

# Genomic analysis of Latin American-Mediterranean family of *Mycobacterium tuberculosis* clinical strains from Kazakhstan

Pavel Tarlykov<sup>1/+</sup>, Sabina Atavliyeva<sup>1</sup>, Arike Alenova<sup>2</sup>, Yerlan Ramankulov<sup>1</sup>

<sup>1</sup>National Center for Biotechnology, Nur-Sultan, Kazakhstan

<sup>2</sup>National Scientific Center for Phthisiopulmonology, Almaty, Kazakhstan

The human-adapted strains of the *Mycobacterium tuberculosis* complex (MTBC) comprise seven phylogenetic lineages originally associated with their geographical distribution. Here, we report the genomes of three drug-resistant clinical isolates of the Latin American-Mediterranean (LAM) family collected in Kazakhstan. We utilised whole-genome sequencing to study the distribution and drug resistance of these isolates. Phylogenetic analysis grouped the genomes described in this study with the sequences from Russia, Uzbekistan, and Kazakhstan belonging to the LAM family. One isolate has acquired extensive drug resistance to seven antituberculosis drugs. Our results suggest at least two multi-drug resistant (MDR)/extensively drug-resistant (XDR)-associated genotypes of the LAM family circulate in Kazakhstan.

Key words: *Mycobacterium tuberculosis* - genome - tuberculosis - phylogeny

*Mycobacterium tuberculosis* is a human pathogen with diverging lineages initially associated with a specific geographic region.<sup>(1)</sup> The lineages present in the human-adapted *M. tuberculosis* complex (MTBC) are distinguished as ancient and modern. The ancestral lineages include lineage 1 (Indo-Oceanic), lineage 5 (West Africa 1), lineage 6 (West Africa 2), while recently discovered lineage 7 (Ethiopia) appears to be intermediate between the ancient and modern ones.<sup>(2)</sup> Modern lineages include lineage 2 (East-Asian), lineage 3 (East-African-Indian), and lineage 4 (Euro-American), which includes the Latin American-Mediterranean (LAM) family. The LAM family was first discovered based on the strain collection whose descent was mainly Latin America and the Mediterranean area.<sup>(3,4)</sup> Human migrations have led to a recent expansion of the LAM family worldwide. In 2014, the LAM family has already been observed in 47 countries with different prevalence rates, including several Central Asian countries (Kazakhstan, Uzbekistan, and Turkmenistan) and neighboring Russia.<sup>(5)</sup> To date, LAM is the most predominant family of *M. tuberculosis* observed in Kazakhstan after the Beijing family.<sup>(6)</sup> Region of difference (RD) loci divided the LAM lineage into several sublineages, namely RD-Rio, RD174, and RD115. RD-Rio sublineage is defined by 26 kb deletion and generally concomitant with large deletion RD174.<sup>(7)</sup> The other sublineage is characterised by deletion RD115, which includes the LAM-RUS branch with a specific insertion of IS6110 into the *plcA* gene.<sup>(8)</sup>

Kazakhstan is among the 30 countries with the highest burden of multidrug-resistant tuberculosis in the world.<sup>(9)</sup> A recent expansion of drug-resistant isolates in the country is linked to the spread of “successful” Central Asian/Russian sublineage of *M. tuberculosis* Beijing genotype.<sup>(10)</sup> At the same time, a number of studies have reported rapid acquisition of drug-resistance in the LAM family including KZN isolates from South Africa and LAM-RUS sublineage widespread in Kazakhstan and its neighboring countries.<sup>(8,11)</sup> In previous studies, the LAM genetic family was found at a proportion of 11% in a sample of 470 *M. tuberculosis* isolates from 12 provinces of Kazakhstan (51/470 = 10.85%).<sup>(6,12)</sup> The other study describes local LAM isolates represented mainly by RD115 LAM-RUS sublineage (29/30 = 96.67%).<sup>(13)</sup> Additionally, recent findings observed a very first isolate with an RD-Rio deletion (LAM RD-Rio) not endemic to Central Asia.<sup>(14)</sup> Nevertheless, genomic data on the LAM isolates circulating in Kazakhstan is very limited. Very few local collections of MTBC isolates were tested for RD loci or IS6110 insertions.<sup>(13)</sup> Furthermore, there are only three whole-genome sequences (WGS) of the LAM isolates from Kazakhstan published to date.<sup>(5)</sup> The additional genomes will provide more data on the genetic variations occurring in drug-resistant LAM family isolates circulating in Kazakhstan and neighboring countries.

