

RESEARCH ARTICLE

Evolutionary Trends of the Transposase-Encoding Open Reading Frames A and B (*orfA* and *orfB*) of the Mycobacterial IS6110 Insertion Sequence

Sara Thabet, Amine Namouchi[‡], Helmi Mardassi*

Unit of Typing and Genetics of Mycobacteria, Laboratory of Molecular Microbiology, Vaccinology, and Biotechnology Development; Institut Pasteur de Tunis, Tunis, Tunisia

[‡] Current address: Department of Microbiology, Unit for Genome Dynamics, Oslo University Hospital, Nydalen, Oslo, Norway

* helmi.merdassi@pasteur.ms.tn



OPEN ACCESS

Citation: Thabet S, Namouchi A, Mardassi H (2015) Evolutionary Trends of the Transposase-Encoding Open Reading Frames A and B (*orfA* and *orfB*) of the Mycobacterial IS6110 Insertion Sequence. PLoS ONE 10(6): e0130161. doi:10.1371/journal.pone.0130161

Academic Editor: Riccardo Manganeli, University of Padova, Medical School, ITALY

Received: April 3, 2015

Accepted: May 17, 2015

Published: June 18, 2015

Copyright: © 2015 Thabet et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Sequences data reported in the present study were deposited in GenBank under accession numbers KP844666 to KP844685 and KP844686 to KP844721.

Funding: HM received funding from the Tunisian Ministry of Higher Education and Scientific Research. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Background

The IS6110 insertion sequence, a member of the IS3 family of insertion sequences, was found to be specific to the *Mycobacterium tuberculosis* complex (MTBC). Although IS6110 has been extensively characterized as a transposable genetic marker, the evolutionary history of its own transposase-encoding sequence has not, to the best of our knowledge, been investigated.

Methodology/Principal Findings

Here we explored the evolution of the IS6110 sequence by analysing the genetic variability and the selective forces acting on its transposase-encoding open reading frames (ORFs) A and B (*orfA* and *orfB*). For this purpose, we used a strain collection consisting of smooth tubercle bacilli (STB), an early branching lineage of the MTBC, and present-day *M. tuberculosis* strains representing the full breadth of genetic diversity in Tunisia. In each ORF, we found a major haplotype that dominated over a flat distribution of rare descendent haplotypes, consisting mainly of single- and double-nucleotide variant singletons. The predominant haplotypes consisted of both ancestral and present-day strains, suggesting that IS6110 acquisition predated the emergence of the MTBC. There was no evidence of recombination and both ORFs were subjected to strict purifying selection, as demonstrated by their dN/dS ratios (0.29 and 0.51, respectively), as well as their significantly negative Tajima's *D* statistics. Strikingly, the purifying selection acting on *orfA* proved much more stringent, suggesting its critical role in regulating the transpositional process. Maximum likelihood analyses further excluded any possibility of positive selection acting on single amino acid residues.

Conclusions/Significance

Taken together our data fit with an evolutionary scenario according to which the observed variability pattern of the IS6110 transposase-encoding ORFs is generated mainly through random point mutations that accrued on a functionally optimal IS6110 copy, whose acquisition predated the emergence of the MTBC complex. Background selection acting against deleterious mutations led to an excess of low-frequency variants.

Introduction

Insertion sequences (ISs) are the smallest autonomously transposable mobile genetic elements widely distributed in bacterial genomes [1–4]. IS elements carry in their sequence the genes encoding for a transposase, thus ensuring their mobility in the bacterial genome. IS mobilization and expansion via transposition contribute significantly to genome diversification and plasticity [4–6]. Indeed, IS sequences have been shown to induce mutations, duplications, deletions, inversions, as well as complex genomic rearrangements [4,7]. In many circumstances, insertion of IS elements modulates the expression of neighbouring genes, which may benefit the organism, thus arguing for a consistent role in adaptive evolution, particularly for those bacteria adopting fastidious, host-restricted lifestyles [1,8,9,10,11,12].

Mycobacterium tuberculosis, a highly adapted and host-restricted pathogen, carries in its genome at least 30 different IS elements, one of which, IS6110, is found exclusively in members of the *M. tuberculosis* complex (MTBC) [13–15]. The latter species comprises *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium microti*, *Mycobacterium africanum*, *Mycobacterium pinnipedii* and *Mycobacterium caprae* [15,16]. Strong evidence suggests that the MTBC group has arisen from a pool of mycobacterial species, referred to as smooth tubercle bacilli (STB), including the species “*Mycobacterium canettii*” [17–21]. IS6110 was originally identified in 1990 and proved very useful as a marker for the molecular detection of MTBC strains [22–31]. The highly polymorphic nature of IS6110 in terms of copy number and transpositional location has provided the most widely used fingerprinting approach, IS6110 RFLP. Owing to this typing method, it has been possible to distinguish recent transmission from reactivation, to uncover TB transmission dynamics, and to delineate the population structure of *M. tuberculosis* in diverse settings worldwide [32–34].

IS6110 belongs to the IS3 family of insertion sequences [24]. It consists of a 1361-bp sequence with imperfect 28 bp terminal inverted repeats, whose transposition generates a 3–4 bp duplication at the insertion point. The IS6110 sequence contains two partially overlapping reading frames (ORFs), *orfA* and *orfB*. Based on sequence similarities with the IS3 element of *E. Coli* [35], it is assumed that a -1 translational frameshifting between *orfA* and *orfB* would generate an *orfAB* transframe protein, the IS6110 transposase [34]. The sequence “UUUAAAG” located directly upstream of the IS6110 *orfB* may be responsible for such a frameshifting event [24]. Like IS3 [33–35], IS6110 *orfA* protein contains a helix-turn-helix DNA-binding domain (residues 16 to 58), while the *orfB* protein displays characteristics reminiscent of retrovirus and retrotransposon integrases [36–38].

Compelling evidence suggests that variations in IS6110 transposition induce new strain-specific phenotypic changes, either by mediating genomic rearrangements, or by acting as a promoter sequence that modulates the expression of neighbouring genes [39–46]. Hence, IS6110 is no more viewed as a passive ‘junk’ or ‘selfish parasite’ DNA sequence, but as a significant contributor to the evolution of *M. tuberculosis* [34].

To date, IS6110 has been extensively studied as a transposable element that induces several mutational changes on the host chromosomal DNA. Despite the fact that the transposition process of IS6110 is mediated by its own transposase, a very few studies have focused on the transposase-encoding sequences. In the present study we sought to explore the genetic variability of the IS6110 transposase-encoding ORFs, and determine the selective pressures acting on their coding sequences.

Materials and Methods

Ethics statement

This investigation involved only DNA from Mycobacteria that have previously been described and published. No sputum or any other samples were collected from patients for the specific needs of this study.

