Whole-Genome Analysis of *Mycobacterium tuberculosis* from Patients with Tuberculous Spondylitis, Russia

Ekaterina Chernyaeva, Mikhail Rotkevich, Ksenia Krasheninnikova, Andrey Yurchenko, Anna Vyazovaya, Igor Mokrousov, Natalia Solovieva, Viacheslav Zhuravlev, Piotr Yablonsky, Stephen J. O'Brien

Whole-genome analysis of *Mycobacterium tuberculosis* isolates collected in Russia (N = 71) from patients with tuberculous spondylitis supports a detailed characterization of pathogen strain distributions and drug resistance phenotype, plus distinguished occurrence and association of known resistance mutations. We identify known and novel genome determinants related to bacterial virulence, pathogenicity, and drug resistance.

Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis*, which typically affects the lungs but can affect other sites. In 2016, an estimated 10.4 million new TB cases and 1.6 million TB-related deaths were documented worldwide (*1*). The Russian Federation reported >120,000 TB cases and ≈13,700 TB deaths in 2016 (*1*). TB strains with multidrug resistance (MDR TB), characterized by resistance to isoniazid or rifampin, are common in the Russian Federation. The estimated rate of MDR TB was 27% among TB case-patients newly diagnosed in 2016 and 65% among previously treated case-patients in 2016 (*1*). Most TB cases are associated with pulmonary localization of the disease; however, in some cases, extrapulmonary TB develops. In Russia, the rate of extrapulmonary TB cases among new TB cases was 3.3% in 2014; most extrapulmonary TB cases are osteoarticular and genitourinary (*2*). Approximately 70% of osteoarticular TB cases are tuberculous spondylitis (TBS), which causes severe specific lesions of ≥1 components of the spine (*2*).

Author affiliations: St. Petersburg State University, St. Petersburg, Russia (E. Chernyaeva, M. Rotkevich, K. Krasheninnikova, A. Yurchenko, P. Yablonsky, S.J. O'Brien); St. Petersburg Research Institute of Phthisiopulmonology, St. Petersburg (E. Chernyaeva, N. Solovieva, V. Zhuravlev, P. Yablonsky); University of Glasgow, Glasgow, Scotland, UK (A. Yurchenko); St. Petersburg Pasteur Institute, St. Petersburg (A. Vyazovaya, I. Mokrousov); Nova Southeastern University, Ft. Lauderdale, Florida, USA (S.J. O'Brien)

DOI: https://doi.org/10.3201/eid2403.170151

We report whole-genome sequencing (WGS) and variant analyses of *M. tuberculosis* isolates from patients treated in Russia for TBS during 2007–2014.

The Study

The isolates were randomly collected from 71 TBS patients who received treatment at clinics of the Research Institute of Phthisiopulmonology in 32 regions of the Russian Federation (Figure 1). In these cases, *M. tuberculosis* isolates were cultured from extrapulmonary clinical material and stored. We assessed the susceptibility of these stored isolates to streptomycin, isoniazid, rifampin, ethambutol, pyrazinamide, ethionamide, ofloxacin, kanamycin, amikacin, cycloserine, capreomycin, and paraaminosalicylic acid according to World Health Organization recommendations (*3*). We isolated genomic DNA from cultured bacteria by using phenol/chloroform extraction and subjected bacterial DNA to WGS by using the MiSeq platform (Illumina, San Diego, CA, USA) to a mean coverage of 47× (range 18×–170×), covering ≥99% of the *M. tuberculosis* H37Rv reference genome (GenBank accession no. NC_000962.3). We deposited WGS reads in the NCBI Sequence Read Archive (accession no. PRJNA352769).

We aligned sequenced reads to the reference genome and called variants (single-nucleotide polymorphisms [SNPs] and short insertions/deletions) by using bioinformatics software: bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml); SAMtools (http://samtools.sourceforge.net); VCFtools (http://vcftools.sourceforge.net); and FreeBayes (https://github.com/ekg/freebayes). We used mutations that had q-scores ≥20 for comprehensive analysis. We used concatenated SNPs for phylogenetic analysis by using the GTRCAT (general time-reversible model with rate heterogeneity accommodated by using discrete rate categories) maximum-likelihood algorithm from the RAxML software package (*4*) to calculate an approximation model and 100 bootstrap replications. To avoid misalignments, we annotated SNPs in repetitive genome regions and in genes encoding proteins that contain proline-glutamate or proline-proline-glutamate motifs and filtered them from analysis. We used PhyTB (*5*) and SpoTyping tools (*6*) for phylogenetic classification of *M. tuberculosis* genomes and verified SpoTyping output by using previously conducted conventional spoligotyping analysis for 20 isolates that were previously described (*7*).