Thus, we provide the whole-genome sequencing data of three LAM isolates from a collection of drug-resistant *M. tuberculosis* isolates collected in Nur-Sultan city, Kazakhstan. Epidemiology of *M. tuberculosis* including lineage 4 is routinely studied by the analysis of single nucleotide polymorphisms (SNP), RD loci, IS6110-restriction fragment length polymorphisms (RFLP), spoligotyping, and/or mycobacterial interspersed repetitive units (MIRU)-typing.<sup>(15)</sup> Some of these conventional molecular genotyping methods are laborious and have various limitations.<sup>(16)</sup> In this study, we have implemented whole genome sequencing with *in silico* analysis of the epidemiological characteristics of the local LAM isolates.

doi: 10.1590/0074-02760200215

Financial support: This work was funded by Ministry of Education and Science of the Republic of Kazakhstan (grants AP05130238 and AP05132131).

+ Corresponding author: tarlykov@biocenter.kz

https://orcid.org/0000-0003-2075-307X

Received 06 May 2020

Accepted 20 August 2020



A collection of 28 drug-resistant *M. tuberculosis* strains was isolated in Nur-Sultan, Kazakhstan, from sputum of patients with clinically suspected tuberculosis. The drug susceptibility was tested using a Bactec MGIT 960 culture system (Becton, Dickinson) according to the manufacturer's protocol. DNA was extracted using the cetyltrimethylammonium (CTAB) procedure.<sup>(17)</sup> The quality of the DNA was checked using a Qubit double-stranded DNA (dsDNA) high-sensitivity (HS) assay kit (Thermo, Massachusetts, USA) and a Qubit 2.0 fluorometer (Thermo). A SNP real-time polymerase chain reaction (PCR) assay for detection of G to A transition in *fbpC* codon 103 (*Rv0129c*) was carried out using the CFX96 Touch System (Bio-Rad, California, USA). Three LAM isolates were SNP-confirmed for sequencing (no. 3538, 4142, and 4330).

Isolates no. 3538 and 4142 were sequenced using a MiSeq platform (Illumina, California, USA) and the other LAM isolate no. 4330 using an Ion Torrent platform (Thermo). For the MiSeq sequencing, libraries with an average fragment size of 600 bp were prepared using a Nextera DNA Flex Library Prep kit (Illumina) according to the manufacturer's instructions. A barcoded library for the isolate no. 4330 was prepared using an Ion Xpress Plus fragment library kit and an Ion Xpress barcode adapters 1-16 kit (Thermo). The median library size of 480 bp for the adapter-ligated 400-base-read library was size-selected with E-Gel SizeSelect II Agarose Gel (Invitrogen, California, USA). The sequencing was conducted on the Ion Torrent PGM sequencing platform using a Hi-Q sequencing kit (Thermo) and a 318 Chip (Thermo) as described previously.<sup>(18)</sup> The quality of the raw sequencing data was checked using FastQC v.0.11.9.<sup>(19)</sup> Raw sequence reads filtered with Trimmomatic v.0.38 (Phred score > 20) were used for further nucleotide variation analysis and *de novo* assembly with SPAdes v.3.14.1.<sup>(20,21)</sup> PhyResSe online tool was used to check for heteroresistance in the obtained WGS data.<sup>(22)</sup> The annotation of genomes was performed by either NCBI Prokaryotic Genome Annotation Pipeline (paired-end reads) or Prokka annotation pipeline v.1.14.5 (single-end reads).<sup>(23,24)</sup> Default parameters were used for all software unless otherwise specified.