M. tuberculosis strains

This study was performed using 63 *M. tuberculosis* (MTB1 to MTB63) clinical strains representing diverse genotypes circulating in Tunisia between 2001 and 2005. This clinical strain collection consisted of the following spoligotypes: 1 Latin American Mediterranean 1 (LAM1) (ST20), 4 LAM4 (ST60), 1 LAM4 (ST828), 2 LAM5 (ST93), 12 LAM9 (ST42), 2 LAM9 (ST177), 1 LAM9 (ST398), 2 LAM9 (ST822), 2 LAM9 (ST1064), 6 Haarlem 1 (H1) (ST47), 1 H1 (ST883), 1 Haarlem 3 (H3) (ST49), 7 H3 (ST50), 1 H3 (ST56), 1 H3 (ST180), 1 H3 (ST871), 1 H3 (ST121), 1 H3 (ST764), 1 T1 (ST7), 10 T1 (ST53), 1 T1 (ST281), 1 T2 (ST52), 1 MANU2 (ST54), 1 S (ST34), and 1 S (ST1536). Assignment of these genotypes was performed based on the SITVITWEB database (http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/). We also included 20 smooth tubercle bacilli (STB1 to STB20) covering 6 genotypes: B (STB1), C/D (STB2-STB15), E (STB16), F (STB17), G (STB18), and H (STB19-STB20). These STB strains have been described earlier [19,20]. Details regarding the origin and year of isolation of the mycobacterial isolates are listed in [S1 Table](#).

PCR amplification and cloning

DNA extraction was performed by the phenol chloroform method [23]. The primer pair IS6110F (5'-ATCTGAACCGCCCCGGCATGTCCGG-3') and IS6110R (5'-ATCTGAACCGCCCCGGTGAGTCCGG-3') was used for PCR amplification of the IS6110 sequence. The amplification reaction mixture contained 20 ng of template genomic DNA, 10 µl of 10x buffer (Qiagen), 10 µl DMSO, 2 µl of 10 mM nucleotide mix (Amersham Biosciences), 2 µl of each primer (20 µM stock), 0.25 µl (1.25 U) of HotStar *Taq* DNA polymerase (Qiagen) and sterile nuclease-free water (Amersham Biosciences) to 50 µl total reaction volume. Cycling was carried out in a 2720 thermocycler (Applied Biosystems) with an initial denaturation step of 10 min at 96°C followed by 35 cycles consisting of 1 min at 95°C, 1 min at 60°C and 2 min at 72°C. The amplification ended with a final elongation step of 7 min at 72°C.

One to 2 µL of each PCR product was directly cloned into the pCR2.1 plasmid vector using the TA cloning kit, following the manufacturer's instructions (Life Technologies). The ligation product was then transformed into chemically competent TOP10 *E. coli*, and 200 µL were plated for blue/white screening using LB agar plates containing 40 µl of X-Gal (40 mg/mL) and ampicillin (100 µg/mL). Ten white colonies from each ligation were allowed to grow overnight in 2YT (containing 100 µg/mL ampicillin), then plasmid DNA was recovered using the QIA-prep spin miniprep kit (Qiagen). Recombinant clones were confirmed by *EcoRI* digestion.

Sequencing

A single IS6110 copy from each mycobacterial strain was subjected to nucleotide sequencing on both strands of the recombinant pCR2.1 plasmid using the M13 reverse and forward primers. Sequencing was performed with the BigDye Terminator Cycle Sequencing Kit. The reaction consisted of 1.5 μ l of BigDye terminator cycle sequencing reagents, 4 μ l of BigDye terminator cycle sequencing buffer, 1 μ l of 20 μ M concentrations of primers, and UltraPURE Distilled DNase/RNase-FreeWater (Gibco/Invitrogen) to make a 20- μ l reaction. Cycle sequencing was performed using a 2720 thermocycler (Applied Biosystems) programmed for 25 cycles at 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min. The template DNA was ethanol-precipitated, washed, and subjected to automated sequencing on an ABI Prism 3130 genetic analyzer (Applied Biosystems) according to the manufacturer's protocol.

Genetic polymorphism and diversity

Sequence data of the 83 IS6110 copies were edited and aligned using BioEdit (<https://www.bioedit.com/>) and ClustalW [47]. The DnaSP software package (version.5.10) [48] was used to carry out several population genetic analyses. For both IS6110 *orfA* and *orfB*, we determined the number of haplotypes (h), the number of polymorphic sites (S), the nucleotide diversity (π), and the per-site population mutation rate, θ ($2N_e\mu$). To test for adaptive selection, we determined the nucleotide substitution changes and the ratio of nonsynonymous (dN) to synonymous (dS) substitutions per site (dN/dS), using the analysis developed by Nei-Gojobori [49] after Jukes-Cantor correction for multiple substitutions.

Phylogenetic analysis

Maximum likelihood (ML) methods were used to infer phylogenetic relationships. ML analyses were performed using PhyML 3.0 [50]. Bootstrap confidence levels were based on 1000 resampling. The trees were visualized using FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

Prior to ML analyses, we determined for each dataset the best-fit model of nucleotide substitution (see S2 Table) using jModelTest 2 [51].

To test for the topological congruence between trees, we computed the *Icong* index, which is based on maximum agreement subtrees (MAST) [52]. This method determines the minimum number of leaves that have to be removed in each phylogeny to render the trees identical. Computation of *Icong* and of the associated *P*-value was performed on line at <http://max2.ese.u-psud.fr/bases/upresa/pages/devienne>.

Tests of recombination

As a first indication of the eventual existence of recombination, we performed a split decomposition analysis to generate phylogenetic networks [53]. The networks were generated in Splitstree 4.6 [54]. Evidence of recombination, indicated by the presence of cycles in the networks, was assessed by the pairwise homoplasy index (PHI). Significance of the PHI statistic is assessed with the normal approximation of a permutation test where, under the null hypothesis of no recombination, sites along the alignment are randomly permuted to obtain the null distribution of PHI. *P* values < 0.05 indicate significant presence of recombination [55]. To confirm the results obtained with the split decomposition analysis, we used two additional recombination detection algorithms:

- a. Hudson and Kaplan's R_{\min} [56]: The Hudson and Kaplan's lower bound on the minimal number of recombination events in an infinite site model was computed using the DnaSP software package (version.5.10) [48].

- b. The Web-based service GARD (genetic algorithm for recombination detection) [57]: a model-based approach that searches for putative breakpoints delimiting sequence regions having distinct phylogenies. Briefly, GARD compares a nonrecombinant model in which the sequence data are fitted to a single phylogeny to models in which breakpoints partition the sequence data into two or more regions having varying phylogenies. Site-by-site substitution rate was assumed to be constant between sites. The identified breakpoints were further confirmed using the akaike information criterion (AIC) score and Kishino-Hasegawa topological incongruence test.

