

# Unusual Large-Scale Chromosomal Rearrangements in *Mycobacterium tuberculosis* Beijing B0/W148 Cluster Isolates

Egor A. Shitikov<sup>1\*</sup>, Julia A. Bespyatykh<sup>1</sup>, Dmitry S. Ischenko<sup>1,2</sup>, Dmitry G. Alexeev<sup>1,2</sup>, Irina Y. Karpova<sup>1</sup>, Elena S. Kostryukova<sup>1</sup>, Yulia D. Isaeva<sup>3</sup>, Elena Y. Nosova<sup>3</sup>, Igor V. Mokrousov<sup>4</sup>, Anna A. Vyazovaya<sup>4</sup>, Olga V. Narvskaya<sup>4</sup>, Boris I. Vishnevsky<sup>5</sup>, Tatiana F. Otten<sup>5</sup>, Valery Y. Zhuravlev<sup>5</sup>, Peter K. Yablonsky<sup>5</sup>, Elena N. Ilina<sup>1</sup>, Vadim M. Govorun<sup>1</sup>

**1** Research Institute of Physical - Chemical Medicine, Moscow, Russian Federation, **2** Moscow Institute of Physics and Technology, Dolgoprudny, Russian Federation, **3** Moscow Scientific-Practical Center of Treatment of Tuberculosis of Moscow Healthcare, Moscow, Russian Federation, **4** St. Petersburg Pasteur Institute, St. Petersburg, Russian Federation, **5** Research Institute of Phthisiopulmonology, St. Petersburg, Russian Federation

## Abstract

The *Mycobacterium tuberculosis* (MTB) Beijing family isolates are geographically widespread, and there are examples of Beijing isolates that are hypervirulent and associated with drug resistance. One-fourth of Beijing genotype isolates found in Russia belong to the B0/W148 group. The aim of the present study was to investigate features of these endemic strains on a genomic level. Four Russian clinical isolates of this group were sequenced, and the data obtained was compared with published sequences of various MTB strain genomes, including genome of strain W-148 of the same B0/W148 group. The comparison of the W-148 and H37Rv genomes revealed two independent inversions of large segments of the chromosome. The same inversions were found in one of the studied strains after deep sequencing using both the fragment and mate-paired libraries. Additionally, inversions were confirmed by RFLP hybridization analysis. The discovered rearrangements were verified by PCR in all four newly sequenced strains in the study and in four additional strains of the same Beijing B0/W148 group. The other 32 MTB strains from different phylogenetic lineages were tested and revealed no inversions. We suggest that the initial largest inversion changed the orientation of the three megabase (Mb) segment of the chromosome, and the second one occurred in the previously inverted region and partly restored the orientation of the 2.1 Mb inner segment of the region. This is another remarkable example of genomic rearrangements in the MTB in addition to the recently published of large-scale duplications. The described cases suggest that large-scale genomic rearrangements in the currently circulating MTB isolates may occur more frequently than previously considered, and we hope that further studies will help to determine the exact mechanism of such events.

**Citation:** Shitikov EA, Bespyatykh JA, Ischenko DS, Alexeev DG, Karpova IY, et al. (2014) Unusual Large-Scale Chromosomal Rearrangements in *Mycobacterium tuberculosis* Beijing B0/W148 Cluster Isolates. PLoS ONE 9(1): e84971. doi:10.1371/journal.pone.0084971

**Editor:** Philip Supply, Institut Pasteur de Lille, France

**Received:** March 15, 2013; **Accepted:** November 28, 2013; **Published:** January 8, 2014

**Copyright:** © 2014 Shitikov et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was partly supported by 16.522.11.2003 grant of the Ministry of Education and Science of the Russian Federation; and by the European Commission (ORCHID project, Grant Agreement No. 261378). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Igor Mokrousov is Academic Editor in PLOS ONE. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

\* E-mail: egorshtkv@gmail.com

## Introduction

The Beijing genotype of *Mycobacterium tuberculosis* (MTB) has been shown to be globally spread all over the world [1]. In Russia, half of the local MTB population belongs to Beijing genotype, and one-fourth of these strains belong to the so-called B0/W148 clonal group [2]. Members of this group possess a specific 17-band IS6110 restriction fragment length polymorphism (RFLP) pattern, which was originally identified in Russia in the 1990s [3,4]. In comparison with other Beijing genotypes, B0/W148 strains demonstrated an increased virulence in the macrophage model [5], a stronger association with multidrug resistance [6], and an increased transmissibility [7,8]. The Beijing B0/W148 has been defined as a 'successful Russian clone' of *M. tuberculosis*, and its pathobiology and phylogeography have recently been reviewed

and discussed [2]. Together, these findings have led to assumption that Beijing B0/W148 strains possess unique genomic features that gave them evolutionary advantages.

To date, a small amount of whole genome sequencing data for B0/W148 MTB strains has been uploaded into the international databases, including one genomic scaffold of the W-148 strain (GL877853.1) and raw sequencing data in the NCBI Sequence Read Archive for several strains from the Samara region in Russia [9]. The aim of this work was to get more profound knowledge regarding the properties of Beijing B0/W148 strains using comparative genomics approach. All newly sequenced genomes were shown to be similar to the W-148 strain. Whole genome alignment between W-148 and the reference H37Rv MTB strain revealed two large chromosomal inversions in the W-148 genome. The largest inversion changed the orientation of the three

megabase (Mb) segment of the chromosome. The second one occurred in the previously inverted region and partly restored the orientation of the large inner segment. These two inversions were flanked by partial or complete copies of mobile genetic element *IS6110* and touched large parts of genome. Detailed PCR analysis of our sequenced strains ( $n = 4$ ) revealed the rearrangements in their genomes identical to those ones found in W-148 strain.

Remarkably, only two cases of large-scale genome rearrangement events in the MTB have been reported until now. First case was reported by Domenech P. *et al.* [10] describing the duplication of a 350 kilobase (Kb) region spanning Rv3128c to Rv3427c in the strains belonging to the W/Beijing family of MTB lineage 2. The second case was described by Weiner B. *et al.* [11], showing the independent duplication events occurred in the MTB lineages 2 and 4 were found. We have found another example of chromosomal rearrangement, i.e. inversions of large DNA segments. Large inversions were previously detected in some bacteria [12,13], but not in MTB.

Here we report two large-scale genome inversions characteristic exclusively for the members of MTB Beijing B0/W148 cluster and further hypothesize that these events occurred in their progenitor. This is the first report of a large-scale inversion in the MTB genome, and we hope that it will be one more step in filling the gap in the knowledge of a history and of an evolution of this pathogen.

## Results

### Genome sequencing of four clinical *M. tuberculosis* isolates belonged to Beijing B0/W148 cluster

Four Russian MTB isolates SP1, SP7, SP21, and MOS11 of the Beijing B0/W148 cluster were selected for whole genome re-sequencing (Table 1). Genomes were sequenced up to 98% completion using 454 pyrosequencing with more than 10-fold of coverage. To determine the taxonomy relationship between our strains and previously sequenced Beijing MTB strains deposited in GenBank, we have performed a phylogenetic analysis using polymorphisms relative to the reference genome of H37Rv MTB strain. The CTRI-4 strain previously sequenced in our laboratory and representing ancestral Beijing sublineage [17] was additionally included into analysis. Phylogenetic tree was built based on overall SNPs extracted from genomic DNA sequences after excluding SNPs for PE-PPE and PGRS protein families. This approach does not give us the perfect phylogenetic relationships in a case of fast evolving microorganisms influenced by recombination; however, it can be very efficient for the genetically monomorphic bacteria such as MTB [18]. The resulting phylogeny is shown in Figure 1. Phylogenetic tree demonstrated a close similarity between the genomes of four Beijing B0/W148 strains sequenced in this study and W-148 strain.