An *in silico* spoligotyping was performed by SpolTyping 2.1 and the assignment of regions of deletion was accomplished by TB-Profiler.<sup>(25,26)</sup> The reads also were mapped to the reference genome H37Rv (Genbank accession no. NC\_000962.3) to confirm IS6110 insertions, and spoligoprofiles with a set of primers by Kamerbeek, J et al.<sup>(27)</sup> using Geneious Prime v.2019.2.1.

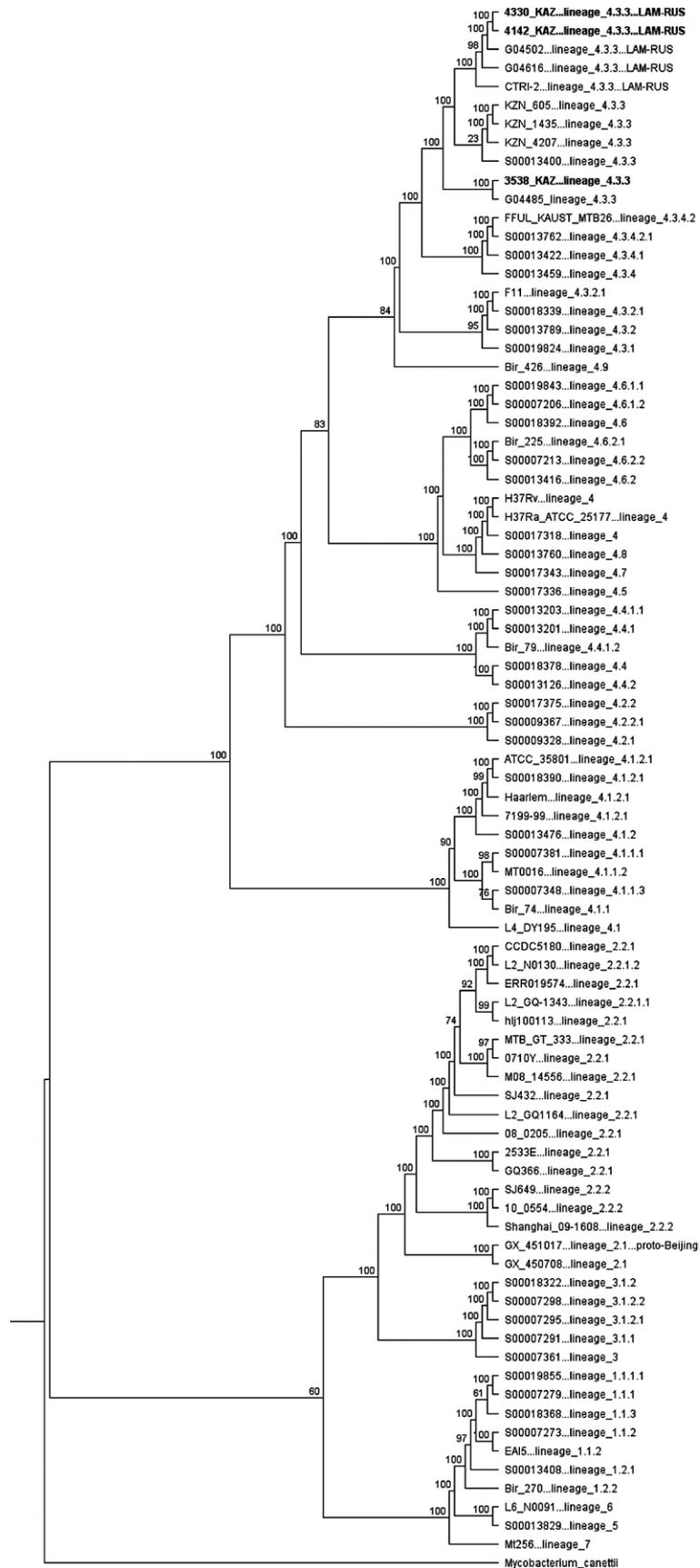
An SNP matrix was produced by comparing SNPs found between the studied genomes, and sequences of 81 MTBC isolates [Supplementary data (Table)]. The filtered sequence reads were mapped using the BWA-MEM program with *M. tuberculosis* H37Rv (NC\_000962.3) reference genome sequence.<sup>(28)</sup> SNPs were called by UnifiedGenotyper pipeline (GATK v.3.8.1.0).<sup>(29)</sup> SNPs located in repetitive genome regions, PE/PPE genes of the reference genome (NC\_000962.3) were filtered by TB Variant Filter v.0.1.3 before compiling the concatenated sequence.<sup>(30,31)</sup> Variant calls with per-base coverage