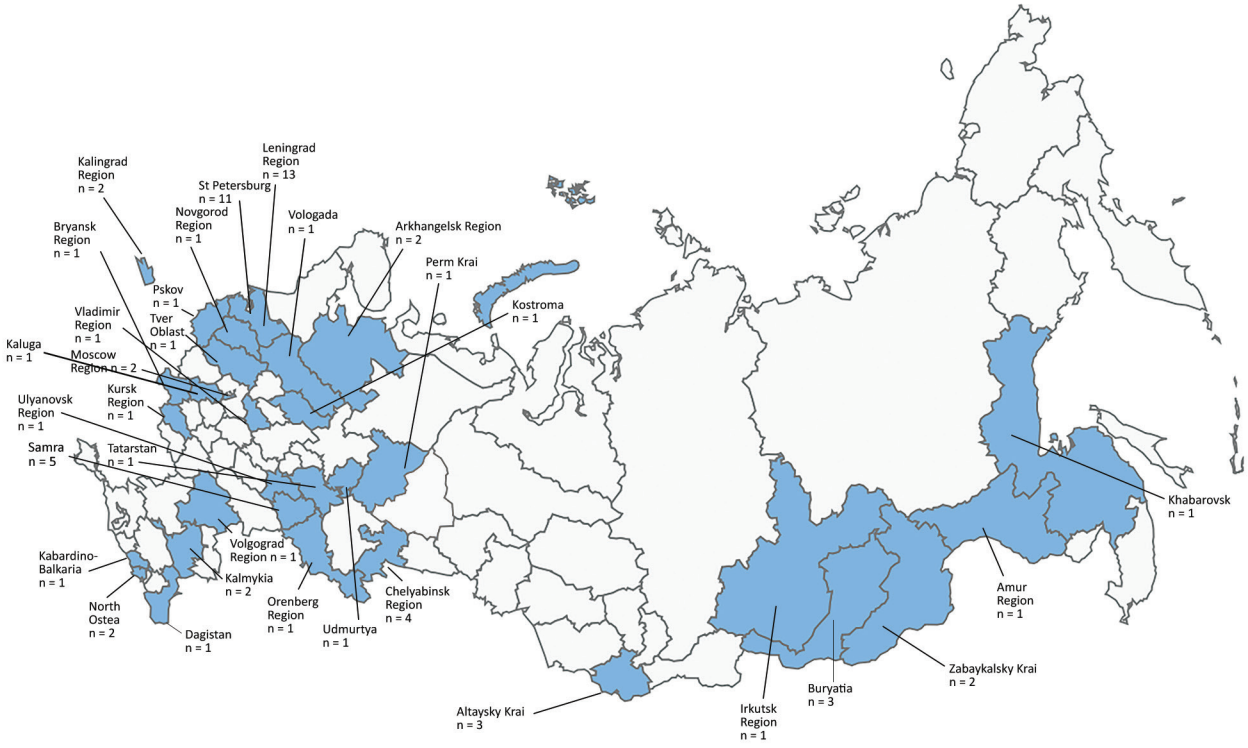

Figure 1. Distribution of *Mycobacterium tuberculosis* isolates randomly collected from 71 patients with tuberculous spondylitis who received treatment at clinics of the Research Institute of Phthisiopulmonology in 32 regions of the Russian Federation, 2007–2014.

We identified 2 principal phylogenetic lineages among *M. tuberculosis* isolates, lineage 2 and lineage 4; further, we detected evolutionary ancient and modern sublineages within major lineage 2 (Beijing; Figure 2) according to previously described classifications (*8*). The largest subgroup within the Beijing clade belonged to the B0/W148 clonal cluster (Figure 2) (*8*). Lineage 4 was represented by 4 genetic families: Ural, 4.2; Latin-American/Mediterranean (LAM), 4.3; and T, 4.1 and 4.8. The 58 Beijing genotype isolates contained 38 MDR (65.5%), 5 extensively drug-resistant (XDR; resistant to isoniazid and rifampin plus any fluoroquinolone and >1 of 3 injectable second-line drugs) (8.6%), 7 polyresistant but not MDR (12%), 1 monoresistant (1.9%), and 7 susceptible (12%) TB isolates. The MDR TB frequency in the Beijing group (65.5%) was higher than that for other genetic groups pooled ($p<0.0096$ by Fisher exact test). The *M. tuberculosis* Beijing B0/W148 cluster was represented by 1 susceptible (4.8%), 2 polyresistant (9.5%), 3 XDR (14.3%), and 15 MDR (71.4%) TB isolates. The B0/W148 genetic group demonstrates an association with MDR TB ($p = 0.03$), shown previously (*9*,*10*). The other genetic groups (T, LAM, and Ural) included too few isolates to test for association with MDR TB.

Specimens of 50 TBS patients were HIV negative; 21 were HIV positive. Although we found no significant association of *M. tuberculosis* genetic groups to HIV infection, 42% of patients infected by B0/W148 strains were HIV positive whereas among patients infected by non-B0/W148 Beijing strains, only 22% were HIV positive. Further, only 14% patients infected with non-Beijing *M. tuberculosis* strains were HIV positive (online Technical Appendix Table 1, https://wwwnc.cdc.gov/EID/article/24/3/17-0151-Techapp1.pdf).