### Rearrangements in W-148 chromosomal DNA

Similarity of the genome sequences of the studied strains and W-148 gave us an opportunity to analyze structural genomic rearrangements within this group. Start of the W-148 genome was changed in relation to base 1 of the MTB H37Rv genomic sequence. This whole genome alignment of W-148 and H37Rv chromosomal DNA sequences revealed the presence of two large inversions in the W-148 genome. The Mauve 2.3.1 program highlighted these chromosomal rearrangements by subdividing the W-148 genome into five local collinear blocks (LCBs) (Table 2). This analysis demonstrated that the first, third and fifth LCBs were conserved whereas the second and fourth were inverted and rearranged in W-148 with respect to H37Rv (Figure 2).

### Chromosomal rearrangements in SP21 MTB strain confirmed by NGS

Based on similarity of genomic DNA sequences between SP1, SP7, SP21, MOS11 and W-148 strains, we expected to find the discovered inversions in the other strains as well. To confirm this, we additionally sequenced the SP21 strain using the mate-pair library strategy. The assembly of the SP21 genome sequence was performed by combining 454 (70 K reads, mean length 540 bp) and Ion Torrent data (650 K reads, 180 bp, mate-pair) that together represented more than 40-fold coverage of the genome. Initial assembly was performed by using the GS de novo Assembler and produced 391 contigs which length ranged from 500 to 69,788 bp. Further scaffolding resulted in 12 scaffolds with a total length of 4.45 Mb (AOUF00000000.1). The alignment of H37Rv, W-148, and SP21 chromosomal DNA sequences revealed the presence of identical large-scale inversions in both SP21 and W-148 strains (Figure S1).

### Chromosomal rearrangements in SP21 MTB strain confirmed by RFLP

The inversions observed in SP21 genome relative to H37Rv were verified by RFLP analysis. Based on analysis of the H37Rv restriction endonuclease map, the *MluI* was chosen for DNA digestion because its recognition sites were close to recombination junctions. The DNA probes specific to genome regions flanking the recombination junctions were obtained by PCR with specific primers (Table 1 in supplementary Text S1).

The RFLP analysis was performed for both SP21 and H37Rv *M. tuberculosis* strains and revealed the rearrangements in the SP21 genome sequence relative to that of H37Rv (Figure 3). In case of H37Rv, the hybridization signals from A&B, C&D, E&F, and G&H probes perfectly matched each other (Figure 3A) and the size of RFLP fragments corresponded to the expected one (Figure 1 in supplementary Text S1). In case of SP21, the RFLP pattern was different (Figure 3B). The signals from alternative combinations of probes (A&G, F&D, E&C and B&H) matched each other, which indicated the presence of this specific inversion (Figure 3C). The RFLP fragments corresponded to those expected in the inverted genome.

### The presence of inversions in other members of Beijing B0/W148 and non-Beijing B0/W148 groups

To verify the presence of the discovered inversions in another Beijing B0/W148 MTB strains, we developed a set of PCR primers flanking the sites of inversions. All primers were designed on the basis of W-148 genome (Table 3). Two pairs (P1&P2, P3&P4) flanked the ends of the external inverted region (between LCB I&IV and LCB II&V, respectively); and other pairs (P5&P6, P7&P8) flanked the ends of the internal region (between LCB IV&III and LCB III&II, respectively) (Figure 2). The primers were designed in such a way that the same primers used in different combinations would be suitable for analysis of genomic arrangement in other, non-B0/W148 strains. The size of expected PCR products is shown in Table 3.

We applied the developed amplification systems followed by sequencing of PCR products to the analysis of 40 MTB strains from the phylogenetic lineages 2 (East-Asian/Beijing) and 4 (Euro-American). Among them, 20 strains belonged to the Beijing family, including 8 Beijing B0/W148-cluster isolates, four strains belonged to Ural, and eight strains belonged to LAM. Primer pairs 1, 2, 5, and 6 (Table 3) amplified PCR fragments exclusively in the Beijing B0/W148 strains yielding 1021, 2215, 1761, and 2527 bp amplicons, respectively (see Figure 4A, lanes 1, 2, 5, 6, as

**Table 1.** Genotyping and drug resistance data of the B0/W148 strains sequenced in this study.

Strain	Drug resistance <sup>1</sup>	IS6110 RFLP <sup>2</sup>	Spoligotype clade <sup>3</sup> (SIT)	24-VNTR <sup>4</sup>	Location, year
SP1	RIF <sup>R</sup> , INH <sup>R</sup> , EMB <sup>R</sup> , STR <sup>R</sup> , PZA <sup>S</sup> , AMI <sup>S</sup> , CAP <sup>S</sup> , ETH <sup>R</sup> , OFL <sup>R</sup>	B0/W148	Beijing (1)	223325173533424672454433	Saint Petersburg (North West Russia), 2011
SP7	RIF <sup>R</sup> , INH <sup>R</sup> , EMB <sup>R</sup> , STR <sup>R</sup> , PZA <sup>R</sup> , ETH <sup>R</sup> , AMI <sup>R</sup> , CAP <sup>R</sup> , OFL <sup>R</sup>	B0/W148	Beijing (1)	223325173533424672454433	Saint Petersburg (North West Russia), 2011
SP21	RIF <sup>R</sup> , INH <sup>R</sup> , EMB <sup>R</sup> , STR <sup>R</sup> , PZA <sup>S</sup> , AMI <sup>R</sup> , CAP <sup>R</sup> , ETH <sup>S</sup> , OFL <sup>R</sup>	B0/W148	Beijing (1)	223325173533424672454433	Saint Petersburg (North West Russia), 2010
MOS11	RIF <sup>R</sup> , INH <sup>R</sup> , EMB <sup>R</sup> , STR <sup>R</sup> , PZA <sup>S</sup> , AMI <sup>S</sup> , CAP <sup>S</sup> , ETH <sup>R</sup> , OFL <sup>R</sup>	B0/W148	Beijing (1)	223325173533424672454433	Moscow (Central Russia), 2011

<sup>1</sup>RIF - rifampicin, INH - isoniazid, EMB - ethambutol, STR - streptomycin, PZA - pyrazinamide, ETH - ethionamide, AMI - amikacin, CAPR - capreomycin, OFL - ofloxacin.

<sup>2</sup>B0 designation according to Narvskaia *et al.* [6], W148 according to Bifani *et al.* [14].

<sup>3</sup>SITVITWEB was used for identification of data [15].

<sup>4</sup>24 - VNTR: s154, s580, s960, s1644, s2059, s2531, s2687, s2996, s3007, s3192, s4348, s802, s2165, s2461, s577, s2163, s4052, s4156, s424, s1955, s2347, s2401, s3171, s3690 [16].

doi:10.1371/journal.pone.0084971.t001

an example). No PCR products were obtained for 32 non- B0/W148 strains (Figure 4B, lanes 1, 2, 5, 6).