of fewer than 10x coverage depth or a Phred score below 20 were removed. The obtained high-confidence SNPs were subsequently written to a multi-FASTA alignment. A maximum-likelihood tree of all concatenated SNPs was generated using RAxML v8.2.11 with 100 bootstrap iterations.<sup>(32)</sup> We used the General Time Reversible (GTR) model of nucleotide substitution implemented in RAxML. The phylogeny was rooted using *Mycobacterium canettii* as an out-group. The phylogenetic tree was visualised with FigTree software v.1.4.4.

Characteristics of the draft whole-genome sequences of the LAM isolates are listed in Table I. The whole-genome shotgun sequencing data gave an average 110-fold genome coverage. MiSeq instrument generates sequence reads from both ends of a fragment (paired-end reading); while Ion Torrent produces single-end reads. Sequencing from both ends of fragment produce reads capable of accurate detection of genomic rearrangements and repetitive sequence elements.<sup>(33)</sup> The library preparation procedure took two days for both platforms, while MiSeq sequencing took a longer time to perform than single-end Ion Torrent technology (38 h versus 8 h). The lowest coverage (~ 84x) was obtained for single-end sequenced DNA from isolate no. 4330 that have resulted in a higher number of the assembled contigs, as shown in Table I.

Three isolates harbored high-confidence mutations in various genes associated with the drug resistance compared to the genome of reference strain H37Rv (GenBank accession no. NC\_000962.3) (Table II).<sup>(34)</sup> As follows from Table II, the studied isolates were not having a mixture of wild-type and mutant alleles also known as heteroresistance. No mixed calls were assigned in the alleles associated with the drug resistance with minority alleles composed more than 5% of the read depth. The isolates' phenotypic susceptibility was tested for isoniazid (INH), rifampin (RIF), streptomycin (SM), ethambutol (EMB), amikacin (AMI), kanamycin (KAN), and ofloxacin (OFX). WGS has confirmed phenotyping results for multi-drug resistant (MDR) isolates no. 3538 and 4330 and extensively drug-resistant (XDR) isolate no. 4142. Genotypic prediction of the *M. tuberculosis* susceptibility to anti-TB agents was found to correlate with phenotypic susceptibility.<sup>(35)</sup> It confirms predictive WGS-based drug-resistance profiling as a valuable tool for clinical use.

The prevalence rate of the LAM isolates in the current study corresponds to previously reported data (3/28 = 10.71%). The deletion of a large genomic region RD115 was observed in the three studied isolates (no. 3538, 4142, and 4330). Two out of three isolates were assigned to the LAM-RUS branch based on IS6110 insertion in the *plcA* gene (no. 4142 and 4330). Prevalence of the LAM-RUS isolates in local samples is especially intriguing since Lineage 4 is the most heterogeneous lineage of *M. tuberculosis*, consisting of 10 different sublineages, determined by the absence of specific RD loci called RD115, RD122, RD174, RD182, RD183, RD193, RD219, RD724, RD726, and RD761.<sup>(36)</sup> All three WGS genomes of the local LAM isolates sequenced by Stucki et al.<sup>(5)</sup> were also defined by the large deletion RD115 (Table III). Two of them were LAM-RUS isolates with a specific insertion