Tests of selection

To search for signals of positive selection acting on IS6110 ORFs, we first measured the ratio of nonsynonymous substitutions to synonymous substitutions, dN/dS ratio (or ω), over the entire length of each ORF. Evidence of positive selection for amino acid replacements is suggested when $\omega > 1$ (adaptive evolution; nonsynonymous changes are favored because they confer a fitness advantage and are fixed at a higher rate than synonymous changes), purifying selection is inferred when $\omega < 1$ (strong functional constraint; nonsynonymous changes are deleterious for protein function and are fixed at a lower rate than synonymous changes), whereas neutral evolution is assumed when $\omega = 1$ (relaxed selective constraint; nonsynonymous changes have no associated fitness advantage and are fixed at the same rate as synonymous changes). Then we performed a codon-by-codon analysis using codeml as implemented in the software package PAML (Phylogenetic Analysis by Maximum Likelihood) v. 4.4e [58]. For this purpose we used “site models” where codon sites are allowed to fall into categories depending on their ω values. First, we compared a “nearly neutral model”, M1a, to a “positive selection” model, M2a. The model M1a allows 2 categories of codon sites in p_0 , and p_1 proportions, with $\omega_0 < 1$ and, $\omega_1 = 1$, whereas M2a adds an additional category of codons (p_2), with ω_2 that is free to vary above 1. In addition to M1a and M2a, we compared several additional site models, M7, M8, and M8a. M7 specifies a neutral model similar to M1a, but the sites affected by negative selection approximate a beta distribution with parameters (p and q) estimated from the data. M7 is compared to M8 (selection) for which the category of sites with a dN/dS > 1 is added. We also compared the model M8 to M8a. In the latter model the extra ω is fixed at 1. Previous studies have shown that the M8-M8a comparison is more robust than the M7-M8 comparison and produces less false positives [59,60].

The comparison between models was assessed using Likelihood-Ratio Tests (LRTs). A significantly higher likelihood of the alternative model than that of the null model indicates positive selection in the data set examined. For models comparisons, we used degree of freedom, $df = 2$. For each analysis, correction for multiple testing (Bonferroni correction) was applied. Only in cases where LRT was significant, we used the Bayes empirical Bayes (BEB) procedure to calculate the posterior probabilities (PPs) to identify sites under positive selection [61].

Nucleotide sequence accession numbers

IS6110 ORFs A and B haplotype nucleotide sequences were deposited in GenBank under accession numbers KP844666 to KP844685 and KP844686 to KP844721, respectively.

Results

Sequence diversity of IS6110 transposase-encoded ORFs A and B

To appreciate the inter-strain genetic diversity of IS6110 transposase-encoded *orfA* and *orfB*, we sequenced a single IS6110 copy from each strain of STB and *M. tuberculosis*. The statistics

Table 1. Patterns of molecular genetic diversity of IS6110 *orfA* and *orfB*.

	No. of polymorphic sites (S)	No. of Haplotypes (h)	Haplotype diversity (Hd)	Nucleotide diversity (π)	per-site population mutation rate (θ)	No. of nsSNPs	No. of sSNPs	dN/dS (ω)	Tajima test statistic (<i>D</i>)
<i>orfA</i> (STB)	13	11	0.763	0.00398	0.01121	8	5	0.58743	-2.34049*
<i>orfA</i> (MTB)	12	10	0.2939	0.00135	0.00779	4	8	0.14906	-2.35117*
<i>orfB</i> (STB)	15	12	0.8105	0.00152	0.00428	9	6	0.39413	-2.38105*
<i>orfB</i> (MTB)	36	25	0.6395	0.00125	0.00775	24	11	0.58168	-2.74469**
<i>orfA</i> (STB + MTB)	25	20	0.4258	0.00199	0.01532	12	13	0.29613	-2.65444**
<i>orfB</i> (STB + MTB)	48	36	0.6820	0.00132	0.00976	32	15	0.51483	-2.79601**

**P* < 0.01

***P* < 0.001

doi:10.1371/journal.pone.0130161.t001

are reported in [Table 1](#). Overall, the *orfA* showed slightly higher nucleotide diversity (π) compared to *orfB* (0.00199 vs 0.00132), as well as a higher population mutation rate (θ) (0.01532 vs 0.00976). As expected, much of the nucleotide diversity in *orfA* was associated with the STB group, which is three fold higher than in today's *M. tuberculosis* strains (0.00398 vs 0.00135). In both ORFs, mutations were distributed throughout the entire sequence ([Fig 1](#)). Only in one case (strain MTB20), did a nonsense mutation occurred, yielding an *orfB* product lacking the 13 C-terminal last residues.

Twenty and 36 haplotypes were identified for IS6110 ORFs A and B (Ha1 to Ha20 and Hb1 to Hb36). Both ORFs displayed low haplotype diversity (0.42 vs 0.68) and showed similar haplotype distribution patterns. Indeed, in each ORF we observed a predominant haplotype, involving 76.19% and 47.22% of the entire strain collection in *orfA* and *orfB*, respectively ([Fig 2](#)). These frequent haplotypes dominate over a flat distribution of rare descendent haplotypes that differ mainly by a single or two mutations. In both ORFs, the predominant haplotype included ancestral (STB) and present-day (MTB) strains ([Fig 2](#)). It is noteworthy that *orfA* and *orfB* sequences of an IS6110 copy of the reference strain H37Rv [[13](#)], each belongs to the respective predominant haplotype.

Tajima's *D* statistics proved significantly negative for both ORFs ([Table 1](#)), which is consistent with their observed haplotype distribution pattern that is characterized by an excess of rare variants (singletons).

Furthermore, both *orfA* and *orfB* evolved under strict purifying selection, since their dN/dS values are < 1 (0.29 and 0.51, respectively). The *orfA* is likely to be the subject of more stringent purifying action compared to *orfB* ([Table 1](#)). Indeed, in *orfB* the ratio nsSNP/sSNP is 2.13 (32/15), which is in the range known for *M. tuberculosis* genome, while this ratio is about 0.92 for *orfA*, suggesting that non synonymous changes tend to be purged during its evolution.

Evaluation of the congruence of *orfA* and *orfB* phylogenies

To better appreciate the evolutionary trends of both IS6110 transposase-encoded ORFs, we constructed their ML trees and assessed their congruence. The ML trees, shown in a polar

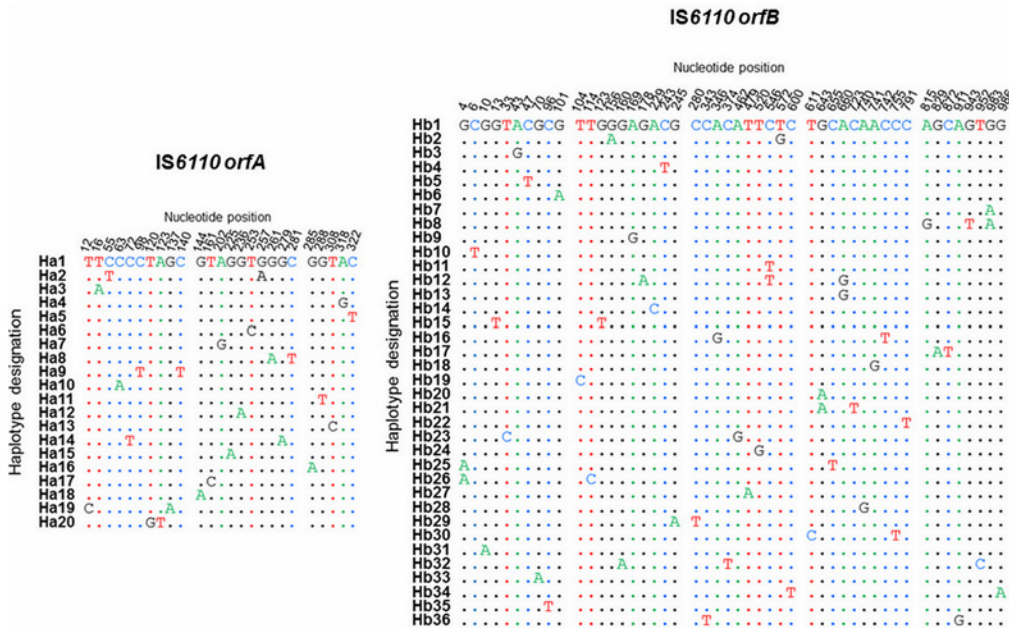


Fig 1. Sequence alignment of IS6110 *orfA* and *orfB* haplotypes.

