Check for updates

Prolonged survival of a patient with active MDR-TB HIV co-morbidity: insights from a *Mycobacterium tuberculosis* strain with a unique genomic deletion

Mor Rubinstein[1†], Andrei Makhon[1†], Yelena Losev[1], Gal Zizelski Valenci[1], Yair E. Gatt[2], Hanah Margalit[2], Ephraim Fass[1], Ina Kutikov[1], Omer Murik[3], David A. Zeevi[3], Michal Savyon[4], Luba Tau[5,6], Hasia Kaidar Shwartz[1], Zeev Dveyrin[1], Efrat Rorman[1] and Israel Nissan[1*]

[1]National Public Health Laboratory, Public Health Directorate, Ministry of Health, Tel Aviv, Israel, [2]Department of Microbiology and Molecular Genetics, Institute for Medical Research Israel-Canada, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel, [3]Translational Genomics Laboratory, Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem, Israel, [4]Tel Aviv District Health Office, Ministry of Health, Tel Aviv, Israel, [5]Tel Aviv Sourasky Medical Center, Tel Aviv, Israel, [6]Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

Coinfection of HIV and multidrug-resistant tuberculosis (MDR-TB) presents significant challenges in terms of the treatment and prognosis of tuberculosis, leading to complexities in managing the disease and impacting the overall outcome for TB patients. This study presents a remarkable case of a patient with MDR-TB and HIV coinfection who survived for over 8 years, despite poor treatment adherence and comorbidities. Whole genome sequencing (WGS) of the infecting *Mycobacterium tuberculosis* (*Mtb*) strain revealed a unique genomic deletion, spanning 18 genes, including key genes involved in hypoxia response, intracellular survival, immunodominant antigens, and dormancy. This deletion, that we have called "Del-X," potentially exerts a profound influence on the bacterial physiology and its virulence. Only few similar deletions were detected in other non-related *Mtb* genomes worldwide. *In vivo* evolution analysis identified drug resistance and metabolic adaptation mutations and their temporal dynamics during the patient's treatment course.

Introduction

Tuberculosis (TB) is a major cause of morbidity and one of the leading causes of death worldwide (1). TB/HIV coinfection and drug resistant TB pose challenges to the TB elimination efforts. Effective treatment of TB, particularly multidrug-resistant TB (MDR-TB) and extensively drug resistant TB (XDR-TB) is long, expensive and toxic. The long treatment duration [6–22 months (2)], and severe adverse effect of anti MDR-TB medications often lead to poor

patient adherence to the treatment, putting not only these patients, but also the people in their close proximity, at risk (3, 4). Discontinuous treatment and partial treatment are two leading causes for drug resistance (5).

Coinfection of TB and HIV exacerbates the symptoms of both diseases in a synergetic manner, as both pathologies impair the immune system's ability to defend against these pathogens (6). Coinfection with HIV significantly increases mortality among TB patients (7). For example, a study conducted in Eastern Europe reported a median survival time of 5.9 years following diagnosis with MDR-TB or XDR-TB, which decreases to 1.9 years for patients also diagnosed with HIV (8). Additional risk factors for death among TB patients include alcoholism, male gender and old age, among others (7, 9).

Mycobacterium tuberculosis (Mtb), the major causative agent of TB, can exist inside the human host either in an active, replicating state, or in a dormant state. Dormancy is a programed response which enables *Mtb* to survive inside macrophage phagosomes under conditions of starvation, hypoxia, low pH, and oxidative and nitrosative stress. During dormancy, bacteria cease replication, downregulate central metabolism, and switch to anaerobic metabolism (10). The induction of dormancy is regulated by the dormancy survival DosR regulon (11, 12). This dormant state allows latent infection that can persist for many years or even a lifetime without clinical symptoms. However, if a critical number of bacteria manage to evade the host immune system and resume an active, replicating state, the patient develops an active, symptomatic TB disease (13, 14).