Maximum-likelihood phylogeny of three *Mycobacterium tuberculosis* isolates from this study and 81 representative genomes of *M. tuberculosis* complex (MTBC) strains. Sublineages are labeled according to Coll et al.<sup>(2)</sup> Branch lengths are proportional to nucleotide substitutions and the topology is rooted with *Mycobacterium canettii*. Three isolates from this study are highlighted. Bootstrap values are shown.

TABLE I  
Characteristics of three *Mycobacterium tuberculosis* genome assemblies

Isolate	Platform	SRA accession No.	Genome size (bp)	GC content (%)	Coverage (x)	No. of contigs	Total no. of CDSs	SNPs called*
3538	MiSeq**	SRR11241401	4,427,028	65.24	130	404	4,392	1455
4142	MiSeq	SRR11241400	4,351,634	65.58	116	243	4,237	1185
4330	Ion Torrent***	SRR11241402	4,344,962	65.00	84	925	4,495	989

\*: compared to the genome of reference strain H37Rv (GenBank accession no. NC\_000962.3); \*\*: MiSeq libraries were prepared using a Nextera DNA Flex Library Prep Kit and the sequencing was performed using MiSeq Reagent Kit v3; \*\*\*: Ion Torrent library was prepared using an Ion Xpress Plus fragment library kit and the sequencing was conducted using a 318 Chip. CDSs: coding sequences; GC: guanine-cytosine; SRA: sequence read archive.

TABLE II  
Mutations observed in drug resistance-associated loci of the *Mycobacterium tuberculosis* isolates

Isolate	Drug resistance							
	AMI	EMB	INH	OFX	PZA	RIF	SM	ETH
3538								
Gene			<i>katG</i>			<i>rpoB</i>		
AA change			Ser315Thr			Ser450Leu		
Base change			G/C 944			C/T 1349		
% of reads			100			100		
4142								
Gene	<i>rrs</i>	<i>embB</i>	<i>katG</i>	<i>gyrA</i>		<i>rpoB</i>	<i>rrs</i>	<i>fabG</i>
AA change	-	Met306Ile	Ser315Thr	Asp94Tyr		His445Leu	-	-
Base change	A/G 1401	G/C 918	G/C 944	G/T 280		A/T 1334	A/C 514	C/T -15
% of reads	100	100	100	100		99	96	100
4330								
Gene		<i>embB</i>	<i>katG</i>			<i>rpoB</i>	<i>rrs</i>	<i>fabG</i>
AA change		Met306Ile	Ser315Thr			His445Leu	-	-
Base change		G/C 918	G/C 944			A/T 1334	A/C 514	C/T -15
% of reads		98	99			100	99	100

AMI: amikacin; EMB: ethambutol; INH: isoniazid; OFX: ofloxacin; PZA: pyrazinamide; RIF: rifampin; SM: streptomycin; ETH: ethambutol.

of IS6110 into the *plcA* gene (G04493 and G04546).<sup>(8)</sup> Noteworthy, RD115 was present in all LAM-RUS isolates (Table III), while two isolates with RD115 did not belong to the LAM-RUS family. Geographic mapping of the LAM-RUS family shows its prevalence across Northern Eurasia, including Russia and Kazakhstan. It was also reported in other countries, such as Brazil, Venezuela, Ethiopia, and Sierra Leone.<sup>(37)</sup>

