



# Sequencing and Analysis of Three *Mycobacterium tuberculosis* Genomes of the B0/N-90 Sublineage

Natalia V. Zakharevich,<sup>a</sup> Marina V. Zaychikova,<sup>a</sup> Kirill V. Shur,<sup>a</sup>  Olga B. Bekker,<sup>a</sup>  Dmitry A. Maslov,<sup>a</sup> Valery N. Danilenko<sup>a</sup>

<sup>a</sup>Vavilov Institute of General Genetics Russian Academy of Sciences, Moscow, Russia

**ABSTRACT** We report the draft genome sequences of three *Mycobacterium tuberculosis* isolates belonging to the B0/N-90 sublineage, EKB34, EKB53, and EKB79. The B0/N-90 sublineage belongs to the prevalent (in Russia) and highly virulent Beijing-B0/W148 sublineage. Isolates EKB34 and EKB79 were obtained from people with immune deficiency.

Recently, tuberculosis has become more common. The probability of disease manifestation depends on several factors, chief among which are the host immune status and virulence of the pathogen, which are different in different lines (1, 2). Currently, there is a suggestion that the B0/N-90 sublineage (a group belonging to the prevalent [in Russia] and highly virulent Beijing-B0/W148 sublineage of the Beijing family) mainly affects immunocompromised people (3, 4).

In this article, we describe the following isolates from people living in Russia: *M. tuberculosis* EKB34, isolated from a 49-year-old male, and *M. tuberculosis* EKB53 and EKB79, isolated from 16- and 27-year-old immunodeficient males, respectively. All isolates are part of the collection described in reference 4 and were genotyped as belonging to the B0/N-90 sublineage by nonsynonymous single-nucleotide polymorphism (SNP) analysis described in reference 3. *M. tuberculosis* was cultured in Middlebrook 7H9 medium with the addition of oleic acid-albumin-dextrose-catalase (OADC; HiMedia, India) at 37°C for 4 weeks. Genomic DNA was isolated and purified by phenol-chloroform-isoamyl alcohol extraction after enzymatic cell lysis, as described by Belisle et al. (5). The quality of DNA was checked using gel electrophoresis and a Bioanalyzer 2100 system. Genomic DNA libraries were prepared using the NEBNext Ultra II DNA library prep kit for Illumina (New England BioLabs, USA). The raw sequencing data were obtained using a HiSeq 2500 platform (Illumina, USA) in rapid run mode with the HiSeq Rapid SBS kit v2 2 × 100 bp (Illumina). A total of 17,966,286 pairs of reads with an average read length of 111 bases were obtained. Quality check of the reads was done using FastQC v0.11.7 (6). *De novo* genome assembly was performed using SPAdes v3.11 (7), while assembly metrics were calculated with QUAST v5.0.2 (both programs were used with default parameters) (8). The automatic functional annotation results were obtained using the NCBI Prokaryotic Genome Annotation Pipeline. The characteristics of the three sequenced genomes are listed in Table 1.

To identify meaningful differences between isolates of the B0/N-90 sublineage and Beijing-B0/W148 sublineage, the unassembled reads were aligned to the reference genome of *M. tuberculosis* W-148 (GenBank accession number CP012090) using BWA-MEM, with default settings (9). The pileup was generated by mpileup (-B -f) in SAMtools (10). We used the mpileup2snp (-min-avg-qual 30 -min-var-freq 0.80 -p-value 0.01 -output-vcf 1) command in VarScan 2.3.9 (11) for calling single-nucleotide variants. We found 37 SNPs common to all the B0/N-90 strains and absent in the reference genome, 7 SNPs present in 2 out of 3 strains, as well as a number of unique SNPs present only in one strain (6 for EKB34, 8 for EKB53, and 11 for EKB79). A comparative analysis of virulence-associated genes from the previously developed catalogue (3, 12, 13) showed

**Citation** Zakharevich NV, Zaychikova MV, Shur KV, Bekker OB, Maslov DA, Danilenko VN. 2019. Sequencing and analysis of three *Mycobacterium tuberculosis* genomes of the B0/N-90 sublineage. *Microbiol Resour Announc* 8:e00796-19. <https://doi.org/10.1128/MRA.00796-19>.

**Editor** Julie C. Dunning Hotopp, University of Maryland School of Medicine

**Copyright** © 2019 Zakharevich et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Natalia V. Zakharevich, [zakharevich@yandex.ru](mailto:zakharevich@yandex.ru).

N.V.Z. and M.V.Z. contributed equally to this work.