The chronic nature of infection with *Mtb* allows for within-host evolution of the infecting bacteria over time. This process involves the accumulation of mutations, which are believed to contribute to the pathogen's adaptation to the environmental conditions within the host. Indirect evidence supporting the selective advantage of these mutations is the observation that they often occur in the same genes across different patients, suggesting homoplastic evolution. A major mycobacterial function influenced by these mutations is drug resistance, frequently accompanied by compensatory mutations that mitigate the fitness cost of drug resistance mutations. Genes linked to virulence, lipid and cell wall metabolism and glycerol-3-phosphate metabolism have also been demonstrated to accumulate mutations during host infection (15, 16).

Compared to other bacteria, the genome of *Mtb* is highly conserved. Modern *Mtb* strains do not typically acquire new genomic features through horizontal gene transfer (17, 18). However, *Mtb* strains often lose fragments of their genomes. Baena et al. (19) found 4,111 genomic deletions longer than 1,000 bp in 522 *Mtb* isolates of L4 lineage. One of those deletions, that is 6,479 bp long and results in the loss of 10 genes, was observed in 249 genomes from four different countries. Other large deletions were also detected in multiple strains. Most of these strains were obtained from sputum, which is produced exclusively during active respiratory tuberculosis. Assuming each unique deletion originated from a distinct event, it is therefore evident that some large genomic deletions do not impede pathogeny and transmission.

Israel has maintained ongoing surveillance of all TB cases since 1997. Patients with suspected pulmonary TB in Israel are requested to provide sputum or other biological specimens for culture. All isolates are sent by primary laboratories to the National Mycobacterium Reference Laboratory (NMRL) in Tel Aviv, Israel. NMRL identifies the strain, performs drug susceptibility tests, and maintains a bank of strains as part of routine work. Since 2019, all new TB strains are fully sequenced (WGS) for epidemiological purposes and for drug susceptibility prediction. Between 2010 and 2019, NMRL received 35 TB samples from a patient diagnosed with both TB and HIV. These samples were found to belong to a single clone of MDR-TB characterized by a unique genomic deletion spanning 18 genes. In spite of the patient's long progressive AIDS and active TB, incomplete treatment, and multiple risk factors, this patient survived for almost 9 years. Here we describe this exceptional case, and discuss the distinct features of the infecting pathogen that may have contributed to this survival.

Results

Case description

A man in his late 50's was diagnosed with multidrug resistant tuberculosis (MDR-TB) in 2010, 5 years after being diagnosed with HIV. Over the next eight and a half years, he suffered from this severe and notorious co-morbidity, and eventually passed away in 2019 at his late 60s (Figure 1). The recorded causes of deaths included TB, alcoholism, AIDS, and acute respiratory failure.

The patient, who will be referred to as "Mr. X" from this point on, was born in Russia and immigrated to Israel. Shortly after his immigration he was diagnosed with tuberculosis. His compliance with medical treatment was poor and sporadic, therefore his medical records were sparse, difficult to track and incomplete.

Throughout the years, Mr. X started treatment multiple times, either in community settings or during hospitalization. Importantly, he never completed any of his therapies and often disappeared from the medical authorities' radar. The 35 TB positive sputum samples collected in each of his therapy were sent to NMRL and tested for anti-TB drugs susceptibility (Figure 1; Supplementary Table S1). Eleven samples were subjected to Whole Genome Sequencing (WGS), which confirmed that the infection comprised a single clone of *Mtb* Beijing lineage throughout the entire illness duration.

Initially, Mr. X was treated using standard therapy for sensitive TB, i.e., a combination of rifampicin (RIF), isoniazid (INH), ethambutol (EMB), and pyrazinamide (PZA; Figure 1). However, after the laboratory reported that the infecting strain is resistant to RIF, INH and PZA, these drugs were replaced, and within 2 months of the first TB diagnosis, his therapy regimens included primarily second line drugs. Overall, he was treated with 13 different anti-tuberculosis antibiotics (Figure 1).

Despite his HIV infection, Mr. X firmly refused antiretroviral therapy (ART). In 2017, his CD4 count was around 80/μL, corresponding to 10% of lymphocytes, with CD4/CD8 ratio lower than 0.2 and a viral load near 350,000 per mL. In the following year, his tests revealed a CD4 count around 80 per μL (10%), CD4/CD8 ratio lower than 0.2, and viral load around 900,000 per mL. These tests indicate an active AIDS, with a severely compromised immune system.