doi:10.1371/journal.pone.0130161.g001

format, are depicted in Fig 3. Consistent with the haplotype diversity patterns of both ORFs, the tree topography shows virtually no deep branching; the majority of tips being nearly at equivalent genetic distances.

When the strain collection was analyzed as a whole (STB + MTB), the phylogenies of *orfA* and *orfB* were found congruent ($I_{cong} = 3.121$; $P = 2.045e^{-20}$). However, a conflict between *orfA* and *orfB* trees was detected in the STB group ($I_{cong} = 1.199$; $P = 0.083$), but not in today's MTB strains ($I_{cong} = 2.898$; $P = 6.474e^{-19}$).

Lack of evidence for recombination

As a repeat sequence, IS6110 could be involved in recombination-mediated genomic rearrangement events. Therefore, we searched for signals of recombination, and for this purpose, we first performed a split decomposition analysis. The tree topography (Fig 4) and the statistical PHI test did not provide evidence for recombination. This finding was further confirmed with two additional approaches, the Hudson and Kaplan's R_{min} and GARD, which found no evidence for breakpoints.

IS6110 transposase-encoding *orfA* and *orfB* evolved under strict purifying selection

Despite the fact that both IS6110 *orfA* and *orfB* were clearly evolving under purifying selection, as suggested by their dN/dS ratios and their negative Tajima's D statistic, one should not dismiss the possibility that positive selection could have operated on specific codons. For this purpose, we performed a codon-by-codon maximum likelihood test using the program codeml (PAML package). As shown in Table 2, in no case was the difference of the likelihood ratio test significant for both models comparisons (M1a vs M2a and M8a vs M8), indicating absence of positive selection acting at specific codons.

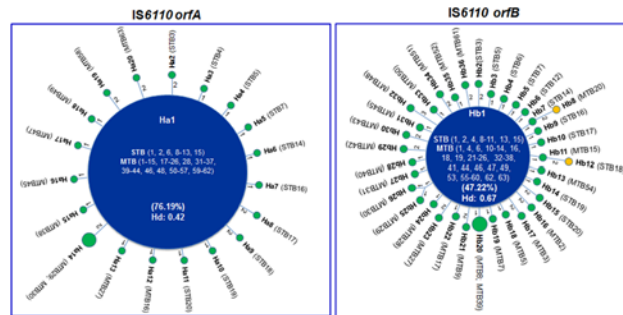


Fig 2. Parsimonious relationships among haplotypes of IS6110 *orfA* and *orfB*. Each IS6110 transposase-encoding haplotype is represented with a filled circle whose size reflects its frequency. The predominant haplotype is shown in blue, while the descendent rare haplotypes are shown in green and orange. Haplotype designations (Ha1 to Ha20 or Hb1 to Hb36 for *orfA* and *orfB*, respectively) and their associated mycobacterial strains are indicated within or next to the circles. The number of mutations dividing the haplotypes is shown next to the lines connecting the circles. Frequency of the predominant haplotype in each ORF is indicated within parenthesis. Hd: haplotype diversity.

doi:10.1371/journal.pone.0130161.g002

Discussion

In the present study we explored the genetic variability and the selective forces acting on the IS6110 transposase-encoding ORFs, *orfA* and *orfB*. For this purpose, we used a strain collection consisting of smooth tubercle bacilli (STB), an early branching lineage of the MTBC, and present-day *M. tuberculosis* strains representing the full breadth of genetic diversity in Tunisia. Based on the roles of the *E. Coli* IS3 insertion sequence [62], it has been suggested that IS6110 *orfA* and *orfB* products may act as negative regulators to limit the number of IS6110 transposition events, thereby minimizing an eventual deleterious effect to the mycobacterial host genome [34]. By testing several models of transposition Tanaka et al. [63] provided evidence for selection against IS6110 in *M. tuberculosis*, a finding in line with the possible roles assigned to the products of *orfA* and *orfB*.

Taken together, our data converge towards an evolutionary scenario according to which IS6110 had been acquired early before emergence of the MTBC. This conclusion stems from the fact that for both ORFs a single haplotype predominated, consisting of admixtures of ancestral and present-day strains. Early acquisition of IS6110 is also suggested by the fact that all strains of the ancestral STB pool tested thus far contain the insertion element IS6110, the copy

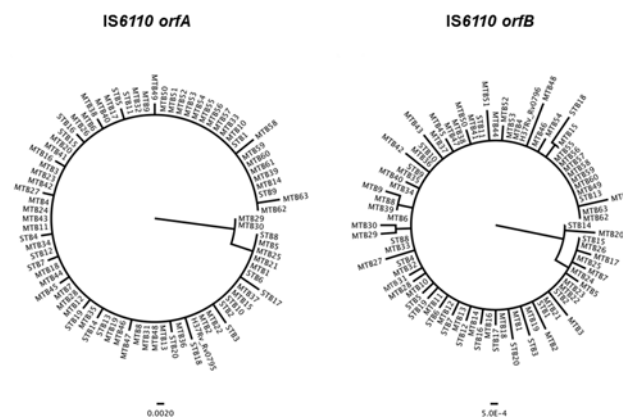


Fig 3. ML Phylogenetic trees showing the relationships of STB and MTB strains based on the IS6110 transposase-encoding *orfA* and *orfB*.