We examined *M. tuberculosis* isolates for the presence of published variants associated with resistance to TB drugs (Table). We found a high level of concordance of phenotypic and genetic data for reported isoniazid- and rifampin-resistant isolates. We detected mutations in *rpsL*, *gid*, and *rrs* genes in 96.4% of streptomycin-resistant isolates; 81.8% of ofloxacin-resistant isolates had mutations in *gyrA* gene (there were no mutations in *gyrB* gene). Most ethambutol-resistant isolates (72.7%) showed mutations in the *embA* promoter region or *embB* region between codons 296 and 497. However, 3 ethambutol-susceptible isolates had mutations M306I (n = 1) and G406A (n = 2) in the *embB* gene. We detected mutations in genes *pncA* (*11*) and *rpsA*, associated with pyrazinamide resistance in 55.6% of pyrazinamide-resistant isolates, and 55% of kanamycin-resistant strains had mutations in the *eis* promoter. We detected no mutations in *alr* and *ddl* genes among cycloserine-resistant isolates, nor in *thyA* gene among paraaminosalicylic acid–resistant isolates.

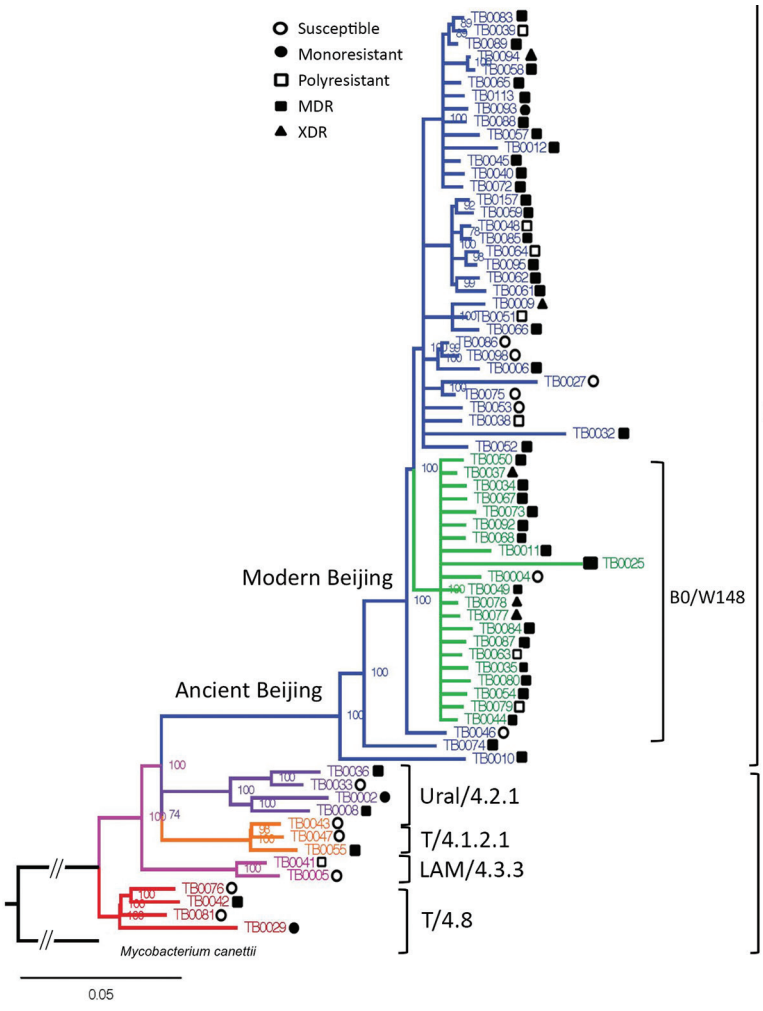

Figure 2. Phylogenetic analysis of *Mycobacterium tuberculosis* isolates obtained from patients with tuberculous spondylitis, Russian Federation, 2007–2014. Symbols indicate drug resistance: susceptible to all tested drugs; monoresistant to 1 drug; polyresistant, resistant to multiple drugs other than isoniazid or rifampin; MDR, resistant to isoniazid or rifampin; XDR, resistant to isoniazid or rifampin plus any fluoroquinolone and >1 of 3 injectable second-line drugs. Number of isolates and parsimony informative sites by genetic group: All groups, 71/2,656; Beijing, 58/919; Beijing B0/W148, 21/213; Beijing non-B0/W148, 37/722; T (4.8), 4/38; Ural, 4/204. The truncated root branch connects the studied *M. tuberculosis* lineages with an outgroup represented by *M. cannetii*. Scale bar indicates nucleotide substitutions per site.

Our analysis for small insertions and deletions detected 15 and 9, respectively, among the Beijing group (online Technical Appendix Table 2). A deletion in *kdpD* and an insertion in *Rv1258c* were previously described (*12,13*). The other 22 mutations are novel: 18 were specific to the Beijing group; 2 to the modern Beijing group; 1 to the ancient Beijing group; and 3 to the B0/W148 group. We identified most mutations in genes encoding membrane-associated proteins, although several mutations were in regulatory genes, genes involved in cell metabolism, probable transposase genes, and genes with unknown function.