In light of his HIV and MDR-TB comorbidity, his poor compliance to treatment, his age, male gender and alcoholism, the survival of this patient around 9 years is very unusual.
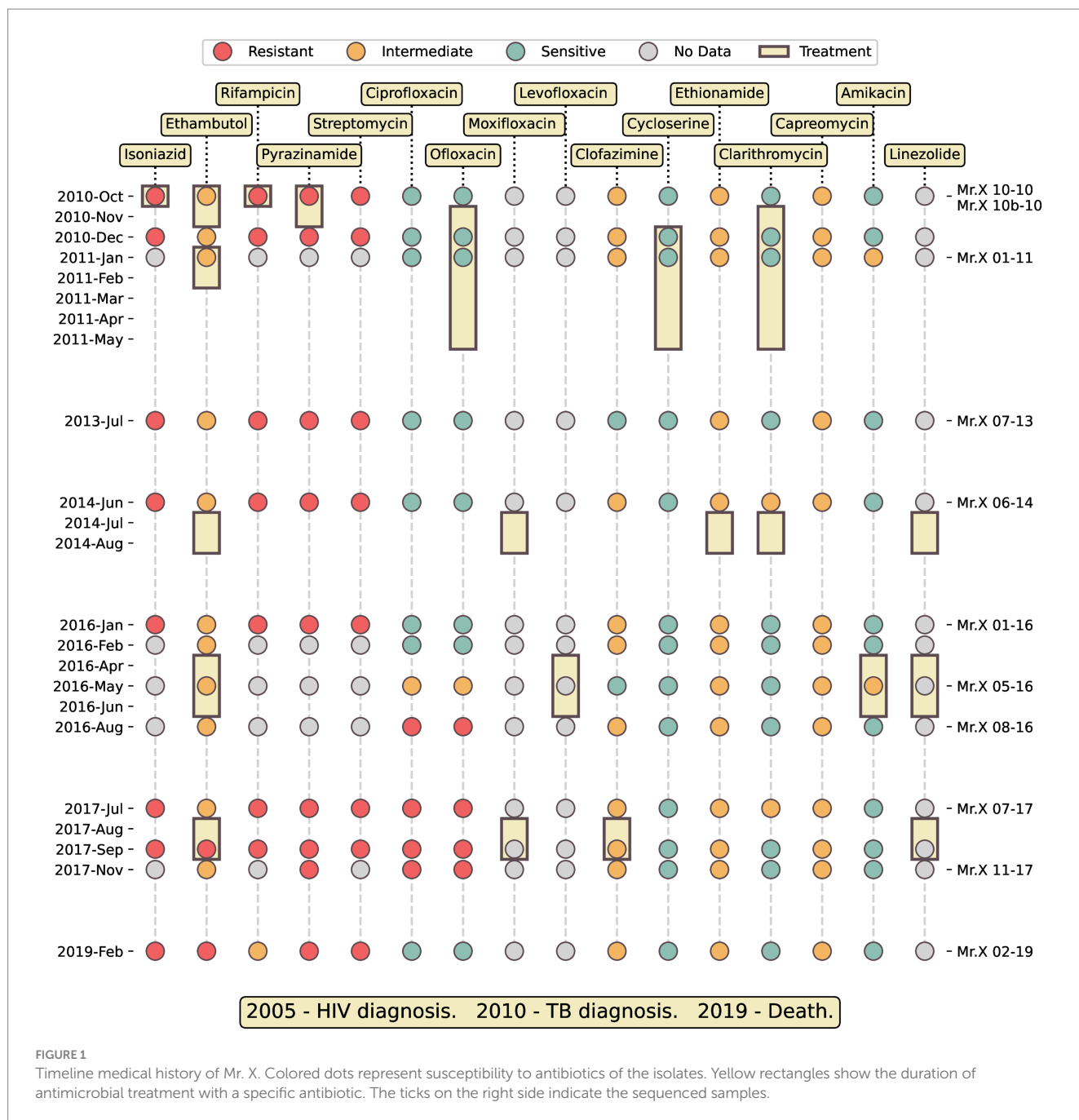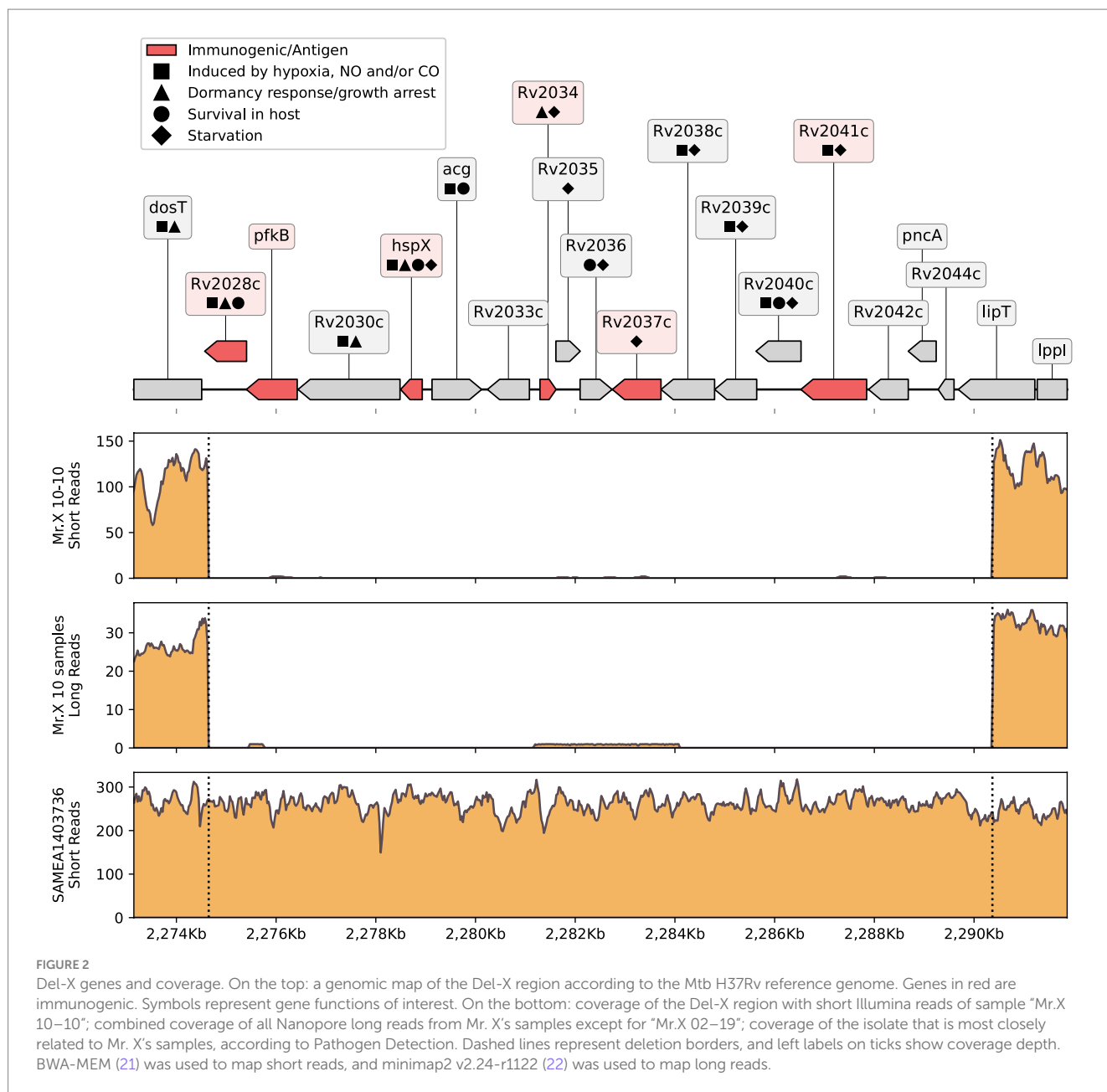
FIGURE 1
Timeline medical history of Mr. X. Colored dots represent susceptibility to antibiotics of the isolates. Yellow rectangles show the duration of antimicrobial treatment with a specific antibiotic. The ticks on the right side indicate the sequenced samples.