doi:10.1371/journal.pone.0130161.g003

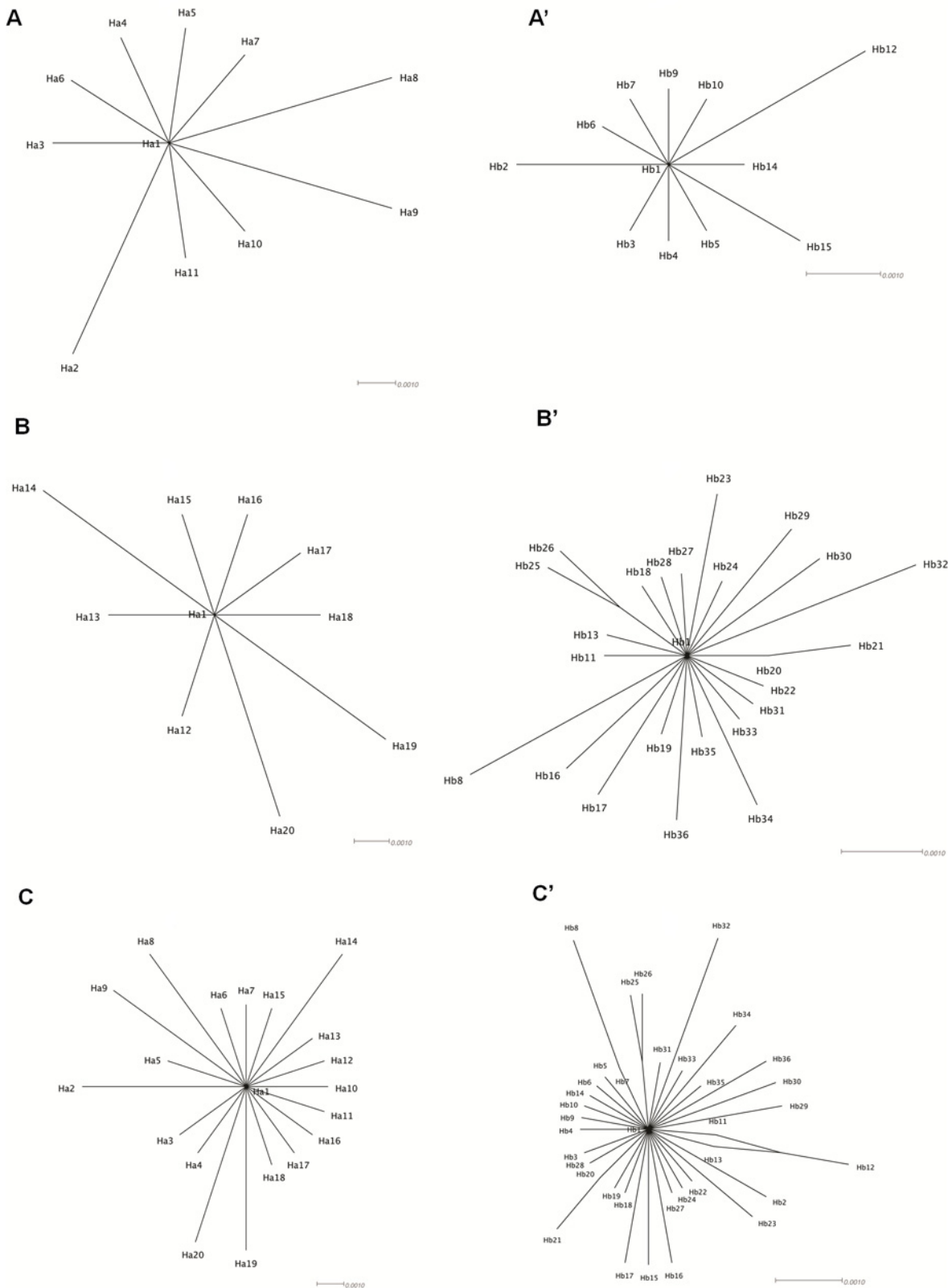


Fig 4. Split decomposition analysis based on IS6110ORFs A and B haplotype sequences. Panels A, C, and D represent the *orfA*-based phylogenetic networks of STB, MTB, and the whole collection (STB + MTB), respectively. Panels A', B', and C' depict the *orfB*-based phylogenetic networks of STB, MTB, and the whole collection (STB + MTB), respectively.

doi:10.1371/journal.pone.0130161.g004

Table 2. Likelihood ratio test (LRT) analysis of models comparisons M1a vs M2a and M8 vs M8a.

	M1a	M2a	M8	M8a
<i>orfA</i> (MTB)	-516.184909	-514.152318	-514.155047	-514.155030
LRT <i>p</i> -value		0.13		0.999984
<i>orfA</i> (STB)	-514.941268	-514.941267	-514.942165	-514.942159
LRT <i>p</i> -value		1		0.999995
<i>orfA</i> (MTB+STB)	-611.532742	-611.532720	-611.541343	-611.540493
LRT <i>p</i> -value		0.999979		0.999151
<i>orfB</i> (MTB)	-1531.882528	-1531.882524	-1532.269623	-1531.885179
LRT <i>p</i> -value		0.999997		0.68083
<i>orfB</i> (STB)	-1400.206582	-1400.206579	-1400.208005	-1400.207932
LRT <i>p</i> -value		0.999998		0.999928
<i>orfB</i> (MTB-STB)	-1659.837451	-1659.837445	-1659.840249	-1659.840053
LRT <i>p</i> -value		0.999995		0.999805

doi:10.1371/journal.pone.0130161.t002

number of which varies from 1 (groups E and F) to 10 (groups G and H) [20]. Hence, the predominant haplotype identified herein, which is common to both STB and present-day *M. tuberculosis* strains, could represent the very early copy inherited from the common ancestor of the STB population. The ancestral origin of IS6110, as confirmed in the present study, is consistent with the identification of IS6110-like sequences from environmental mycobacterial species (*M. smegmatis*, *Mycobacterium* sp. JLS) [64], thus lending further support to the hypothesis that the ancestor of the MTBC could have originated from an environmental mycobacteria [21].

Increased sequence variability of IS6110 ORFs was observed among STB strains compared to *M. tuberculosis*. This finding was somehow expected based on previously published data with other sequence categories, such as houses keeping genes [19,20], PE_PGRS [65], as well as whole genome comparisons of selected STB genomes [21].

Evolution of IS6110 *orfA* and *orfB* appears to have been driven mainly by random point mutations rather than recombination, as witnessed by the distribution pattern of mutations along the two ORF sequences. We could demonstrate no signal of recombination despite the fact that several deletions in *M. tuberculosis* were shown to have resulted from recombinational events between two adjacent IS6110 copies [39,41,42]. However, these events should have involved nearly identical IS6110 sequence copies and therefore no imprints of recombination are left.

The genealogies of *orfA* and *orfB* were congruent in today's *M. tuberculosis* strains, a finding in line with their cooperative role in regulating transpositional recombination [62]. By contrast, a conflict between the two genealogies was observed in STB, the underlying mechanism of which remains to be determined in the absence of demonstrable recombination.

Both IS6110 ORFs were found to evolve under strict purifying selection. Indeed, the majority of variants consisted of singletons that differed mainly by one or two mutations, with a clear tendency to purge non synonymous changes ($dN/dS < 1$). Furthermore, PAML analysis could not detect any positively selected residue along the amino acid sequence of both ORFs. Given the ancestral origin of IS6110, these findings strongly suggest that the MTBC ancestor could have acquired, from the outset, a functionally optimal IS6110 copy that does not tolerate further amino acid changes. Therefore, most of the non synonymous changes that were detected are likely to be neutral or slightly deleterious [66].

The low dN/dS rate ratios in several bacterial transposase genes may not only be linked to purifying selection, but could also result from periodic extinction events of IS elements

followed by the acquisition of evolutionarily young copies [67]. In *Wolbachia*, the low nucleotide divergence rates of IS sequences has been proposed to be due to recent import of IS sequences, via horizontal gene transfer, coupled to subsequent bursts of transposition [68]. These mechanisms could not have accounted in the evolution of the IS6110 transposase in MTBC species, since they have undergone a clonal evolution with virtually no genetic exchanges, thus pointing to the prominent role of purifying selection.