In the current study, an *in silico* spoligotyping has classified all three isolates as a spoligotype international type (SIT) 42. SIT42 is a major ancestral spoligotype of the LAM family also known as LAM prototype spoli-

goprofile.<sup>(37)</sup> These recent findings by Mokrousov et al.<sup>(37)</sup> indicate that SIT42 was highly dominated by the LAM-RUS in the collection of *M. tuberculosis* isolates from Russia, Belarus, and Kazakhstan. Additionally, we conducted *in silico* spoligotyping for the three local LAM isolates sequenced by Stucki et al.<sup>(5)</sup> Our *in silico* analysis suggests at least four populations of the LAM family sublineages, namely, SIT42, SIT42/LAM-RUS, SIT254/LAM-RUS, and SIT444/LAM-RUS, circulate in Kazakhstan based on the WGS data (Table III). Of these, two genotypes (SIT42 and SIT42/LAM-RUS) are associated with MDR/XDR phenotype.

TABLE III

Latin American-Mediterranean(LAM) sublineages and spoligotype international types (SITs) in *Mycobacterium tuberculosis* isolates from Kazakhstan based on the whole-genome sequences (WGS) data

Isolate	Year	Sublineage*	RD	SIT	IS6110 ( <i>plcA</i> gene)		Reference
					LAM-RUS	Resistance	
3538	2014	L4.3.3	RD115	42	-	MDR	This study
4142	2014	L4.3.3	RD115	42	+	XDR	This study
4330	2014	L4.3.3	RD115	42	+	MDR	This study
G04485	2002	L4.3.3	RD115	42	-	MDR	Stucki et al.
G04546	2007	L4.3.3	RD115	254	+	S	Stucki et al.
G04493	2011	L4.3.3	RD115	444	+	S	Stucki et al.

\*: sublineage classification based on Coll et al.<sup>(2)</sup>; MDR: multi-drug resistant isolate; XDR: extensively drug-resistant isolate; S: sensitive isolate.

In addition to spoligotyping, the presence of specific SNPs is another characteristic of the LAM isolates.<sup>(2,38)</sup> Three studied isolates harbored the same set of LAM family-specific SNPs including G/A transition in *fbpC* codon 103 (*Rv0129c*), G/A substitution in position 8,040 (*Rv0006*), C/T substitution in position 403,364 (*Rv0338c*), G/A substitution in position 2,518,919 (*Rv2245*) and C/G substitution in position 3,426,795 (*Rv3062*). As a result, lineage and sublineage of three studied isolates were assigned as 4.3.3 based on *in silico* spoligotyping results together with phylogenetic SNPs analysis.

Finally, we conducted a maximum-likelihood phylogenetic analysis. A phylogenetic tree was constructed based on overall SNPs extracted from 84 MTBC genomic DNA sequences including WGS data from this study [Figure, Supplementary data (Table)]. This data set included fifty published genomic sequences representing phylogeny of *M. tuberculosis* lineage 4 according to Coll et al.<sup>(2)</sup> classification. As a result, the LAM sublineage 4.3.3 was divided into several branches. The sequences no. 4142 and 4330 have formed a separate 4.3.3 LAM-RUS branch together with the published genomic sequences of isolates G04616 (Uzbekistan), G04502, and CTRI-2 (Russia), while no. 3538 has grouped with isolate G04485 (Kazakhstan). The phylogeny of the local LAM-RUS isolates fits an earlier hypothesis on the founding bacterial/human population in Northern Eurasia that was disseminated by large-scale human migration in the former Soviet Union.<sup>(37)</sup>

Contribution of the present data when comparing to earlier studies is WGS-based genotypic prediction of drug-resistance combined with epidemiological genotyping that covered *in silico* spoligotyping, assignment of RD loci as well as IS6110 insertions. The obtained genomes have provided new data on the genetic variations occurring in the drug-resistant LAM family isolates circulating in Central Asia. We found two MDR/XDR-associated genotypes of the LAM family circulating in Kazakhstan (SIT42 and SIT42/LAM-RUS). We are deeply concerned with the acquired extensive and multi-drug resistance in the local LAM family isolates. Further expansion of these drug-resistant clones rep-

resents a threat to regional TB control and the effectiveness of standardised treatment strategy. One of the limitations of the current research was the collection of genomic sequences that is not comprehensive to confirm the origin and migration pathways of the studied LAM isolates. A further comparison of the obtained WGS data with the other circulating LAM sublineage 4.3.3 isolates is needed to reveal the genetic differences responsible for the pathogenicity and transmission success of the LAM isolates.