Drug resistance mutations

Drug Susceptibility Testing (DST) results are shown in Figure 1; Supplementary Table S1. Antibiotic resistance was also predicted based on whole genome sequencing (WGS), by mapping short reads of each TB isolate to the reference genome, and comparing the variations identified to a database of known resistance mutations (20). The genomic resistance prediction generally agreed with the phenotypic resistance determined by DST. We identified mutations that are known to confer resistance to INH, RIF, EMB, and STM (Supplementary Table S2). The situation was different for PZA resistance. The gene *pncA* encodes an enzyme required for the transformation of PZA into its active form, pyrazinoic acid. Genetic

resistance to PZA is usually caused by mutations in this gene (20). However, while DST results showed PZA resistance, no mutation was found that could account for it. Upon careful inspection, we noticed there was no coverage of the genomic region around *pncA* by the strain's short reads. This uncovered region proved to be a genomic deletion of size 15.7 kb (Del-X, see next section, Figure 2). The lack of functional *pncA* gene accounted for the PZA resistance observed by DST. The acquisition of an A90V mutation in *gyrA*, explains the transition to fluoroquinolones (FQs) resistance, shown in Figure 1. This mutation, which is known to confer FQ resistance, was not present in any of the isolates until May 2016. It first appeared in August 2016 and was present in the 2017 isolates, but disappeared in the genome of the last, FQs sensitive, sample taken in 2019. A

FIGURE 2
Del-X genes and coverage. On the top: a genomic map of the Del-X region according to the Mtb H37Rv reference genome. Genes in red are immunogenic. Symbols represent gene functions of interest. On the bottom: coverage of the Del-X region with short Illumina reads of sample "Mr.X 10−10"; combined coverage of all Nanopore long reads from Mr. X's samples except for "Mr.X 02−19"; coverage of the isolate that is most closely related to Mr. X's samples, according to Pathogen Detection. Dashed lines represent deletion borders, and left labels on ticks show coverage depth. BWA-MEM (21) was used to map short reads, and minimap2 v2.24-r1122 (22) was used to map long reads.

frameshift mutation in the gene *ethA*, classified by the WHO as associated with "ethionamide resistance—interim," was found in the genomes of 8 out of the 11 sequenced samples. A careful inspection revealed a significant coverage of this mutation in some reads of an additional sequenced sample. The significance of *ethA* to ethionamide (ETO) resistance is similar to that of *pncA* to PZA resistance: it alters this pro-drug to its active form. Nevertheless, this mutation only provided intermediate resistance (Figure 1). Two of the sequenced genomes did not have this mutation, but had another *ethA* mutation of uncertain significance, which was not shared by the other samples (Supplementary Table S2). In contrast to genomic prediction, no differences in ETO DST were detected among the different samples. No clofazimine (CFZ) resistance mutation was identified to explain the intermediate resistance observed in most isolates. We identified one mutation with uncertain significance regarding capreomycin (CAP) and amikacin (AMK) resistance in all samples. We were unable

to explain the differences among AMK DST results using genomic data. Genomic-based susceptibility states were not predicted for cycloserine (CYC) and clarithromycin (CLR).

Del-X

As mentioned above, WGS of the *Mtb* strain from Mr. X revealed a genomic deletion of size 15.7 kb, spanning bases 2,274,648–2,290,386 when compared to the reference *Mycobacterium tuberculosis* strain H37Rv (RefSeq id NC_000962.3, Figure 2). We have designated this deletion as Del-X. The precise boundaries of Del-X were verified using both short and long read sequencing technologies (Figure 2). The reference genome contains 18 genes in this region. These genes play important roles in the bacterial response to hypoxia, antibiotics susceptibility, interaction with the host, intracellular survival, and