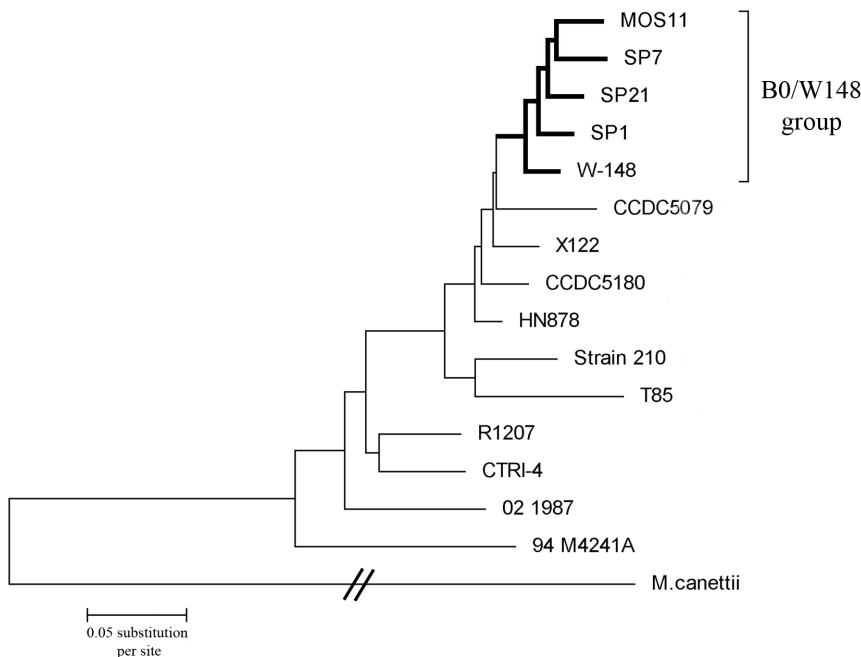
On the contrary, the primer pairs 3, 4, 7, and 8 using the same primers in different combinations (Table 3) amplified expected PCR products in non-B0/W148 strains (Figure 4B, lanes 3, 4, 7, and 8), whereas no PCR products were obtained for Beijing B0/W148 strains (Figure 4A, lanes 3, 4, 7, and 8). The differences in length of amplicons produced by primer sets 4, 7, and 8 for groups of non-B0/W148 Beijing and non-Beijing (Ural and LAM) strains (Table 3) is related to the presence of a complete copy of IS6110 in the analyzed region in non-B0/W148 Beijing strains in contrast to LAM and Ural strains. The specificity of produced PCR products was confirmed by Sanger sequencing in all cases.

These results were additionally verified by using the alternative primer sets selected in the similar way. Primer sequences, expected amplicons' lengths, and electrophoregram of PCR products obtained for Beijing B0/W148 and non-Beijing B0/W148 strains are presented in the Text S2.

Thus, we demonstrated the presence of identical inversions in chromosomal DNA of the studied Beijing B0/W148 strains ( $n = 8$ ), which appears to be a unique event specific to this clonal cluster.

### The hypothetical reconstruction of recombination events in Beijing B0/W148 progenitor

Further we tried to reconstruct the order of rearrangements occurred in a hypothetical W-148 progenitor genome. We suggested that the order and orientation of LCBs in the genome



**Figure 1. Comparative phylogenetic analysis of strains under study and 12 whole genomes from the NCBI database.** Phylogenetic tree based on all SNPs of genomes was constructed using the Neighbor-Joining algorithm. Evolutionary distances were calculated using p-distance method.

doi:10.1371/journal.pone.0084971.g001

**Table 2.** Description of five local collinear blocks (LCBs) between H37Rv and W-148.

LCB	Type	H37Rv position	W-148 position
I	Conserved	1-704243	1-769418
II	Rearranged	704246-1262963	3247025-3814187
III	Conserved*	1262961-3379027	1108934-3245669
IV	Rearranged	3379025-3710381	773801-1107578
V	Conserved	3711737-4411532	3814961-4539434

\*- LCB III conserved according to initial mechanistic analysis, but re-arranged according to our hypothesis

doi:10.1371/journal.pone.0084971.t002

of the W-148 progenitor 1 (P1) was the same as in H37Rv and in other Beijing strains and designed it *in silico* (Figure 5). During the evolution, the first 3 Mb inversion occurred symmetrically across the replication axis and affected LCBs II, III and IV with the formation of progenitor 2 (P2). This recombination event rearranged chromosomal DNA between Rv0609a and Rv3327 genes relative to H37Rv (Figure 2). However, in other Beijing strains, the region between Rv3326 and Rv3327 genes was already disrupted by integration of IS6110. Interestingly, we found only parts of IS6110 in recombination junctions of the inverted region in the genome of W-148. The 812-bp and 543-bp fragments of IS6110 were detected at the boundaries of the LCBs I&II and LCBs IV&V, respectively. These two parts were inverted and formed together a perfect whole sequence of IS6110. We suppose that P1 had two inverted copies of IS6110, which were integrated into sites equidistant from the terminus of a replication (*ter*) region.

According to our hypothesis next step of recombination occurred in Progenitor 2 genome and affected LCB III between Rv3020c and disrupted Rv1135c genes. This inversion restored an original orientation of this segment to the initial form like in P1 and H37Rv and led to the formation of W-148. The inversion of

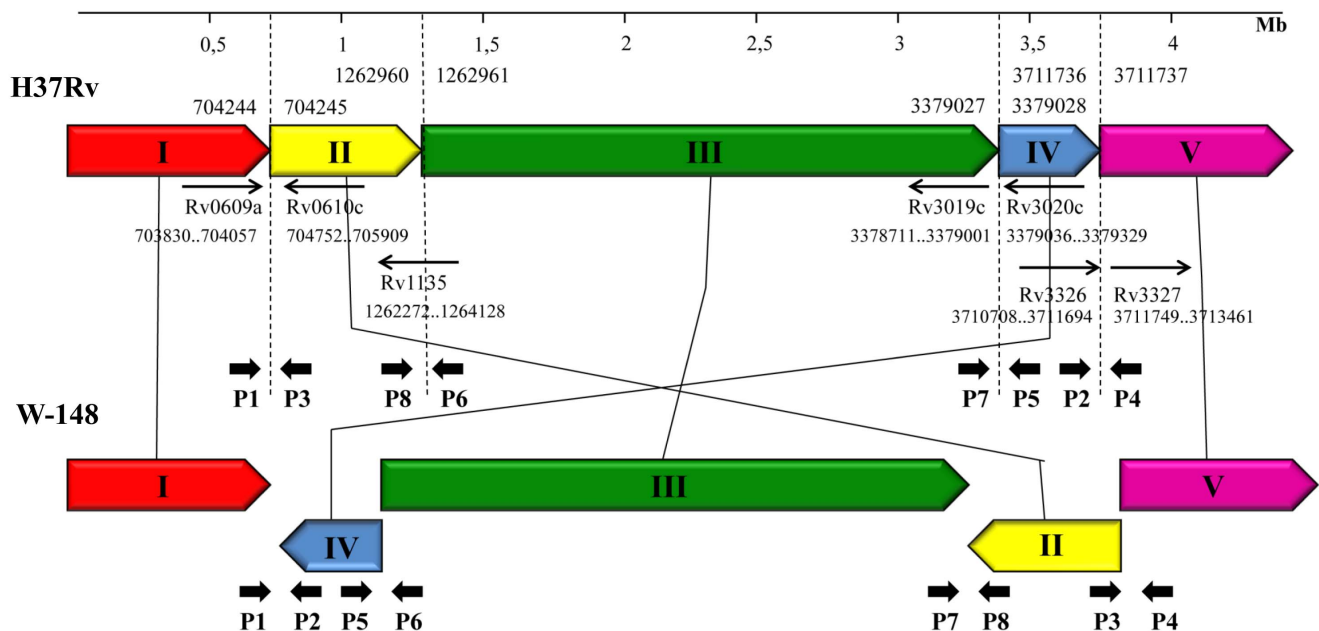
this LCB was most probably mediated by two inverted complete copies of IS6110, which were found on the borders of this LCB. Remarkably, all Beijing strains in the NCBI database have a complete copy of IS6110 between LCBs II and III (between Rv3019 and Rv3020c genes), between LCBs III and IV (disrupted the Rv1135c gene), and between LCBs IV and V (between Rv3326 and Rv3327 genes), while they do not have it between LCBs I and II (between Rv0609 and Rv0610c genes).