One insertion and 2 deletions were significantly associated with B0/W148 genetic group in *kdpD, mmr* and Rv1995 ($p = 2.5 \times 10^{-17}$) genes. Merker et al. (*12*), who proposed a pathogenic influence for B0/W148 strains, described a frameshift deletion in *kdpD* among Beijing B0/W148 strains. Deletion in the *kdpD* gene can lead to the formation of nonfunctional proteins KdpD and KdpE. Parish et al. showed that *M. tuberculosis* lacking KdpD and KdpE function express increased virulence in a mouse model of infection (*14*), which supports that the *kdpD* deletion detected in our study may influence Beijing B0/W148 strain's rapid expansion and virulence. A mutation in the promoter region of the *mmr* gene (*Rv3065*), encoding multidrug-transport integral membrane protein, might contribute to drug resistance in Beijing B0/W148 strains. Sriraman et al. recently showed that *mmr* is upregulated in rifampin-resistant and MDR TB strains, even in the presence of target gene mutations (*15*). Insertion in *Rv1258c* is common to all Beijing strains except ancient TB0010. In their study, Villellas et al. found the cytosine nucleotide insertion between positions 580 and 581 in the *Rv1258c* gene in all Beijing isolates among streptomycin-resistant *M. tuberculosis* strains (*13*).

In conclusion, we examined the phylogenetic and drug-resistance properties of *M. tuberculosis* isolates collected from 71 TBS patients in 32 locales across Russia. Our analyses confirmed the phylogenetic separation of pathogenic *M. tuberculosis* strains and support the prevalence of Beijing strains showing high levels of multidrug

Table. Mutations associated with drug resistance detected in *Mycobacterium tuberculosis* genomes

Tuberculosis drug	Gene variant	No. strains with confirmed drug resistance	Drug-resistant strains, %
Isoniazid	*katG* S315T	52	100
	fabG-inhA-15	1	
	katG S315T	2	
	fabG-inhA-15	2	
Rifampin	*rpoB* S450L	40	100
	rpoB D435V	2	
	rpoB H445D	1	
	rpoB H445R	1	
	rpoB D574E	1	
Streptomycin	*rpsL* K43R	33	96.4
	rpsL K88R	5	
	gid G48G	1	
	gid G34G	1	
	*rrs*516	12	
Ofloxacin	*gyrA* D94N	1	81.8
	gyrA D94Y	1	
	gyrA D94G	5	
	gyrA D94A	2	
Ethambutol	*embA*-16	4	72.7
	embA-8	1	
	embB M3061	2	
	embB S3471	1	
	embB N399T	1	
	embB G406D	2	
	embB G406A	2	
	embB A453A	1	
	embB Q497R	2	
Pyrazinamide	*rpsA* D123A	1	55.6
	rpsA A412V	1	
	pncA L159R	1	
	pncA C138R	1	
	pncA T135P	1	
	pncA V130E	1	
	pncA Q122Stop	1	
	pncA Y103Stop	2	
	pncA G97S	1	
Kanamycin	*eis*-37	4	55.0
	eis-14	3	
	eis-10	4	

resistance among TBS isolates. Further, we found known SNP variants that had high concordance with suggested drug resistance. Finally, novel insertions/deletions were apparent, which we suggest are candidates for conferring drug resistance pending independent replication studies. Our analysis of WGS data identified known and novel genetic determinants that could or do influence bacterial virulence, pathogenicity, and drug resistance.

This work was supported by the Russian Foundation for Basic Research (grant 16-34-60163) and St. Petersburg State University (grants 1.38.253.2015 and 1.52.1647.2016).

E.C., V.Z., and S.O.B. designed the experiment. N.S., V.Z., and P.Y. provided *M. tuberculosis* genomic DNA, information on drug resistance, and clinical data. E.C. performed whole-genome sequencing. E.C., M.R., K.K., and A.Y. performed bioinformatics and statistical analysis of whole-genome sequencing data. A.V. and I.M. conducted *M. tuberculosis* spoligotyping. E.C., I.M., and S.O.B. wrote the manuscript.

About the Author

Dr. Chernyaeva is a senior researcher at Theodosios Dobzhansky Center for Genome Bioinformatics at St. Petersburg State University, St. Petersburg, Russia. Her main research focuses on *M. tuberculosis* comparative genomics to discover genetic markers associated with microbiological and clinical features.

References

1. World Health Organization. Global tuberculosis report 2017. Geneva: The Organization; 2017 [cited: Nov. 27, 2017]. http://apps.who.int/iris/bitstream/10665/259366/1/9789241565516-eng.pdf?ua=1
2. Yablonsky P., Mushkin A., Belilovsky E., Galkin V. Extrapulmonary tuberculosis. In: Tuberculosis in the Russian Federation 2012/2013/2014 [in Russian]. Moscow. Triada; 2015; 129–35 [cited Nov. 27, 2017].
3. World Health Organization. Guidelines for surveillance of drug resistance in tuberculosis. 4th Edition. Geneva: The Organization; 2009.
4. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.