However, positive selection was shown to act on IS transposases of other bacterial species, such as *E. coli* and the cyanobacterium *Crocospaera watsonii* [69,70]. In *E. coli*, positive selection was found to operate on the IS30 and IS1 transposase genes. In IS30, evidence of positive selection could be detected in 16 sites, of which 14 occur in the N-terminal helix-turn-helix motif 1 (HTH1), thus favoring particular sites to be frequently targeted by transposition. As far as could be ascertained, IS6110 does not have a known target for insertion, and despite the existence of insertion hotspots, it tends to integrate the mycobacterial genome at random [71–77]. This is consistent with the absence of positive selection acting on its transposase sequence. However, one should not dismiss the possibility that the IS6110 transposase sequence could have been the subject of positive selection early in its evolutionary history.

In line with its critical role in IS6110 transpositional control, the *orfA* was found to be the subject of more stringent purifying selection compared to *orfB* (nsSNP/sSNP: 0.92 vs 2.13). Indeed, in the IS3 element of *E. coli*, both ORFs were shown to act as inhibitors of transpositional recombination. It has been demonstrated that such inhibition could be mediated by the *orfA* product alone, while *orfB* by itself has no inhibitory activity, but enhances the inhibitory activity of *orfA*, probably by interacting with *orfA* to form a complex. In other words, *orfB* exerts its negative effect on transpositional recombination only when *orfA* is produced, and thus functions in cooperation with *orfA* [62]. Consequently, for an optimal control of IS6110 transposition activity, background selection should act more frequently on *orfA*, which we demonstrate herein.

Conclusions

The haplotype distribution characterizing the mycobacterial IS6110 transposase-encoding ORFs A and B strongly suggests its ancestral origin, which most likely predated emergence of the MTBC. These ORFs evolved essentially by point mutations under strict purifying selection acting against deleterious mutations, thus leading to an excess of low-frequency variants. Aside from purging non synonymous changes, no imprints of positive selection acting on single amino acid residues could be detected in both STB and present-day *M. tuberculosis* strains, arguing that the IS6110 copy they had inherited was functionally optimal. Finally, the fact that *orfA* was the subject of more stringent purifying selection compared to *orfB*, lends further support to its essential role in regulating the IS6110 transpositional process.

Supporting Information

S1 Table. Characteristics of STB and *M. tuberculosis* strain collections used in this study. (DOCX)

S2 Table. The best-fit model of nucleotide substitution as determined by jModelTest 2 [50]. (DOCX)

Acknowledgments

The authors are indebted to Cristina Gutierrez and Michel Fabre for providing DNAs of STB strains. The authors thank Maherzia Ben Fadhel for her assistance with nucleotide sequencing. We thank Nédra Meftahi for her helpful inputs with figures.

This study received financial support from the Tunisian Ministry of Higher Education and Scientific Research.

Author Contributions

Conceived and designed the experiments: HM. Performed the experiments: ST. Analyzed the data: ST AN HM. Contributed reagents/materials/analysis tools: ST AN HM. Wrote the paper: ST HM.

References

1. Mahillon J, Chandler M (1998) Insertion sequences. *Microbiol Mol Biol Rev* 62:725–774. PMID: [9729608](#)
2. Siguier P, Filée J, Chandler M (2006) Insertion sequences in prokaryotic genomes. *Curr Opin Microbiol* 9:526–531. PMID: [16935554](#)
3. Filée J, Siguier P, Chandler M (2007) Insertion sequence diversity in archaea. *Microbiol Mol Biol Rev* 71:121–157. PMID: [17347521](#)
4. Siguier P, Gourbeyre E, Chandler M (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev* 38:865–891. doi: [10.1111/1574-6976.12067](#) PMID: [24499397](#)
5. Charlier D, Piette J, Glansdorff N (1982) IS3 can function as a mobile promoter in *E. coli*. *Nucleic Acids Res* 10:5935–5948. PMID: [6292860](#)
6. Lysnyansky I, Calcutt MJ, Ben-Barak I, Ron Y, Levisohn S, Methé BA, et al. (2009) Molecular characterization of newly identified IS3, IS4 and IS30 insertion sequence-like elements in *Mycobacterium bovis* and their possible roles in genome plasticity. *FEMS Microbiol Lett* 294:172–182. doi: [10.1111/j.1574-6968.2009.01562.x](#) PMID: [19416360](#)
7. Preston A, Parkhill J, Maskell DJ (2004) The Bordetellae: lessons from genomics. *Nat Rev Micro* 2: 379–390.
8. Parkhill J, Thomson N (2003) Evolutionary strategies of human pathogens. *Cold Spring Harb Symp Quant Biol* 68:151–158. PMID: [15338613](#)
9. Schneider D, Lenski RE (2004) Dynamics of insertion sequence elements during experimental evolution of bacteria. *Res Microbiol* 155:319–327. PMID: [15207863](#)
10. Biémont C, Vieira C (2006) Genetics: junk DNA as an evolutionary force. *Nature* 443:521–524. PMID: [17024082](#)
11. Ooka T, Ogura Y, Asadulghani M, Ohnishi M, Nakayama K, Terajima J, et al. (2009). Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in *Escherichia coli* O157 genomes. *Genome Res* 19:1809–1816. doi: [10.1101/gr.089615.108](#) PMID: [19564451](#)
12. Stoebel DM, Dorman CJ (2010) The effect of mobile element IS10 on experimental regulatory evolution in *Escherichia coli*. *Mol Biol Evol* 27:2105–2112. doi: [10.1093/molbev/msq101](#) PMID: [20400481](#)
13. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393: 537–544. PMID: [9634230](#)
14. Gordon SV, Heym B, Parkhill J, Barrell B, Cole ST (1999) New insertion sequences and a novel repeated sequence in the genome of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 145:881–892. PMID: [10220167](#)
15. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. (2002) A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A* 99: 3684–3689. PMID: [11891304](#)
16. Wirth T, Hildebrand F, Allix-Béguec C, Wölbelling F, Kubica T, Kremer K, et al. (2008) Origin, spread and demography of the *Mycobacterium tuberculosis* complex. *PLoS Pathog* 4: e1000160. doi: [10.1371/journal.ppat.1000160](#) PMID: [18802459](#)
17. Canetti G (1970) Infection by atypical mycobacteria and antituberculous immunity. *Lille Med* 15: 280–282. PMID: [5446090](#)
18. van Soolingen D, Hoogenboezem T, de Haas PE, Hermans PW, Koedam MA, Teppema KS, et al. (1997) A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. *Int J Syst Bacteriol* 47: 1236–1245. PMID: [9336935](#)
19. Fabre M, Koeck JL, Le Flèche P, Simon F, Hervé V, Vergnaud G, et al. (2004) High genetic diversity revealed by variable-number tandem repeat genotyping and analysis of hsp65 gene polymorphism in a