## Discussion

This study focused on the genomic characterization of the MTB strains of the Beijing B0/W148 cluster, endemic for Russia and representing the epidemiologically successful variant of MTB [2,5,6]. Recently, Mokrousov [2] hypothesized that “B0/W148 likely originated in Siberia, and its primary dispersal was driven by a massive population outflow from Siberia to European Russia in the 1960–80s” and “a historically recent, phylogenetically demonstrated successful dissemination of the Beijing B0/W148 strain was triggered by an advent and wide use of the modern anti-TB drugs and was due to its remarkable capacity to acquire drug resistance”. For this reason, we sequenced genomes of four Beijing B0/W148 MTB clinical strains isolated in Russia in 2010–2011.

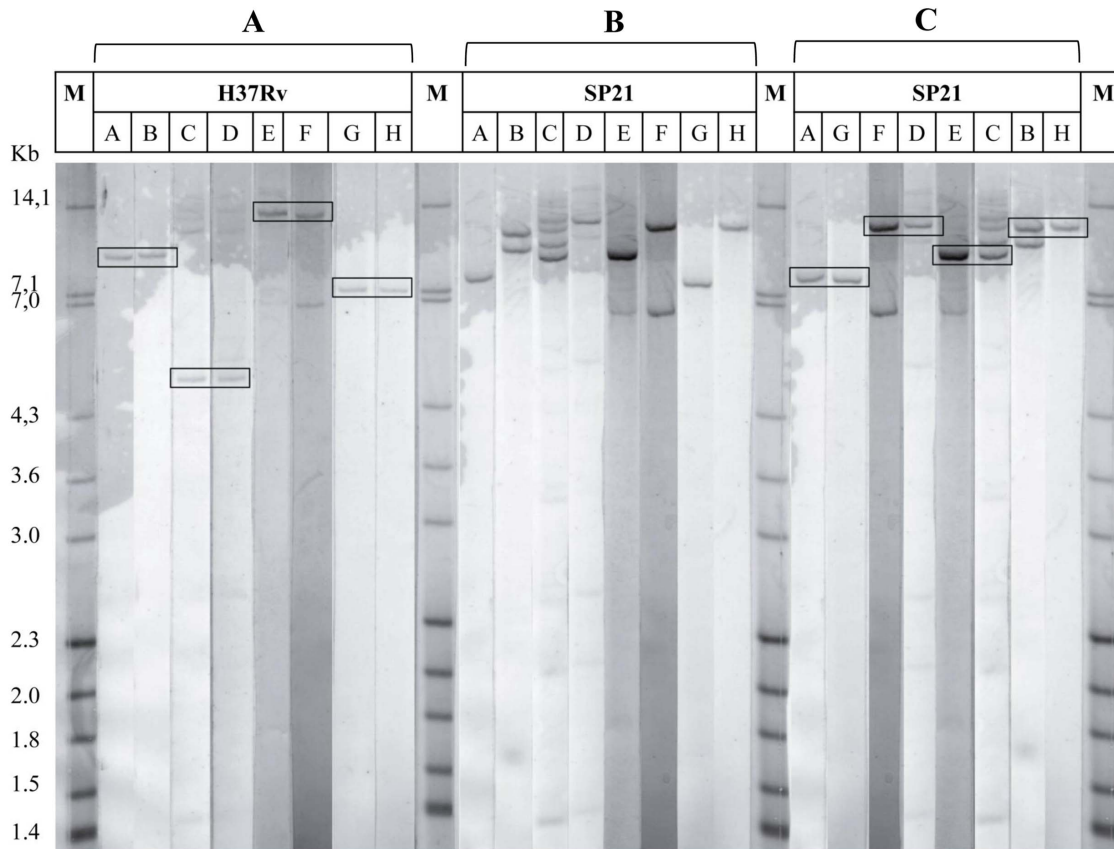
We used the 454 pyrosequencing technology, which produces the long reads (up to 800 bp). This gave us a good opportunity to see indels, and to identify most of the LSPs (large sequence polymorphisms) present in the studied genomes. Additionally, the genome sequence of W-148 strain represented in GenBank was included in analysis.

Comparing genomes of H37Rv and W-148 strains, we detected two large-scale inversions in the genome of W-148, which were confirmed to be present in all studied strains of Beijing B0/W148-cluster. Notably, the presence of large-scale chromosomal rearrangements within mycobacteria genus was recently shown by *in silico* analysis [20]. The genome of *Mycobacterium smegmatis* mc(2) 155 contains a 56 Kb duplicated region when compared with ATCC 607 progenitor. This duplication is flanked by two copies of



**Figure 2. Schematic representation of genome rearrangements and primer design strategy.** Each local collinear block (LCB) corresponds to Table 2 and is represented by a different color.

doi:10.1371/journal.pone.0084971.g002



**Figure 3. Southern blot analysis of H37Rv and SP21 MTB strains.** Genomic DNA was digested with *MluI* and hybridized with the fluorescent labeled probes obtained by amplification. The probes are listed at the top of the lanes (from A to H). (A) Hybridization patterns of H37Rv strain. The order of probes corresponds to the order of complementary sequence sites in the genome of H37Rv. (B) Hybridization pattern of SP21 strains. The order of probes corresponds to the order of complementary sequence sites in the genome of H37Rv. (C) Hybridization patterns of SP21 strain. The order of probes is rearranged in accordance with the expected order of complementary sequence sites in the inverted genome (Supplementary Text S1). The merged bands from probes complementary to the boundaries of recombination junctions are boxed. M, marker strain Mt14323 (Mycobacterial Reference Laboratory, National Public Health Institute (Turku, Finland)).  
doi:10.1371/journal.pone.0084971.g003

an *IS1096* element [21]. Comparative genomics revealed two large tandem chromosomal duplications, DU1 and DU2, in *Mycobacterium bovis* BCG strain. DU1 was found only in BCG Pasteur, while four different forms of DU2 were found in different BCG strains [22]. Two cases of large duplications occurred in the MTB belonged to lineages 2 and 4 have been reported to date [10,11]. Some of the duplicated regions were flanked by *IS6110* elements supporting a general assertion that the majority of genomic rearrangements are mediated by the mobile genetic elements or repeats [23].