Bioinformatics. 2006;22:2688–90. http://dx.doi.org/10.1093/bioinformatics/btl446
5. Benavente ED, Coll F, Furnham N, McNerney R, Glynn JR, Campino S, et al. PhyTB: Phylogenetic tree visualisation and sample positioning for M. tuberculosis. BMC Bioinformatics. 2015;16:155. http://dx.doi.org/10.1186/s12859-015-0603-3
6. Xia E, Teo YY, Ong RT. SpoTyping: fast and accurate *in silico* mycobacterium spoligotyping from sequence reads. Genome Med. 2016;8:19. http://dx.doi.org/10.1186/s13073-016-0270-7
7. 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;35:907–14.
8. Mokrousov I. Insights into the origin, emergence, and current spread of a successful Russian clone of *Mycobacterium tuberculosis*. Clin Microbiol Rev. 2013;26:342–60. http://dx.doi.org/10.1128/CMR.00087-12
9. Vyazovaya A, Mokrousov I, Solovieva N, Mushkin A, Manicheva O, Vishnevsky B, et al. Tuberculous spondylitis in Russia and prominent role of multidrug-resistant clone *Mycobacterium tuberculosis* Beijing B0/W148. Antimicrob Agents Chemother. 2015;59:2349–57. http://dx.doi.org/10.1128/AAC.04221-14
10. Lasunskaia E, Ribeiro SC, Manicheva O, Gomes LL, Suffys PN, Mokrousov I, et al. Emerging multidrug resistant *Mycobacterium tuberculosis* strains of the Beijing genotype circulating in Russia express a pattern of biological properties associated with enhanced virulence. Microbes Infect. 2010;12:467–75. http://dx.doi.org/10.1016/j.micinf.2010.02.008
11. Yadon AN, Maharaj K, Adamson JH, Lai YP, Sacchettini JC, Ioerger TR, et al. A comprehensive characterization of PncA polymorphisms that confer resistance to pyrazinamide. Nat Commun. 2017;8:588. http://dx.doi.org/10.1038/s41467-017-00721-2
12. Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, et al. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. Nat Genet. 2015; 47:242–9. http://dx.doi.org/10.1038/ng.3195
13. Villellas C, Aristimuño L, Vitoria MA, Prat C, Blanco S, García de Viedma D, et al. Analysis of mutations in streptomycin-resistant strains reveals a simple and reliable genetic marker for identification of the *Mycobacterium tuberculosis* Beijing genotype. J Clin Microbiol. 2013;51:2124–30. http://dx.doi.org/10.1128/JCM.01944-12
14. Parish T, Smith DA, Kendall S, Casali N, Bancroft GJ, Stoker NG. Deletion of two-component regulatory systems increases the virulence of *Mycobacterium tuberculosis*. Infect Immun. 2003;71:1134–40. http://dx.doi.org/10.1128/IAI.71.3.1134-1140.2003
15. Sriraman K, Nilgiriwala K, Saranath D, Chatterjee A, Mistry N. Deregulation of genes associated with alternate drug resistance mechanisms in *Mycobacterium tuberculosis*. Curr Microbiol. 2017. http://dx.doi.org/10.1007/s00284-017-1393-9

Address for correspondence: Ekaterina Chernyaeva, St. Petersburg State University, Theodosius Dobzhansky Center for Genome Bioinformatics, 41 Sredniy Prospect, St. Petersburg 199034, Russian Federation; email: echernya@gmail.com

Article DOI: https://doi.org/10.3201/eid2403.170151

Genome-wide Analysis of *Mycobacterium tuberculosis* from Patients with Tuberculous Spondylitis, Russia