- large collection of "*Mycobacterium canettii*" strains indicates that the *M. tuberculosis* complex is a recently emerged clone of "*M. canettii*". J Clin Microbiol 42: 3248–3255. PMID: [15243089](#)
20. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omaïs B, Marmiesse M, et al. (2005) Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. PLoS Pathog 1: e5. PMID: [16201017](#)
 21. Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, et al. (2013) Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. Nat Genet 45: 172–179. doi: [10.1038/ng.2517](#) PMID: [23291586](#)
 22. Thierry D, Cave MD, Eisenach KD, Crawford JT, Bates JH, Gicquel B, et al. (1990) IS6110, an IS-like element of *Mycobacterium tuberculosis* complex. Nucleic Acids Res 18:188. PMID: [2155396](#)
 23. Hermans PW, van Soolingen D, Dale JW, Schuitema AR, McAdam RA, Catty D, et al. (1990) Insertion element IS986 from *Mycobacterium tuberculosis*: a useful tool for diagnosis and epidemiology of tuberculosis. J Clin Microbiol 28:2051–2058. PMID: [1977765](#)
 24. McAdam RA, Hermans PW, van Soolingen D, Zainuddin ZF, Catty D, van Embden JD, et al. (1990) Characterization of a *Mycobacterium tuberculosis* insertion sequence belonging to the IS3 family. Mol Microbiol 4:1607–1613. PMID: [1981088](#)
 25. Thierry D, Matsiota-Bernard P, Pitsouni E, Costopoulos C, Guesdon JL (1993). Use of the insertion element IS6110 for DNA fingerprinting of *Mycobacterium tuberculosis* isolates presenting various profiles of drug susceptibility. FEMS Immunol Med Microbiol 6:287–297. PMID: [8098974](#)
 26. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. (1993) Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 31:406–409. PMID: [8381814](#)
 27. van Soolingen D, de Haas PE, Hermans PW, Groenen PM, van Embden JD (1993) Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. J Clin Microbiol 31:1987–1995. PMID: [7690367](#)
 28. Dalovisio JR, Montenegro-James S, Kemmerly SA, Genre CF, Chambers R, Greer D, et al. (1996) Comparison of the amplified *Mycobacterium tuberculosis* (MTB) direct test, Amplicor MTB PCR, and IS6110-PCR for detection of MTB in respiratory specimens. Clin Infect Dis 23:1099–1106. PMID: [8922809](#)
 29. Maurya AK, Kant S, Kushwaha RA, Nag VL, Kumar M, Dhole TN, et al. (2011) The advantage of using IS6110-PCR vs. BACTEC culture for rapid detection of *Mycobacterium tuberculosis* from pleural fluid in northern India. Biosci Trends 5:159–164. doi: [10.5582/bst.2011.v5.4.159](#) PMID: [21914951](#)
 30. Maurya AK, Kant S, Nag VL, Kushwaha R, Dhole TN (2012) Detection of 123 bp fragment of insertion element IS6110 *Mycobacterium tuberculosis* for diagnosis of extrapulmonary tuberculosis. Indian J Med Microbiol 30:182–186. doi: [10.4103/0255-0857.96688](#) PMID: [22664434](#)
 31. Aryan E, Makvandi M, Farajzadeh A, Huygen K, Alvandi AH, Gouya MM, et al. (2013) Clinical value of IS6110-based loop-mediated isothermal amplification for detection of *Mycobacterium tuberculosis* complex in respiratory specimens. J Infect 66:487–493. doi: [10.1016/j.jinf.2013.02.005](#) PMID: [23466595](#)
 32. Barnes PF, Cave MD (2003) Molecular Epidemiology of Tuberculosis. N Engl J Med 349:1149–1156. PMID: [13679530](#)
 33. Mathema B, Kurepina NE, Bifani PJ, Kreiswirth BN (2006) Molecular epidemiology of tuberculosis: current insights. Clin Microbiol Rev 19:658–685. PMID: [17041139](#)
 34. McEvoy C R, Falmer AA, Gey van Pittius NC, Victor TC, van Helden PD, Warren RM, et al. (2007) The role of IS6110 in the evolution of *Mycobacterium tuberculosis*. Tuberculosis 87:393–404. PMID: [17627889](#)
 35. Sekine Y., Eisaki N, Ohtsubo E (1994) Translational control in production of transposase and in transposition of insertion sequence IS3. J Mol Biol 235: 1406–1420. PMID: [8107082](#)
 36. Prère MF, Chandler M, Fayet O (1990) Transposition in *Shigella dysenteriae*: isolation and analysis of IS911, a new member of the IS3 group of insertion sequences. J Bacteriol 172:4090–4099. PMID: [2163395](#)
 37. Fayet O, Ramond P, Polard P, Prère MF, Chandler M (1990) Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences? Mol Microbiol 4:1771–1777. PMID: [1963920](#)
 38. Khan E, Mack JPG, Katz RA, Kulkosky J, Skalka AM (1991) Retroviral integrase domains: DNA binding and the recognition of LTR sequences. Nucl Acids Res 19:851–860. PMID: [1850126](#)
 39. Brosch R, Philipp W J, Stavropoulos E, Colston M J, Cole S T, Gordon SV, et al. (1999) Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra strain. Infect Immun 67:5768–5774. PMID: [10531227](#)