dormancy. Many of them are considered immunogenic (Supplementary Table S3). One of these genes is *pncA*, the absence of which explains the pathogen's resistance to PZA, as discussed previously. Another key gene in Del-X is *hspX*, which encodes an alpha-crystallin—a small heat-shock protein, which is essential for long-term viability during latent, asymptomatic infections (23). The deletion of these 18 genes is thus expected to affect the pathogen's physiology, in particular to impose a pressure against intra-phagosome existence, as further discussed in the discussion section.
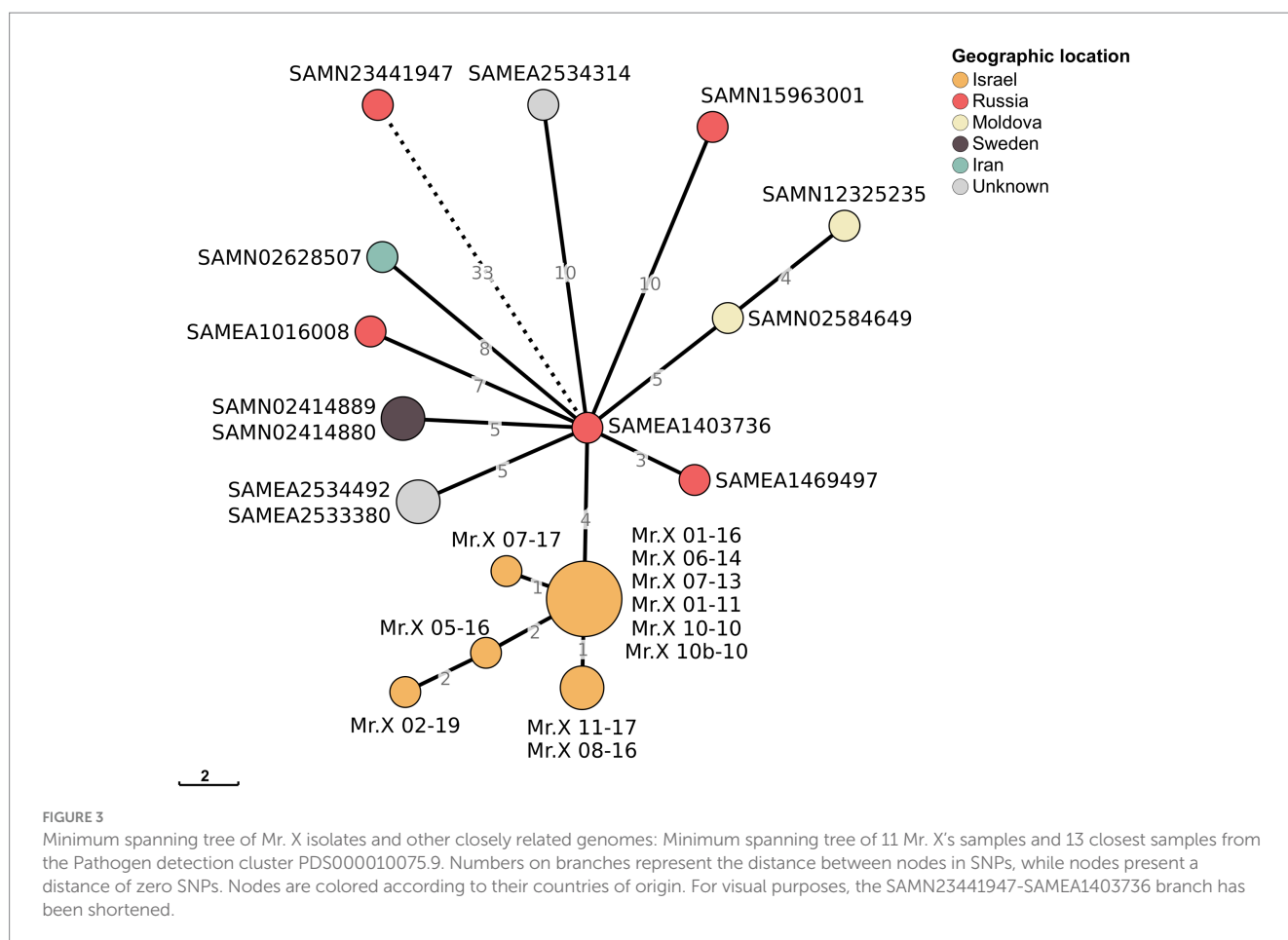
Closely related genomes