Technical Appendix

Technical Appendix Table 1. *Mycobacterium tuberculosis* isolate data

Strain No.	Major lineage	Genotype/ PhyTB Barcode	SpoTyping SIT	Verified SIT (if available)	Geographic region	HIV status	Year of isolate extraction	SM	INH	RIF	ETH	EMB	KM	OFL	PAS	CS	CM	AM	PZA
TB0002	4	Ural/4.2.1	777	ND	St. Petersburg	–	2011	S	S	S	S	R	S	S	S	S	S	S	S
TB0004	2	Beijing (B0/W148)/ 2.2.1	1	1	St. Petersburg	+	2011	S	S	S	S	S	S	S	S	S	S	S	S
TB0005	4	LAM/4.3.3	ND	ND	Leningrad Region	+	2011	S	S	S	S	S	S	S	S	S	S	S	S
TB0006	2	Beijing/2.2.1	1	1	Chelyabinsk Region	–	2011	R	R	R	R	R	S	S	S	S	S	S	R
TB0008	4	Ural/4.2.1	262	262	Kalmykia	–	2009	R	R	R	R	R	S	S	S	S	S	S	R
TB0009	2	Beijing/2.2.1	1	1	Kaliningrad Region	+	2011	R	R	R	R	S	S	R	S	S	R	R	R
TB0010	2	Beijing/2.2.1	269	269	Buryatia	–	2010	R	R	R	S	S	ND	S	ND	ND	S	S	R
TB0011	2	Beijing (B0/W148)/ 2.2.1	1	1	Volgograd Region	–	2010	R	R	R	R	R	R	S	S	S	R	R	R
TB0012	2	Beijing/2.2.1	1	1	Arkhangelsk Region	–	2010	R	R	R	R	S	R	S	S	S	S	S	R
TB0025	2	Beijing (B0/W148)/ 2.2.1	1	1	Leningrad Region	+	2011	R	R	R	R	R	R	S	S	S	R	R	S
TB0027	2	Beijing/2.2.1	1	ND	Novgorod Region	+	2010	S	S	S	S	S	S	S	S	S	S	S	S
TB0029	4	T/4.8	53	53	Ulyanovsk Region	+	2011	S	S	S	S	S	S	S	ND	ND	R	S	S
TB0032	2	Beijing/2.2.1	1	ND	Zabaykalsky Krai	+	2012	R	R	R	S	S	S	S	ND	ND	S	S	R

Strain No.	Major lineage	Genotype/ PhyTB Barcode	SpoTyping SIT	Verified SIT (if available)	Geographic region	HIV status	Year of isolate extraction	SM	INH	RIF	ETH	EMB	KM	OFL	PAS	CS	CM	AM	PZA
TB0033	4	Ural/4.2.1	1050	1050	Kalmykia	–	2011	S	S	S	S	S	S	S	S	S	S	ND	ND
TB0034	2	Beijing (B0/W148)/ 2.2.1	1	1	St. Petersburg	–	2011	R	R	R	R	R	R	S	R	S	R	R	R
TB0035	2	Beijing (B0/W148)/ 2.2.1	1	1	Leningrad Region	+	2010	R	R	R	R	S	S	S	S	S	S	S	R
TB0036	4	Ural/4.2.1	1134	ND	Leningrad Region	–	2011	R	R	R	S	S	S	S	R	S	S	S	ND
TB0037	2	Beijing (B0/W148)/ 2.2.1	1	1	Kabardino-Balkaria	–	2011	R	R	R	R	R	R	R	S	S	S	S	R
TB0038	2	Beijing/2.2.1	1	1	Amur Region	–	2010	R	R	S	R	S	R	S	S	S	R	R	ND
TB0039	2	Beijing/2.2.1	1	1	Kaluga	–	2009	R	R	S	R	R	R	S	S	S	ND	ND	S
TB0040	2	Beijing/2.2.1	1	ND	Samara	–	2012	R	R	R	R	S	R	S	S	S	S	R	R
TB0041	4	LAM/4.3.3	42	42	Leningrad Region	–	2008	R	R	S	S	R	S	S	R	S	S	S	S
TB0042	4	T/4.8	53	53	Vladimir Oblast	–	2012	R	R	R	S	R	S	S	S	S	R	S	R
TB0043	4	T/ 4.1.2.1	37	ND	St. Petersburg	–	2008	S	S	S	S	S	S	S	ND	ND	ND	ND	ND
TB0044	2	Beijing (B0/W148)/ 2.2.1	1	1	North Osetia	–	2011	R	R	R	S	S	R	S	R	S	S	S	R
TB0045	2	Beijing/2.2.1	1	1	Leningrad Region	–	2008	R	R	R	R	S	R	S	S	S	S	ND	S
TB0046	2	Beijing/2.2.1	1	1	St. Petersburg	–	2009	S	S	S	S	S	S	S	ND	ND	ND	ND	ND
TB0047	4	T/ 4.1.2.1	334	ND	Udmurtya	–	2008	S	S	S	S	S	S	S	S	S	ND	ND	ND
TB0048	2	Beijing/2.2.1	1	ND	St. Petersburg	–	2007	R	R	S	R	S	S	S	S	S	ND	ND	ND
TB0049	2	Beijing (B0/W148)/ 2.2.1	1	ND	St. Petersburg	–	2007	R	R	R	S	R	S	S	S	S	ND	ND	ND
TB0050	2	Beijing (B0/W148)/ 2.2.1	1	ND	Samara	–	2007	R	R	R	S	S	R	S	S	S	ND	ND	ND
TB0051	2	Beijing/2.2.1	1	ND	Kostroma	–	2009	R	R	S	S	S	S	S	S	S	S	ND	ND
TB0052	2	Beijing/2.2.1	1	ND	Leningrad Region	+	2009	R	R	R	R	S	R	S	S	S	S	S	blank
TB0053	2	Beijing/2.2.1	1	ND	St. Petersburg	–	2008	S	S	S	S	S	S	S	ND	ND	ND	ND	ND
TB0054	2	Beijing (B0/W148)/ 2.2.1	1	ND	Khabarovsk	–	2007	R	R	R	S	R	S	S	R	S	ND	ND	ND
TB0055	4	T/ 4.1.2.1	53	ND	Pskov	–	2010	R	R	R	S	R	S	S	S	S	ND	ND	ND