**Received** 4 July 2019

**Accepted** 31 August 2019

**Published** 26 September 2019

**TABLE 1** Characteristics of three *M. tuberculosis* isolates belonging to the B0/N-90 sublineage

Isolate	WGS (GenBank) accession no.	SRA accession no.	Genome size (bp)	GC content (%)	Coverage (×)	No. of contigs	$N_{50}$ value (bp)	No. of contigs >500 bp	Total no. of CDSs <sup>a</sup>
EKB34	<a href="https://www.ncbi.nlm.nih.gov/nuccore/SRLI000000000">SRLI000000000</a>	<a href="https://www.ncbi.nlm.nih.gov/sra/SRR8439237">SRR8439237</a>	4,356,659	65.2	46	160	92,498	130	4,269
EKB53	<a href="https://www.ncbi.nlm.nih.gov/nuccore/SRLJ000000000">SRLJ000000000</a>	<a href="https://www.ncbi.nlm.nih.gov/sra/SRR8363334">SRR8363334</a>	4,358,808	64.4	169	180	73,246	149	4,268
EKB79	<a href="https://www.ncbi.nlm.nih.gov/nuccore/SRLK000000000">SRLK000000000</a>	<a href="https://www.ncbi.nlm.nih.gov/sra/SRR8374277">SRR8374277</a>	4,357,388	64.6	223	177	72,890	149	4,270

<sup>a</sup> CDSs, coding sequences.

that besides the previously known lineage-specific mutations (genes *mce3F* [rv1971],  $A_{1229}C$  [Asp→Ala]; *vapC46* [rv3384c],  $C_{113}G$  [Ala→Gly]; and *irtB* [rv1349],  $G_{523}A$  [Ala→Thr]) (3), all three isolates also had a mutation in the Rv3132c gene. This gene is part of the *dosRS* two-component system involved in response to stress conditions; in particular, it regulates the expression of the *hspX* gene, which encodes a protein that plays an important role in regulating the growth of *M. tuberculosis* under hypoxic conditions (14). However, this mutation is a synonymous one and does not result in an amino acid change. On the contrary, SNPs in the three genes described above may lead to changes in the properties of the encoded proteins, resulting in changes to virulence levels (15, 16).

**Data availability.** The whole-genome shotgun (WGS) assemblies have been deposited in NCBI GenBank. The versions described in this paper are the first versions. The read archives have been deposited in the NCBI SRA. The WGS (GenBank) and SRA accession numbers are listed in Table 1. All of the data are part of BioProject number [PRJNA509547](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA509547).

## ACKNOWLEDGMENT

This work was supported by the Project of the Ministry of Science and Higher Education of the Russian Federation, “The role of region-specific polymorphisms of virulence genes in the formation of drug resistance of *Mycobacterium tuberculosis*” (project identifier [ID] RFMEFI61317X0068).

## REFERENCES

- Coscolla M, Gagneux S. 2014. Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Semin Immunol* 26:441–444. <https://doi.org/10.1016/j.jsmim.2014.09.012>.
- Danilenko VN. 2019. Developing a technological platform to create innovative TB drugs active against multidrug-resistant strains. *Her Russ Acad Sci* 89:144–150. <https://doi.org/10.1134/S1019331619020035>.
- Mikhecheva NE, Zaychikova MV, Melerzanov AV, Danilenko VN. 2017. A nonsynonymous SNP catalog of *Mycobacterium tuberculosis* virulence genes and its use for detecting new potentially virulent sublineages. *Genome Biol Evol* 9:887–899. <https://doi.org/10.1093/gbe/evx053>.
- Shur KV, Umpeleva TV, Bekker OB, Maslov DA, Zaychikova MV, Vakhru-sheva DV, Danilenko VN. 2018. Compilation of the *Mycobacterium tuberculosis* Beijing-B0 lineage sample and identifying predictors of immune dysfunction in source patients. *Bull RSMU* 3:23–28. <https://doi.org/10.24075/brsmu.2018.040>.
- Belisle JT, Mahaffey SB, Hill PJ. 2009. Isolation of *Mycobacterium tuberculosis* species genomic DNA, p 1–12. In Parish T, Brown AC (ed), *Mycobacteria protocols*, 2nd ed. Humana Press, Totowa, NJ.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyskhin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595. <https://doi.org/10.1093/bioinformatics/btp698>.
- Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>.
- Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22:568–576. <https://doi.org/10.1101/gr.129684.111>.
- Forrellad MA, Klepp LI, Gioffré A, Sabio y García J, Morbidoni HR, de la Paz Santangelo M, Cataldi AA, Bigi F. 2013. Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* 4:3–66. <https://doi.org/10.4161/viru.22329>.
- Prozorov AA, Fedorova IA, Bekker OB, Danilenko VN. 2014. The virulence factors of *Mycobacterium tuberculosis*: genetic control, new conceptions. *Russ J Genet* 50:775–797. <https://doi.org/10.1134/S1022795414080055>.
- Kendall SL, Movahedzadeh F, Rison SC, Wernisch L, Parish T, Duncan K, Betts JC, Stoker NG. 2004. The *Mycobacterium tuberculosis dosRS* two-component system is induced by multiple stresses. *Tuberculosis (Edinb)* 84:247–255. <https://doi.org/10.1016/j.tube.2003.12.007>.
- Gupta A, Venkataraman B, Vasudevan M, Gopinath Bankar K. 2017. Co-expression network analysis of toxin-antitoxin loci in *Mycobacterium tuberculosis* reveals key modulators of cellular stress. *Sci Rep* 7:5868. <https://doi.org/10.1038/s41598-017-06003-7>.
- Zaychikova MV, Mikhecheva NE, Belay YO, Alekseeva MG, Melerzanov AV, Danilenko VN. 2018. Single nucleotide polymorphisms of Beijing lineage *Mycobacterium tuberculosis* toxin-antitoxin system genes: their role in the changes of protein activity and evolution. *Tuberculosis (Edinb)* 112:11–19. <https://doi.org/10.1016/j.tube.2018.06.011>.