To find the origin of the *Mtb* strain that infected Mr. X, we utilized the NCBI Pathogen Detection platform (24) to identify closely related sequenced isolates. NCBI Pathogen Detection Project clusters related pathogen genome sequences in order to identify transmission networks and outbreaks. Based on genomic similarity, 13 *Mtb* isolates closely related to the strain from Mr. X were identified (Figure 3; Supplementary Table S4). These isolates were collected between 2007 and 2015 in Russia, Sweden, Moldova, and Iran. The genome most closely related to the clone from Mr. X was collected in 2009 in Russia, differing from it by only four SNPs (Figure 3). None of these genomes contains Del-X, suggesting that the deletion event might have occurred in Mr. X immediately following his infection, or in another unknown host prior to his infection. Mr. X immigrated to Israel from

Russia the same year in which he was diagnosed with active tuberculosis. Thus, it is reasonable to assume that he was infected in Russia, where a significant number of the isolates, which belong to the same cluster, were collected.

Other genomes with deletions in Del-X region

In order to estimate the uniqueness of the Del-X deletion, we searched for deletions overlapping the Del-X region in all non-redundant, well-annotated, publicly available *Mtb* genomes. Among the 7,028 RefSeq *Mtb* genomes found, 252 had deletions of various lengths overlapping Del-X (Supplementary File S5). Of these, 51 had a deletion longer than 1 kb, 33 had a deletion longer than 3 kb, and 22 had a deletion longer than 5 kb (Figure 4; Supplementary File S5). These genomes were isolated from various parts of the world, and belonged to different *Mtb* lineages, primarily the Beijing and the LAM lineages. It is important to notice that not all these 22 genomes represent unique deletion events. In fact, nine of them belong to three different clusters of nearly identical genomes. One such pair is particularly interesting: two clonal genomes with a 23.2 kb genomic deletion spanning almost the entire Del-X region, collected from a parent–child pair in Lima, Peru (25) (BioSamples SAMN10697670 and SAMN10697510, rows 4 and 5 in Figure 4). This observation suggests that despite its potential impact on dormancy, hypoxia