Strain No.	Major lineage	Genotype/ PhyTB Barcode	SpoTyping SIT	Verified SIT (if available)	Geographic region	HIV status	Year of isolate extraction	SM	INH	RIF	ETH	EMB	KM	OFL	PAS	CS	CM	AM	PZA
TB0057	2	Beijing/2.2.1	1	ND	Samara	–	2013	R	R	R	S	R	S	R	S	S	S	S	R
TB0058	2	Beijing/2.2.1	1	ND	Moscow Region	–	2013	R	R	R	S	S	S	R	S	S	S	S	ND
TB0059	2	Beijing/2.2.1	1	ND	Kaliningrad Region	–	2013	R	R	R	R	R	S	S	S	S	S	S	S
TB0061	2	Beijing/2.2.1	1	ND	Chelyabinsk Region	–	2013	R	R	R	S	S	S	S	S	S	S	S	ND
TB0062	2	Beijing/2.2.1	1	ND	Chelyabinsk Region	–	2013	R	R	R	S	S	S	S	S	S	S	S	ND
TB0063	2	Beijing (B0/W148)/ 2.2.1	1	ND	North Osetia	–	2013	R	R	S	R	S	S	S	S	S	S	S	ND
TB0064	2	Beijing/2.2.1	1	ND	Vologda Region	–	2013	S	R	S	S	S	S	S	S	S	R	R	S
TB0065	2	Beijing/2.2.1	1	ND	Leningrad Region	–	2013	R	R	R	S	S	R	S	S	S	S	S	S
TB0066	2	Beijing/2.2.1	1	ND	St. Petersburg	–	2013	R	R	R	R	R	S	S	S	S	S	S	S
TB0067	2	Beijing (B0/W148)/ 2.2.1	1	ND	Leningrad Region	+	2013	R	R	R	S	S	R	S	S	S	S	S	S
TB0068	2	Beijing (B0/W148)/ 2.2.1	1	ND	Altaysky Krai	–	2013	R	R	R	R	S	R	S	S	S	R	R	ND
TB0072	2	Beijing/2.2.1	1	ND	Samara	–	2014	R	R	R	S	S	S	S	R	S	S	S	ND
TB0073	2	Beijing (B0/W148)/ 2.2.1	1	ND	Tatarstan	+	2014	R	R	R	S	S	S	S	S	S	S	S	ND
TB0074	2	Beijing/2.2.1	1	ND	St. Petersburg.	+	2014	R	R	R	R	S	R	S	S	S	R	R	R
TB0075	2	Beijing/2.2.1	1	ND	St. Petersburg.	+	2014	S	S	S	S	S	S	S	S	S	S	S	ND
TB0076	4	T /4.8	52	ND	Perm Krai	–	2014	S	S	S	S	S	S	S	S	S	S	S	ND
TB0077	2	Beijing (B0/W148)/ 2.2.1	1	ND	Leningrad Region	+	2014	R	R	R	S	R	R	R	R	S	R	R	ND
TB0078	2	Beijing (B0/W148)/ 2.2.1	1	ND	Tver Oblast	+	2014	R	R	R	R	S	S	R	R	S	R	S	ND
TB0079	2	Beijing (B0/W148)/ 2.2.1	1	ND	Kursk Region	–	2014	R	R	S	R	S	S	S	R	S	S	S	ND

