads from two isolates were assembled into contigs using GS De Novo Assembler. All alignments were done against the *M.tuberculosis* reference strain H37Rv using GS Reference Mapper and Mauve. Comparing isolates against reference strain were done with Sibelia, visualizing data were done with Circos .

**Results.** The whole genome sequencing has performed for two *M.tuberculosis* isolates MTB-476 and MTB-489. 96 M bp with an average read length of 520 bp, approximately 21.8X coverage and 104.2 M bp with an average read length of 589 bp and approximately 23.7X coverage were generated for the MTB-476 and MTB-489, respectively. The genome of MTB-476 consists of 257 contigs, 4204 CDS, 46 tRNAs and 3 rRNAs. MTB-489 has 187 contigs, 4183 CDS, 45 tRNAs and 3rRNAs.

The results of genome assembling have been submitted into NCBI GenBank and available for public access under the accession numbers AZBA00000000 and AZAZ00000000. These genome assemblies can be useful for comparative genome analysis and for identification of novel SNPs and gene variants. Further work is being conducted on detailed analysis of results from whole genomes, genotyping of *M.tuberculosis* isolates that circulated on the territory of Kazakhstan.