40. Beggs ML, Eisenach KD, Cave MD (2000) Mapping of IS6110 insertion sites in two epidemic strains of *Mycobacterium tuberculosis*. *J Clin Microbiol* 38:2923–2928. PMID: [10921952](#)
41. Sampson SL, Warren RM, Richardson M, Victor TC, Jordaan AM, van der Spuy GD, et al. (2003) IS6110-mediated deletion polymorphism in the direct repeat region of clinical isolates of *Mycobacterium tuberculosis*. *J Bacteriol* 185:2856–2866. PMID: [12700265](#)
42. Sampson SL, Richardson M, Van Helden PD, Warren RM (2004) IS6110-mediated deletion polymorphism in isogenic strains of *Mycobacterium tuberculosis*. *J Clin Microbiol* 42:895–898. PMID: [14766883](#)
43. Safi H, Barnes PF, Lakey DL, Shams H, Samten B, Vankayalapati R, et al. (2004) IS6110 functions as a mobile, monocyte-activated promoter in *Mycobacterium tuberculosis*. *Mol Microbiol* 52:999–1012. PMID: [15130120](#)
44. Soto CY, Menendez MC, Perez E, Samper S, Gomez AB, Garcia MJ, et al. (2004). IS6110 Mediates Increased Transcription of the phoP Virulence Gene in a Multidrug-Resistant Clinical Isolate Responsible for Tuberculosis Outbreaks. *J Clin Microbiol* 42:212–219. PMID: [14715755](#)
45. Yesilkaya H, Dale JW, Strachan NJ, Forbes KJ (2005) Natural transposon mutagenesis of clinical isolates of *Mycobacterium tuberculosis*: how many genes does a pathogen need? *J. Bacteriol* 187:6726–6732. PMID: [16166535](#)
46. Alonso H, Aguilo JI, Samper S, Caminero JA, Campos-Herrero MI, Gicquel B, et al. (2011) Deciphering the role of IS6110 in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237. *Tuberculosis (Edinb)* 91:117–126. doi: [10.1016/j.tube.2010.12.007](#) PMID: [21256084](#)
47. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680. PMID: [7984417](#)
48. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452. doi: [10.1093/bioinformatics/btp187](#) PMID: [19346325](#)
49. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3: 418–426. PMID: [3444411](#)
50. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O, et al. (2010) "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0." *Systematic Biology* 59:307–321. doi: [10.1093/sysbio/syq010](#) PMID: [20525638](#)
51. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9:772. doi: [10.1038/nmeth.2109](#) PMID: [22847109](#)
52. de Vienne DM, Giraud T, Martin OC (2007) A congruence index for testing topological similarity between trees. *Bioinformatics* 23: 3119–3124. PMID: [17933852](#)
53. Bandelt HJ, Dress AW (1992) Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Mol Phylogenet Evol* 1: 242–252. PMID: [1342941](#)
54. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23: 254–267. PMID: [16221896](#)
55. Bruen TC, Philippe H, Bryant D (2006) A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172: 2665–2681. PMID: [16489234](#)
56. Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111: 147–164. PMID: [4029609](#)
57. Kosakovskiy SL, Posada D, Gravenor MB, Woelk CH, Frost SD (2006) GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22: 3096–3098. PMID: [17110367](#)
58. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586–1591. PMID: [17483113](#)
59. Swanson WJ, Nielsen R, Yang Q (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol* 20: 18–20. PMID: [12519901](#)
60. Wong WS, Yang Z, Goldman N, Nielsen R (2004). Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 168: 1041–1051. PMID: [15514074](#)
61. Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22: 1107–1118. PMID: [15689528](#)
62. Sekine Y, Izumi K, Mizuno T, Ohtsubo E (1997) Inhibition of transpositional recombination by OrfA and OrfB proteins encoded by insertion sequence IS3. *Genes Cells* 2:547–557. PMID: [9413996](#)
63. Tanaka MM, Rosenberg NA, Small PM (2004) The control of copy number of IS6110 in *Mycobacterium tuberculosis*. *Mol Biol Evol* 21:2195–2201. PMID: [15317877](#)

64. Coros A, DeConno E, Derbyshire KM (2008) IS6110, a *Mycobacterium tuberculosis* complex-specific insertion sequence, is also present in the genome of *Mycobacterium smegmatis*, suggestive of lateral gene transfer among mycobacterial species. *J Bacteriol* 190:3408–3410. doi: [10.1128/JB.00009-08](https://doi.org/10.1128/JB.00009-08) PMID: [18326566](https://pubmed.ncbi.nlm.nih.gov/18326566/)
65. Namouchi A, Karboul A, Fabre M, Gutierrez MC, Mardassi H (2013) Evolution of smooth tubercle Bacilli PE and PE_PGRS genes: evidence for a prominent role of recombination and imprint of positive selection. *PLoS One* 8:e64718. doi: [10.1371/journal.pone.0064718](https://doi.org/10.1371/journal.pone.0064718) PMID: [23705005](https://pubmed.ncbi.nlm.nih.gov/23705005/)
66. Nei M (2005) Selectionism and neutralism in molecular evolution. *Mol Biol Evol* 22:2318–2342. PMID: [16120807](https://pubmed.ncbi.nlm.nih.gov/16120807/)
67. Wagner A (2006). Periodic extinctions of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes. *Mol Biol Evol* 23:723–733. PMID: [16373392](https://pubmed.ncbi.nlm.nih.gov/16373392/)
68. Cerveau N, Leclercq S, Leroy E, Bouchon D, Cordaux R (2011) Short- and long-term evolutionary dynamics of bacterial insertion sequences: insights from *Wolbachia* endosymbionts. *Genome Biol Evol* 3:1175–1186. doi: [10.1093/gbe/evr096](https://doi.org/10.1093/gbe/evr096) PMID: [21940637](https://pubmed.ncbi.nlm.nih.gov/21940637/)
69. Petersen L, Bollback JP, Dimmic M, Hubisz M, Nielsen R (2007) Genes under positive selection in *Escherichia coli*. *Genome Res* 17: 1336–1343. PMID: [17675366](https://pubmed.ncbi.nlm.nih.gov/17675366/)
70. Mes TH, Doeleman M (2006) Positive selection on transposase genes of insertion sequences in the *Crocospaera watsonii* genome. *J Bacteriol* 188:7176–7185. PMID: [17015656](https://pubmed.ncbi.nlm.nih.gov/17015656/)
71. Fang Z, Forbes KJ (1997) A *Mycobacterium tuberculosis* IS6110 preferential locus (ipl) for insertion into the genome. *J Clin Microbiol* 35:479–481. PMID: [9003621](https://pubmed.ncbi.nlm.nih.gov/9003621/)
72. McHugh TD, Gillespie SH (1998) Nonrandom association of IS6110 and *Mycobacterium tuberculosis*: implications for molecular epidemiological studies. *J Clin Microbiol* 36:1410–1413. PMID: [9574716](https://pubmed.ncbi.nlm.nih.gov/9574716/)
73. Hermans P W, van Soolingen D, Bik EM, de Haas PE, Dale JW, van Embden JD, et al. (1991) Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect Immun* 59:2695–2705. PMID: [1649798](https://pubmed.ncbi.nlm.nih.gov/1649798/)
74. Vera-Cabrera L, Hernandez-Vera MA, Welsh O, Johnson WM, Castro-Garza J (2001) Phospholipase region of *Mycobacterium tuberculosis* is a preferential locus for IS6110 transposition. *J Clin Microbiol* 39:3499–3504. PMID: [11574563](https://pubmed.ncbi.nlm.nih.gov/11574563/)
75. Kim EY, Nahid P, Hopewell PC, Kato-Maeda M (2010) Novel hot spot of IS6110 insertion in *Mycobacterium tuberculosis*. *J Clin Microbiol* 48:1422–1424. doi: [10.1128/JCM.02210-09](https://doi.org/10.1128/JCM.02210-09) PMID: [20147648](https://pubmed.ncbi.nlm.nih.gov/20147648/)
76. Reyes A, Sandoval A, Cubillos-Ruiz A, Varley KE, Hernández-Neuta I, Samper S, et al. (2012) IS-seq: a novel high throughput survey of in vivo IS6110 transposition in multiple *Mycobacterium tuberculosis* genomes. *BMC Genomics* 13:249. doi: [10.1186/1471-2164-13-249](https://doi.org/10.1186/1471-2164-13-249) PMID: [22703188](https://pubmed.ncbi.nlm.nih.gov/22703188/)
77. Alonso H, Samper S, Martín C, Otal I (2013) Mapping IS6110 in high-copy number *Mycobacterium tuberculosis* strains shows specific insertion points in the Beijing genotype. *BMC Genomics* 14:422. doi: [10.1186/1471-2164-14-422](https://doi.org/10.1186/1471-2164-14-422) PMID: [23800083](https://pubmed.ncbi.nlm.nih.gov/23800083/)