As far as large-scale chromosomal inversions are concerned, a single event was detected among *M. tuberculosis* KZN strains, and there were several such events in *Mycobacterium avium* evolution. Three KZN strains sequenced by Broad Institute showed a large-scale inversion of nearly 2.5 Mb (spanning coordinates ~1 Mb to ~3.5 Mb, relative to the origin of replication), although the re-sequencing of one of these strains in another laboratory found no evidence for this event [24]. In *M. avium*, large-scale inversions were found between *M. avium* subspecies *hominissuis* and subspecies *paratuberculosis* [25]. The interspecies comparison of genomes of fish *M. marinum* isolates and *M. tuberculosis* also revealed X-shaped chromosomal inversions derived from the accumulation of rearrangements that were symmetrical across the replication axis [26].

In our study, we discovered the large-scale chromosomal rearrangements characteristic for MTB isolates of the Beijing B0/W148-cluster. The presence of these inversions in all members of Beijing B0/W148 group was confirmed by PCR, sequencing and RFLP hybridization analysis. Additionally, we suggest a two-step scenario of evolution for these strains. In the first step, a large-scale inversion of the 3 Mb segment of the chromosome occurred. This assumption is based on the fact that boundaries of inversion are perfectly equidistant from the site of terminus of replication (i.e., symmetrical across the replication axis). There is a lot of data supporting the chromosome rearrangement around the *ter* region in other bacterial genomes [27], and MTB may have probably implemented a similar mechanism. However, the reason why we have found only half of *IS6110* at the boundaries of inversion is not clear. Remarkably, one part of this disrupted *IS6110* contains a site for *PvuII* while its second part contains the sequence used as a probe in *IS6110*-RFLP typing (between LCBs I and II), which is why only one band is detected in the *IS6110*-RFLP profile. This ~7.4 Kb band corresponds to two sites for *PvuII* found in unique regions of the W-148 genome (Figure S2). Using the BioNumerics version 5.1 package we compared a collection of *IS6110* RFLP profiles of different Beijing and non-Beijing genotypes. As a result, only members of the Beijing B0/W148-cluster contained the ~7.4 Kb band demonstrating their unique origine.



**Table 3.** Primers designed for analysis of genomic arrangement in studied MBT strains.

Primers' set No	Primer's name	5'-3' sequences	Product's length		
			B0/W148 Beijing	other Beijing	non-Beijing (Ural, LAM)
External inversion (links between LCB I, II, IV, V)					
1	P1	gtgtgtacattgggcatcg	1021	no reaction	no reaction
	P2	ggtgtacatatcgaagctcg			
2	P3	gctcgacgaagtgagattgc	2215	no reaction	no reaction
	P4	ttggcgatccgatacagtgc			
3	P1	gtgtgtacattgggcatcg	no reaction	320	320
	P3	gctcgacgaagtgagattgc			
4	P2	ggtgtacatatcgaagctcg	no reaction	2920	1845
	P4	ttggcgatccgatacagtgc			
Internal inversion (links between LCB II, III, IV)					
5	P5	ctgccaagcactggacagc	1761	no reaction	no reaction
	P6	caagtctccggtattcaagg			
6	P7	agccttggtcgtctcttacc	2527	no reaction	no reaction
	P8	cacggctctcccaacgtgg			
7	P6	caagtctccggtattcaagg	no reaction	1841	483
	P8	cacggctctcccaacgtgg			
8	P5	ctgccaagcactggacagc	no reaction	2447	1090
	P7	agccttggtcgtctcttacc			

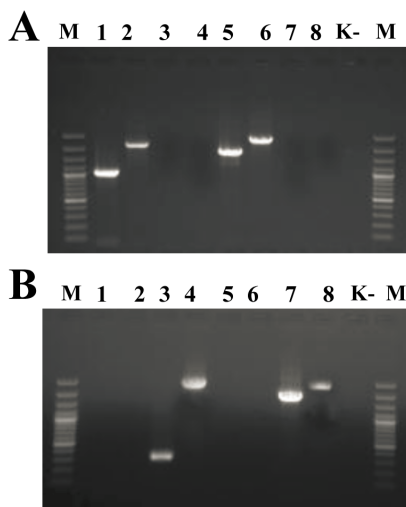
doi:10.1371/journal.pone.0084971.t003

The second inversion occurred with LCB III and partly restored an orientation of the large inner segment. As it was noted above, the *IS6110* flanking LCB III is found in all Beijing strains available in GenBank. One of the characteristics of *IS6110* integration is a duplication of the 3–4 base pair region flanking the inserted element at the insertion site [28]. We checked the presence of

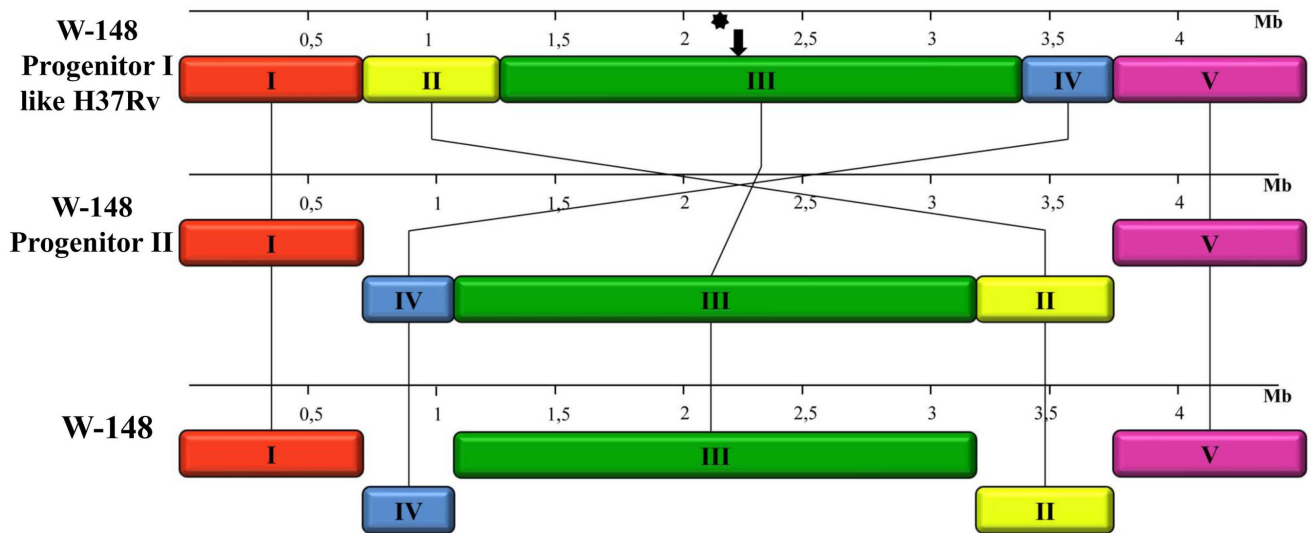
these duplications in the genomes of B0/W148 and non-B0/W148 Beijing strains. In non-B0/W148 strains, the duplication of AGC proximal to the *IS6110* insertion site between LCBs II and III was found, while the CAG was duplicated between LCBs III and IV (Figure 6). In B0/W148 strains, the sequences of duplicated triplets in the LCB III were in the same orientation, while the sequences of triplets in the LCBs II and IV were inverted and rearranged, which corresponded to the origin of W-148 from W-148 Progenitor 2 (Figure 6). In this case, a homologous recombination between *IS6110* elements appears to be the most appropriate mechanism for the inversion.