Strain No.	Major lineage	Genotype/ PhyTB Barcode	SpoTyping SIT	Verified SIT (if available)	Geographic region	HIV status	Year of isolate extraction	SM	INH	RIF	ETH	EMB	KM	OFL	PAS	CS	CM	AM	PZA
TB0080	2	Beijing (B0/W148)/ 2.2.1	1	ND	Bryansk Region	–	2014	R	R	R	R	R	S	S	S	S	S	S	R
TB0081	4	T/4.8	131	ND	Leningrad Region	–	2014	S	S	S	S	S	S	S	S	S	S	S	ND
TB0083	2	Beijing/2.2.1	1	ND	Altaysky Krai	–	2014	R	R	R	R	R	S	S	S	S	S	S	ND
TB0084	2	Beijing (B0/W148)/ 2.2.1	1	ND	Altaysky Krai	+	2014	R	R	R	S	S	R	S	R	S	S	R	ND
TB0085	2	Beijing/2.2.1	1	ND	Leningrad Region	–	2014	R	R	R	S	S	S	S	S	S	S	S	ND
TB0086	2	Beijing/2.2.1	1	ND	Buryatia	–	2014	S	S	S	S	S	S	S	S	S	S	S	ND
TB0087	2	Beijing (B0/W148)/ 2.2.1	1	ND	Irkutsk Region	+	2014	R	R	R	S	S	S	S	R	S	S	S	ND
TB0088	2	Beijing/2.2.1	1	ND	Samara	–	2014	R	R	R	S	S	S	R	S	S	S	S	ND
TB0089	2	Beijing/2.2.1	1	ND	Leningrad Region	+	2014	R	R	R	S	R	R	S	R	S	R	S	ND
TB0092	2	Beijing (B0/W148)/ 2.2.1	1	ND	Chelyabinsk Region	–	2014	R	R	R	R	S	R	R	S	S	S	S	ND
TB0093	2	Beijing/2.2.1	1	ND	Arkhangelsk Region	+	2014	R	S	S	S	S	S	S	S	S	S	S	ND
TB0094	2	Beijing/2.2.1	1	ND	Moscow Region	+	2014	R	R	R	R	R	R	R	S	S	S	S	ND
TB0095	2	Beijing/2.2.1	1	ND	Buryatia	+	2014	R	R	R	S	S	S	S	S	S	S	S	ND
TB0098	2	Beijing/2.2.1	1	ND	Zabaykalsky Krai	–	2014	S	S	S	S	S	S	S	S	S	S	S	S
TB0113	2	Beijing/2.2.1	1	ND	Orenburg Region	–	2014	R	R	R	R	R	S	R	S	S	S	S	R
TB0157	2	Beijing/2.2.1	1	ND	Dagestan	–	2014	R	R	R	S	S	R	S	S	S	R	R	R

*Phy TB, phylogenic analysis of Mycobacterium tuberculosis; spo, spondylitis; SIT, strains isolated in patients; HIV, human immunodeficiency virus; SM, streptomycin; INH, isoniazid; RIF, rifampicin; ETH, ethionamide; EMB, ethambutol; KM, kanamycin; OFL, ofloxacin; PAS, para-aminosalicylic acid; CS, cycloserine; CM, capreomycin; AM, amikacin; PZA, pyrazinamide; + positive; – negative; S ,drug susceptible; R, drug resistant; ND, no data.

Technical Appendix Table 2. Insertions and deletions associated with *M. tuberculosis* genetic clades

Isolate no.	Genetic clade	Gene position	Genome coordinates	Reference sequence	Alternative sequence	Effect	Genome region	Gene name
1	Beijing	682	99162	TC	TCGGTGTGCGC	Deletion	CDS	Rv0090
2	Beijing	511	194305	CGGC	CC	Deletion	CDS	mce1R
3	Beijing	-56	507028	GCCCCCG	GCCCCCCG	Insertion	Intergenic	Rv0420c
4	Beijing	-12	1090188	AGGG	AGG	Deletion	Intergenic	Rv0976c
5	Beijing	336	1168009	GCCCCG	GCCCCCG	Insertion	CDS	Rv1045
6	Beijing	581	1406760	TGC	TC	Deletion	CDS	Rv1258c
7	Beijing	96	1413129	CGGGGGC	CGGGGGGC	Insertion	Non-coding RNA	mcr11
8	Beijing	118	2009290	GCCCCA	GCCCCCA	Insertion	CDS	Rv1775
9	Beijing	872	2241031	AG	AGG	Insertion	CDS	ctpF
10	Beijing	841	2342649	AGGCGTACACACG	AG	Deletion	CDS	Rv2084
11	Beijing	105	2734481	TC	TAC	Insertion	CDS	Rv2437
12	Beijing	53	3021892	GTTCGACGACT	GTTCGACGACTTCGACGACT	Insertion	CDS	Rv2709
13	Beijing	1,308	3194241	CGGC	CGC	Deletion	CDS	Rv2885c
14	Beijing	381	3238119	CGTGC	CGTGGTGC	Insertion	CDS	Rv2923c
15	Beijing	1	3794867	CCAC	CC	Deletion	CDS	dxs2
16	Beijing	-19	4221070	AC	ACC	Insertion	Intergenic	Rv3776
17	Beijing	623	4305063	GAAA	GAA	Deletion	CDS	Rv3830c
18	Beijing	501	4322039	AC	ACC	Insertion	CDS	Rv3847
19	Beijing B0/W148	2,545	1149142	CTGT	CT	Insertion	CDS	kdpD
20	Beijing B0/W148	720	2238861	GAAAAAAG	GAAAAAG	Insertion	CDS	Rv1995
21	Beijing B0/W148	-29	3430358	CATGTACAAA	CATGTACAAATGTACAAA	Deletion	Intergenic	mmr
22	Modern Beijing	91	1418863	CGGGAGCCAG	CGGGAGCCAGGGAGCCAG	Deletion	CDS	Rv1269c
23	Modern Beijing	633	2703902	CGGC	CGC	Deletion	Intergenic	Rv2405
24	Ancient Beijing	-106	809840	GCCCCCCA	GCCCCCCCA	Deletion	Intergenic	Rv0713