**Accession numbers** - This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession no. JABACH000000000, JABACG000000000, and JABACF000000000. The raw Whole Genome Shotgun data was submitted to the NCBI SRA. SRA accession numbers are listed in Table I.

#### AUTHORS' CONTRIBUTION

PT and YR conceived the study; PT and AA designed the study protocol; AA assisted in the culture work; SA and PT performed the sequencing; PT and SA analysed the data; PT and SA wrote the manuscript. The authors declare no conflict of interest, financial, or other.

#### REFERENCES

- de Almeida IN, Vasconcellos SE, Figueredo LJA, Dantas NG, Augusto CJ, Hadaad JP, et al. Frequency of the *Mycobacterium tuberculosis* RD(Rio) genotype and its association with multidrug-resistant tuberculosis. BMC Infect Dis. 2019; 19(1): 556.
- Coll F, McNERNEY R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, et al. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. Nat Commun. 2014; 5: 4812.
- Sola C, Filliol I, Legrand E, Mokrousov I, Rastogi N. *Mycobacterium tuberculosis* phylogeny reconstruction based on combined numerical analysis with IS1081, IS6110, VNTR, and DR-based spoligotyping suggests the existence of two new phylogeographical clades. J Mol Evol. 2001; 53(6): 680-9.
- van Soolingen D, Qian L, de Haas PE, Douglas JT, Traore H, Portaels F, et al. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of East Asia. J Clin Microbiol. 1995; 33(12): 3234-8.
- Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, et al.

- Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat Genet.* 2016; 48(12): 1535-43.
6. Klotoe BJ, Kacimi S, Costa-Conceição E, Gomes HM, Barcellos RB, Panaiotov S, et al. Genomic characterization of MDR/XDR-TB in Kazakhstan by a combination of high-throughput methods predominantly shows the ongoing transmission of L2/Beijing 94-32 central Asian/Russian clusters. *BMC Infect Dis.* 2019; 19(1): 553.
  7. Lazzarini LC, Huard RC, Boechat NL, Gomes HM, Oelemann MC, Kurepina N, et al. Discovery of a novel *Mycobacterium tuberculosis* lineage that is a major cause of tuberculosis in Rio de Janeiro, Brazil. *J Clin Microbiol.* 2007; 45(12): 3891-902.
  8. Dubiley S, Kirillov E, Ignatova A, Stepanshina V, Shemyakin I. Molecular characteristics of the *Mycobacterium tuberculosis* LAM-RUS family prevalent in Central Russia. *J Clin Microbiol.* 2007; 45(12): 4036-8.
  9. WHO - World Health Organization. Global tuberculosis report 2019. Available from: [https://www.who.int/tb/publications/global\\_report/en/](https://www.who.int/tb/publications/global_report/en/).
  10. Merker M, Barbier M, Cox H, Rasigade JP, Feuerriegel S, Kohl TA, et al. Compensatory evolution drives multidrug-resistant tuberculosis in Central Asia. *eLife.* 2018; 7: e38200.
  11. Pillay M, Sturm AW. Evolution of the extensively drug-resistant F15/LAM4/KZN strain of *Mycobacterium tuberculosis* in KwaZulu-Natal, South Africa. *Clin Infect Dis.* 2007; 45(11): 1409-14.
  12. Skiba Y, Mokrousov I, Ismagulova G, Maltseva E, Yurkevich N, Bismilda V, et al. Molecular snapshot of *Mycobacterium tuberculosis* population in Kazakhstan: a country-wide study. *Tuberculosis (Edinb).* 2015; 95(5): 538-46.
  13. Mokrousov I, Vyazovaya A, Narvskaya O. *Mycobacterium tuberculosis* Latin American-Mediterranean family and its sublineages in the light of robust evolutionary markers. *J Bacteriol.* 2014; 196(10): 1833-41.
  14. Skiba Y, Mokrousov I, Nabirova D, Vyazovaya A, Maltseva E, Malakhova N, et al. *Mycobacterium tuberculosis* RD-Rio strain in Kazakhstan. *Emerg Infect Dis.* 2019; 25(3): 604-6.
  15. Ei PW, Aung WW, Lee JS, Choi GE, Chang CL. Molecular strain typing of *Mycobacterium tuberculosis*: a review of frequently used methods. *J Korean Med Sci.* 2016; 31(11): 1673-83.
  16. Mokrousov I. On sunspots, click science and molecular iconography. *Tuberculosis (Edinb).* 2018; 110: 91-5.
  17. van Soolingen D, Hermans PW, de Haas PE, Soll DR, van Embden JD. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J Clin Microbiol.* 1991; 29(11): 2578-86.
  18. Shevtsov A, Tarlykov P, Zholdybayeva E, Shevtsova E, Momykulov D, Sytnik I, et al. Draft genome sequence of the live vaccine strain *Brucella abortus* 82. *Genome Announc.* 2013; 1(6): e01101-13.
  19. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2019. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
  20. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *Journal of computational biology: a journal of computational molecular cell biology.* 2013; 20(10): 714-37.
  21. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014; 30(15): 2114-20.
  22. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, et al. PhyResSE: a web tool delineating *Mycobacterium tuberculosis* antibiotic resistance and lineage from whole-genome sequencing data. *J Clin Microbiol.* 2015; 53(6): 1908.
  23. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014; 30(14): 2068-69.
  24. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 2016; 44(14): 6614-24.
  25. Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med.* 2015; 7(1): 51.
  26. Xia E, Teo Y-Y, Ong RT-H. SpoTyping: fast and accurate *in silico* *Mycobacterium* spoligotyping from sequence reads. *Genome Med.* 2016; 8(1): 19.
  27. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol.* 1997; 35(4): 907-14.
  28. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010; 26(5): 589-95.
  29. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010; 20(9): 1297-303.
  30. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol.* 2004; 5(2): R12.
  31. Sekizuka T, Yamashita A, Murase Y, Iwamoto T, Mitarai S, Kato S, et al. TGS-TB: total genotyping solution for *Mycobacterium tuberculosis* using short-read whole-genome sequencing. *PLoS One.* 2015; 10(11): e0142951.
  32. Stamatakis A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014; 30(9): 1312-13.
  33. Nakazato T, Ohta T, Bono H. Experimental design-based functional mining and characterization of high-throughput sequencing data in the sequence read archive. *PLoS One.* 2013; 8(10): e77910.
  34. Rodwell TC, Valafar F, Douglas J, Qian L, Garfein RS, Chawla A, et al. Predicting extensively drug-resistant *Mycobacterium tuberculosis* phenotypes with genetic mutations. *J Clin Microbiol.* 2014; 52(3): 781-89.
  35. Allix-Beguec C, Arandjelovic I, Bi L, Beckert P, Bonnet M, Bradley P, et al. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. *N Engl J Med.* 2018; 379(15): 1403-15.
  36. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA.* 2006; 103(8): 2869-73.
  37. Mokrousov I, Vyazovaya A, Iwamoto T, Skiba Y, Pole I, Zhdanova S, et al. Latin-American-Mediterranean lineage of *Mycobacterium tuberculosis*: human traces across pathogen's phylogeography. *Mol Phylogenet Evol.* 2016; 99: 133-43.
  38. Homolka S, Projahn M, Feuerriegel S, Ubben T, Diel R, Nubel U, et al. High resolution discrimination of clinical *Mycobacterium tuberculosis* complex strains based on single nucleotide polymorphisms. *PLoS One.* 2012; 7(7): e39855.