Another possible evolutionary scenario suggests that LCBs II and IV have recombined independently of LCB III. According to this hypothesis, LCBs II and IV could recombine simultaneously or stepwise. However, it seems unlikely that blocks II and IV were involved in two independent recombination events simultaneously. Thus, it should have been a sequential recombination process. At first, block II or IV recombined and then the remaining one. Since these blocks are very distantly located from each other, these recombination events most likely were independent. The case where LCBs II and IV have recombined independently of LCB III is also possible but looks improbable.

It has not escaped our notice that the described large rearrangements could have potential consequences for the phenotype as described for other bacteria [29]. Therefore we looked more closely at genes involved into the postulated recombination events. As described in Results section above, the discovered inversions occurred in the proximity of the Rv0609, Rv0610c, Rv1135c, Rv3019, Rv3020c, Rv3326, and Rv3327 genes (Figure 2). However, only in the B0/W148 strains the disrupted part of *IS6110* element is found between Rv0609 and Rv0610c in comparison with other Beijing strains. Both of these genes code for hypothetical proteins, they are located far away



**Figure 4. Results of PCR verification of inversions.** Electropherogram of PCR products obtained for MTB strains during the amplification with primer sets 1–8 (Table 3). (A) SP 21 B0/W148 Beijing strain and (B) SP 5 non-B0/W148 Beijing strain. Lanes 1–8 correspond to primer sets 1–8; M is a marker GeneRuler 100 bp Plus DNA Ladder (Fermentas, SM0324); K- is a negative control. doi:10.1371/journal.pone.0084971.g004



**Figure 5. Genome rearrangements' representation for W-148 Progenitor I like H37Rv, W-148 Progenitor II and W-148 genomes.** Each local collinear block (LCB) I–V is represented by a different color. Upside-down blocks (LCBs II and IV) represent the location of the reverse strand, which means an inversion has occurred. Asterisk indicates a terminus of a replication site. Terminus of a replication site was calculated based on GraphDNA (GC-skew mode) software [19]. doi:10.1371/journal.pone.0084971.g005

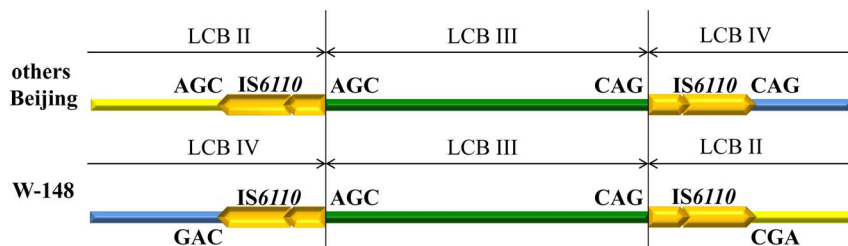
from the site of recombination, and it is hard to assume any influence on the phenotype.

To understand the causes of recombination events in Beijing B0/W148 strains, the complete list of unique cluster-specific SNPs (n = 94) was built (Table S1). We have included only those mutations which were found in at least four of the five isolates under consideration. All of these SNPs were mapped to genes coding the proteins of repair, recombination and replication (3R) system in the MTB [30,31]. No non-synonymous SNPs were found. One synonymous mutation Gly269Gly was found in the RecF protein, which could hardly be associated to large-scale inversions.

To classify the precise genetic sublineage of our sequenced strains, we examined five specific LSPs present in genomes of the East Asian lineage (RD105, RD207, RD181, RD150 and RD142) [32,33]. According to this analysis, the studied strains belong to the Beijing sublineage 3 (RD105, RD207 and RD181 were deleted), as well as the strains with the large duplications recently reported [10,11]. These studies reported the large duplication occurred in the strains within sublineages 3, 4 and 5, which spans 350 Kb of the chromosome from the Rv3128c to the Rv3427c genes. Additionally, this duplication was flanked by two complete copies of IS6110 in the same orientation. After a detailed review of

strains studied, we found no evidence of IS6110 duplication, and the different location of inversions boundaries. Remarkably, according to Weiner *et al.* [11] the strain T67 had downstream boundary at Rv3326, and it corresponded to the boundary of the LCB V in the W-148 genome. Interestingly, this region additionally corresponds to the RvD5 in the H37Rv genome and Rv3326, which is a part of IS6110 flanking it from one side [34,35]. In almost all Beijing strains, the orientation of IS6110 in this region is different from the H37Rv.

In summary, we described a chromosomal rearrangement, inversions of large DNA segments in strains of the MTB Beijing B0/W148-cluster. The members of this group possess particular pathobiological features mentioned above, and further studies are necessary to determine the impact of the found inversions on the biological properties of the pathogen. These and previously described inversions and previously reported duplications in the region from 3 Mb to 4 Mb are intriguing and cause an increased interest in these genomes. These rearrangements may possibly reflect evolution of the global chromosomal DNA topology or local DNA-DNA interactions within this region. We hope that our study and studies of other bacteria concerning large-scale rearrangements will shed light on the understanding of the genome evolution of MTB.



**Figure 6. Schematic presentation of duplicated bases flanking the IS6110 insertion sites between LCBs II&III and III&IV.** doi:10.1371/journal.pone.0084971.g006

## Materials and Methods

### Mycobacterial isolates

A total of 40 MTB strains were obtained from the culture collection of the Research Institute of Phthisiopulmonology (St. Petersburg, Russia) and Moscow Scientific-Practical Center of Treatment of Tuberculosis of Moscow Healthcare (Moscow, Russia). The susceptibility testing was done using a BACTEC™ MGIT™ 960 Culture system (Becton Dickinson, USA) by standard protocol. Standard MTB genotyping methods, including spoligotyping and 24 loci MIRU-VNTR were applied to these strains as previously described [16,36] (Table S2). Of them 28, 8, and 4 isolates had Beijing, LAM, and Ural spoligotype profiles, respectively. For Beijing isolates with spoligotype SIT1 IS6110 RFLP analysis was additionally performed [37]. BioNumerics version 5.1 package (Applied Maths, Belgium) was used for band comparison. According to RFLP analysis eight isolates belonged to Beijing B0/W148 cluster. Four of them were selected for a current whole genome re-sequencing project (Table 1).

### Whole genome sequencing and assembly

DNA extraction was performed as previously described [37]. Four B0/W148 strains, SP1, SP7, SP21, and MOS11 were sequenced by using the Roche 454 Life Sciences Genome Sequencer FLX following the manufacturer's instructions (Roche 454 Life Science, USA). Assembly of raw sequencing reads with an average length of 540 bases was performed by the GS de novo assembly software version 2.8 (Roche 454 Life Science, USA). Raw sequencing data for MTB genomes SP1, SP7, SP21, and MOS11 were deposited in the NCBI Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/Traces/sra/>) under accession numbers SRX216883, SRX216889, SRX216899, and SRX216918.

The SP21 MTB strain was sequenced using the Ion Torrent PGM (Life Technologies, USA). Two Ion Torrent PGM mate-paired libraries with average size 2 to 3 Kb and 3 to 4 Kb were constructed using Ion mate-paired library preparation guide (Life Technologies, USA). For genome scaffolding, a hybrid assembly strategy was chosen. Initially, GS de novo assembler was used for the 454 and Ion Torrent data assembly, and further Ion Torrent reads were used for genome scaffolding by SSpace [38]

### Genomes analysis

All individual reads generated using the 454 platform were mapped to H37Rv [GenBank: AL123456.2] genome using the 454 GS Reference Mapper (Roche 454 Life Science, USA). Consensus sequences were called, and point mutations identified for sites covered by at least 3 reads, with PHRED scores greater than 30. SNPs calling for whole genomes and contigs from GenBank database representative of the Beijing genotype, 02\_1987 (ABLM000000000), 94\_M4241A (ABLL000000000), CCDC5079 (CP001641), CCDC5180 (CP001642), CTRI-4 (AII000000000.1), HN878 (ADNF000000000.1), R1207 (ADNH000000000.1), Strain\_210 (ADAB000000000), T85 (ABOW000000000), W-148 (GL877853.1), and X122 (ADNG000000000.1) was done using MUMmer 3.20 with its nucmer and show-snps functions [39].

Phylogenetic tree was built based on overall SNPs extracted from genomic DNA sequences after excluding SNPs for PE-PPE and PGRS protein families using MEGA4. *M. canettii* was taken as an out-group. A Neighbor-Joining algorithm was used to build a tree. Phylogenetic distance was calculated by using p-distance.

Genomic sequences of four B0/W148 strains under the study and W-148 [GenBank: GL877853.1] were compared to each

other and to the H37Rv MTB strain. Genomes were aligned with the open-source MAUVE aligner, version 2.3.1, using the progressive algorithm [40] (<http://gel.ahabs.wisc.edu/mauve/>).

### PCR verification of inversions

The standard PCR was carried out in 25 µL of reaction mixture. The reaction mixture contained 66 mM Tris-HCl (pH 9.0), 16.6 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 2.5 mM MgCl<sub>2</sub>, 250 µM of each dNTP, 1 U of Taq DNA polymerase (Promega, USA), 2.5 mM betain (SIGMA, USA) and 10 pmol of each primer (Table 3, Text S1). One to ten nanograms of genomic DNA were used as a template for PCR. A universal amplification profile included the following steps: the initial heating step was at 94°C for 2 min, followed by 30 cycles of 94°C for 30 sec, 61°C for 15 sec and 72°C for 20 sec and a final step at 72°C for 5 min. The PCR products were then sequenced by conventional Sanger capillary methods on ABI Prism 3730 Genetic Analyzer (Applied Biosystems, USA; Hitachi, Japan) and compared to the H37Rv and W-148 genomes.

### RFLP analysis

RFLP analysis was performed as recommended by van Embden et al. [37] with modifications. Briefly, the whole genomic DNA of SP21 and H37Rv *M. tuberculosis* strains were treated with 15 units of *Mba*I (Thermo Scientific, USA) in recommend reaction buffer during the night at 37°C. Probes for the Southern analysis were obtained by conventional PCR using Amersham ECL labeling and detection systems (GE Healthcare) with dedicated primers sets (Supplementary Text S2). The obtained profiles on ECL films were scanned and processed with BioNumerics version 5.1 package (Applied Maths, Belgium).

### Supporting Information

**Figure S1 Alignment of genomes of H37Rv, W-148, and SP21 MTB strains represented by Mauve 2.3.1.** Colored outlined blocks surround regions of the genome sequence that aligned to part of another genome (LCBs numbering is the same as in the **Figure 2** of the manuscript). Lines link blocks with homology between genomes. Genomes from top to bottom: H37Rv, W-148, and SP21. Vertical red lines in the SP21 correspond to the boundaries of the scaffolds. Scaffolds 5, 3, and 9 containing sequences of inverted regions are indicated by double-headed arrows. The sequences flanked the sites of inverted regions were found within scaffolds 5, 3, and 9. Scaffold 5 (392,333 bp) includes full sequence of the LCB IV (for LCB numbering and length see Table 2 and Figure 2 in the main text) and parts of the LCB I and III (29 Kb and 16 Kb, respectively). Scaffold 3 (738,393 bp) includes large parts of the LCB III and II (16 Kb and 132 Kb). Scaffold 9 (162,927 bp) includes parts of the LCB II and V (132 Kb and 31 Kb, respectively). (TIF)

**Figure S2 Schematic view of the IS6110RFLP profiles of the B0/W148 strain generated with Bionumerics program.** M- molecular weights marker (*Pvu*II digested DNA of strain Mt14323). Arrow indicates particular fragment of IS6110 RFLP corresponding to the inverted region in the Beijing B0/W148 isolates. (TIF)

**Table S1 B0/W148 cluster specific SNPs.** (XLSX)

**Table S2 Description of *M. tuberculosis* clinical isolates involved in this study.** (XLSX)



**Text S1 RFLP-analysis for confirmation of inversions.**  
(DOCX)

**Text S2 Additional primers designed for confirmation of inversions.**  
(DOCX)

## Acknowledgments

The authors are indebted to G.B. Smirnov for his interest and help in manuscript preparation. We also thank V. Karpov for primer synthesis,

M. Chukin for support with sequencing instrumentation, A. Manulov for bioinformatic analysis.

## Author Contributions

Conceived and designed the experiments: EAS OVN ENI VMG. Performed the experiments: JAB IYK ESK YDI EYN AAV VYZ PKY BIV TFO IVM. Analyzed the data: EAS JAB DSI IVM OVN DGA. Contributed reagents/materials/analysis tools: DSI EYN OVN BIV. Wrote the paper: EAS JAB OVN IVM ENI VMG IVM DGA.

## References

- Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, et al. (2006) *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol* 6: 23.
- Mokrousov I (2013) Insights into the origin, emergence, and current spread of a successful Russian clone of *Mycobacterium tuberculosis*. *Clin Microbiol Rev* 26:342–360.
- Narvskaia OV, Mokrousov IV, Otten TF, Vishnevskii BI (1999) [Genetic marking of polyresistant *Mycobacterium tuberculosis* strains isolated in the north-west of Russia]. *Probl Tuberk*: 39–41.
- Kurepina NE, Ramaswamy S, Shashkina EF, Sloutsky AM, Blinova LN, et al. (2001) The sequence analysis of the *pncA* gene determining the PZA-resistance in the predominant *M. tuberculosis* strains isolated in the Tomsk penitentiary system, western Siberia, Russia. *Int J Tuberc Lung Dis* 5:S41.
- Lasunskaja E, Ribeiro SC, Manicheva O, Gomes LL, Suffys PN, et al. (2010) 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* 12: 467–475. doi: 10.1016/j.micinf.2010.1002.1008. Epub 2010 Mar 1017.
- Narvskaia O, Mokrousov I, Otten T, Vishnevskiy B (2005) Molecular markers: application for studies of *Mycobacterium tuberculosis* population in Russia. In: *Trends in DNA fingerprinting research* (ed MM Read) Nova Science Publishers NY, USA pp 111–125
- Pardini M, Niemann S, Varaine F, Iona E, Meacci F, et al. (2009) Characteristics of drug-resistant tuberculosis in Abkhazia (Georgia), a high-prevalence area in Eastern Europe. *Tuberculosis (Edinb)* 89: 317–324. doi: 10.1016/j.tube.2009.1004.1002. Epub 2009 Jun 1017.
- Narvskaia O, Otten T, Limeschenko E, Sapozhnikova N, Grashchenkova O, et al. (2002) Nosocomial outbreak of multidrug-resistant tuberculosis caused by a strain of *Mycobacterium tuberculosis* W-Beijing family in St. Petersburg, Russia. *Eur J Clin Microbiol Infect Dis* 21: 596–602. Epub 2002 Aug 2015.
- Casali N, Nikolayevskiy V, Balabanova Y, Ignatyeva O, Kontsevaya I, et al. (2012) Microevolution of extensively drug-resistant tuberculosis in Russia. *Genome Res* 22: 735–745. doi: 10.1101/gr.128678.128111. Epub 122012 Jan 128631.
- Domenech P, Kolly GS, Leon-Solis L, Fallow A, Reed MB (2010) Massive gene duplication event among clinical isolates of the *Mycobacterium tuberculosis* W/Beijing family. *J Bacteriol* 192: 4562–4570. doi: 10.1128/JB.00536-00510. Epub 02010 Jul 00516.
- Weiner B, Gomez J, Victor TC, Warren RM, Sloutsky A, et al. (2012) Independent large scale duplications in multiple *M. tuberculosis* lineages overlapping the same genomic region. *PLoS One* 7: e26038. doi: 10.1371/journal.pone.0026038. Epub 0022012 Feb 0026037.
- Coyne S, Courvalin P, Galimand M (2010) Acquisition of multidrug resistance transposon Tn6061 and IS6100-mediated large chromosomal inversions in *Pseudomonas aeruginosa* clinical isolates. *Microbiology* 156: 1448–1458. doi: 10.1099/mic.1440.033639-033630. Epub 032010 Jan 033628.
- Nalbantoglu U, Sayood K, Dempsey MP, Iwen PC, Francesconi SC, et al. (2010) Large direct repeats flank genomic rearrangements between a new clinical isolate of *Francisella tularensis* subsp. *tularensis* A1 and Schu S4. *PLoS One* 5: e9007. doi: 10.1371/journal.pone.0009007.
- Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN (2002) Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. *Trends Microbiol* 10: 45–52.
- Demay C, Liens B, Burguiere T, Hill V, Couvin D, et al. (2012) SITVITWEB—a publicly available international multimer database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect Genet Evol* 12: 755–766. doi: 10.1016/j.meegid.2012.1002.1004. Epub 2012 Feb 1017.
- Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rusch-Gerdes S, et al. (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 44: 4498–4510. Epub 2006 Sep 4427.
- Ilna EN, Shitikov EA, Ikryannikova LN, Alekseev DG, Kamashev DE, et al. (2013) Comparative genomic analysis of *Mycobacterium tuberculosis* drug resistant strains from Russia. *PLoS One*. 2013;8(2):e56577. doi: 10.1371/journal.pone.0056577.
- Achtman M (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* 62:53–70. doi: 10.1146/annurev.micro.1162.081307.162832.
- Thomas JM, Horspool D, Brown G, Tcherepanov V, Upton C (2007) GraphDNA: a Java program for graphical display of DNA composition analyses. *BMC Bioinformatics* 8: 21.
- Garcia-Betancur JC, Menendez MC, Del Portillo P, Garcia MJ (2012) Alignment of multiple complete genomes suggests that gene rearrangements may contribute towards the speciation of *Mycobacteria*. *Infect Genet Evol* 12: 819–826. doi: 10.1016/j.meegid.2011.1009.1024. Epub 2011 Oct 1018.
- Wang XM, Galamba A, Warner DF, Soetaert K, Merkel JS, et al. (2008) IS1096-mediated DNA rearrangements play a key role in genome evolution of *Mycobacterium smegmatis*. *Tuberculosis (Edinb)* 88: 399–409. doi: 10.1016/j.tube.2008.1002.1003. Epub 2008 Apr 1024.
- Brosch R, Gordon SV, Buchrieser C, Pym AS, Garnier T, et al. (2000) Comparative genomics uncovers large tandem chromosomal duplications in *Mycobacterium bovis* BCG Pasteur. *Yeast* 17: 111–123.
- Smirnov GB (2010) [Repeats in bacterial genomes: evolutionary considerations]. *Mol Gen Mikrobiol Virusol*: 11–20.
- Ioerger TR, Koo S, No EG, Chen X, Larsen MH, et al. (2009) Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. *PLoS One* 4: e7778. doi: 10.1371/journal.pone.0007778.
- Hsu CY, Wu CW, Talaat AM (2011) Genome-Wide Sequence Variation among *Mycobacterium avium* subspecies *paratuberculosis* Isolates: A Better Understanding of John's Disease Transmission Dynamics. *Front Microbiol* 2:236. doi: 10.3389/fmicb.2011.00236. Epub 02011 Dec 00239.
- Kurokawa S, Kabayama J, Hwang SD, Nho SW, Hikima J, et al. (2013) Comparative genome analysis of fish and human isolates of *Mycobacterium marinum*. *Mar Biotechnol* 15: 596–605
- Eisen JA, Heidelberg JF, White O, Salzberg SL (2000) Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol* 1: R6. Epub 2000 Dec 4.
- Dale JW (1995) Mobile genetic elements in mycobacteria. *Eur Respir J Suppl* 20: 633s–648s.
- Cui L, Neoh HM, Iwamoto A, Hiramatsu K (2012) Coordinated phenotype switching with large-scale chromosome flip-flop inversion observed in bacteria. *Proc Natl Acad Sci U S A* 109: 1647–1656. Epub 2012 May 29.
- Dos Vultus T, Mestre O, Rauzier J, Golec M, Rastogi N, et al. (2008) Evolution and diversity of clonal bacteria: the paradigm of *Mycobacterium tuberculosis*. *PLoS One* 3: e1538. doi: 10.1371/journal.pone.0001538.
- Muttucumar DG, Parish T (2004) The molecular biology of recombination in *Mycobacteria*: what do we know and how can we use it? *Curr Issues Mol Biol* 6: 145–157.
- Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, et al. (2006) Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* 103: 2869–2873. Epub 2006 Feb 2813.
- Tsolaki AG, Gagneux S, Pym AS, Gouget de la Salmoniere YO, Kreiswirth BN, et al. (2005) Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. *J Clin Microbiol* 43: 3185–3191.
- Brosch R, Philipp WJ, Stavropoulos E, Colston MJ, Cole ST, et al. (1999) Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra strain. *Infect Immun* 67: 5768–5774.
- Fang Z, Doig C, Kenna DT, Smitipat N, Palittapongpim P, et al. (1999) IS6110-mediated deletions of wild-type chromosomes of *Mycobacterium tuberculosis*. *J Bacteriol* 181: 1014–1020.
- Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, et al. (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 35: 907–914.
- van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, et al. (1993) Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 31: 406–409.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding pre-assembled contigs using SSPAGE. *Bioinformatics* 27: 578–579. doi: 10.1093/bioinformatics/btq1683. Epub 2010 Dec 1012.
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5: R12. Epub 2004 Jan 2030.
- Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14: 1394–1403.