FIGURE 3
Minimum spanning tree of Mr. X isolates and other closely related genomes: Minimum spanning tree of 11 Mr. X's samples and 13 closest samples from the Pathogen detection cluster PDS000010075.9. Numbers on branches represent the distance between nodes in SNPs, while nodes present a distance of zero SNPs. Nodes are colored according to their countries of origin. For visual purposes, the SAMN23441947-SAMEA1403736 branch has been shortened.
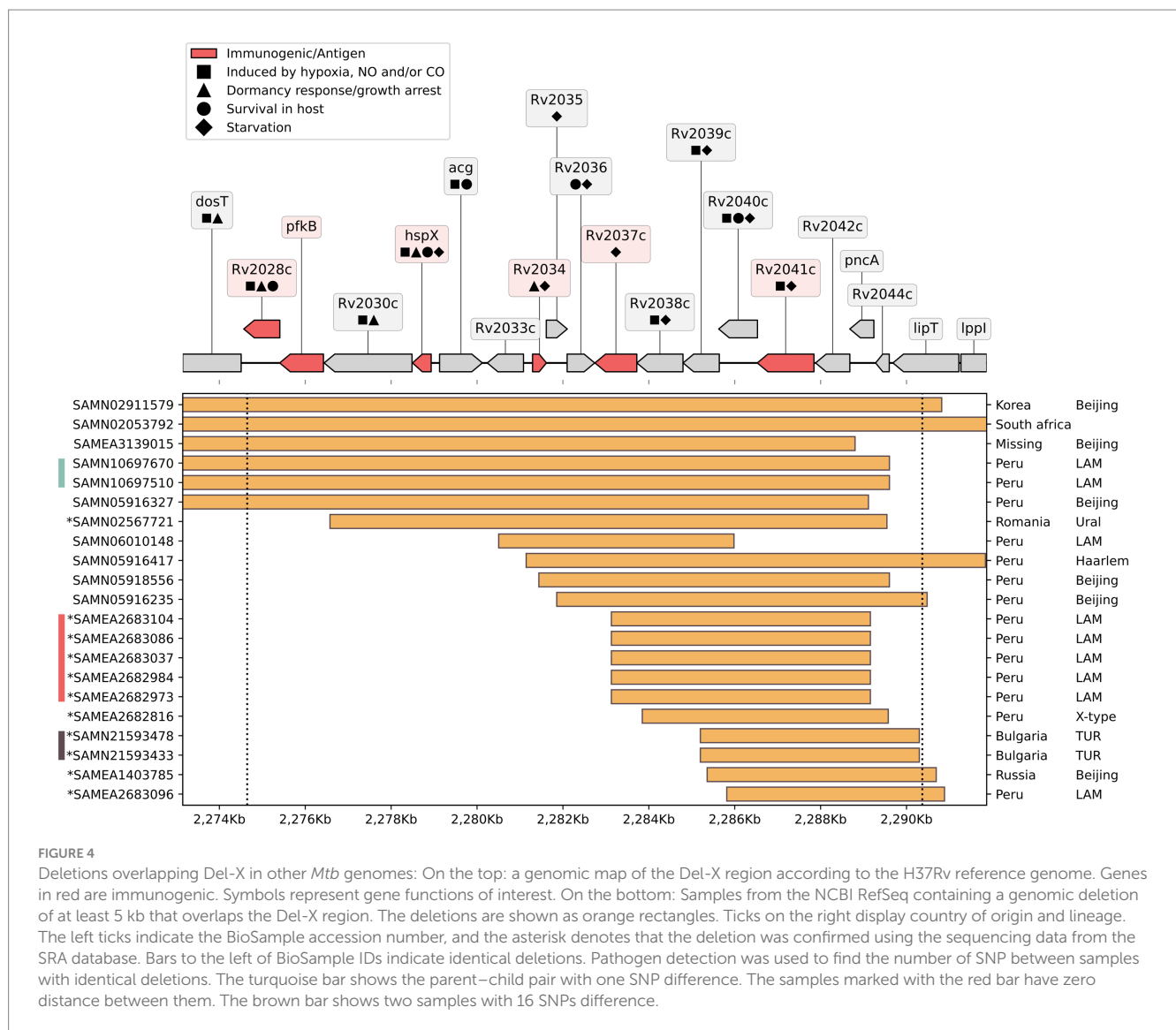
FIGURE 4

Deletions overlapping Del-X in other *Mtb* genomes: On the top: a genomic map of the Del-X region according to the H37Rv reference genome. Genes in red are immunogenic. Symbols represent gene functions of interest. On the bottom: Samples from the NCBI RefSeq containing a genomic deletion of at least 5 kb that overlaps the Del-X region. The deletions are shown as orange rectangles. Ticks on the right display country of origin and lineage. The left ticks indicate the BioSample accession number, and the asterisk denotes that the deletion was confirmed using the sequencing data from the SRA database. Bars to the left of BioSample IDs indicate identical deletions. Pathogen detection was used to find the number of SNP between samples with identical deletions. The turquoise bar shows the parent−child pair with one SNP difference. The samples marked with the red bar have zero distance between them. The brown bar shows two samples with 16 SNPs difference.

response and intracellular survival, Del-X does not prevent host-to-host transmission. This pair of clonal isolates belong to a larger cluster in NCBI Pathogen Detection (24), comprising 223 isolates (cluster PDT000456105.1). Several of these 223 isolates have smaller deletions in Del-X region, such as a 194 bp deletion inside the Rv2028c gene, a deletion of 195 bp inside the *acg* gene (Rv2032), and five clonal isolates with a deletion spanning seven genes inside Del-X, including *pncA*. However, the larger 23.2 kb genomic deletion is not likely to have emerged from any of these smaller deletions, as this region is intact in the genome, which is closest to the pair with this large deletion (BioSample SAMEA2683069). Similarly, BioSample SAMEA2682984, a strain that is clonal (0 SNPs in NCBI Pathogen Detection) to the cluster of strains with large deletions appearing in rows 12–16 of Figure 4, has no deletion in Del-X region.
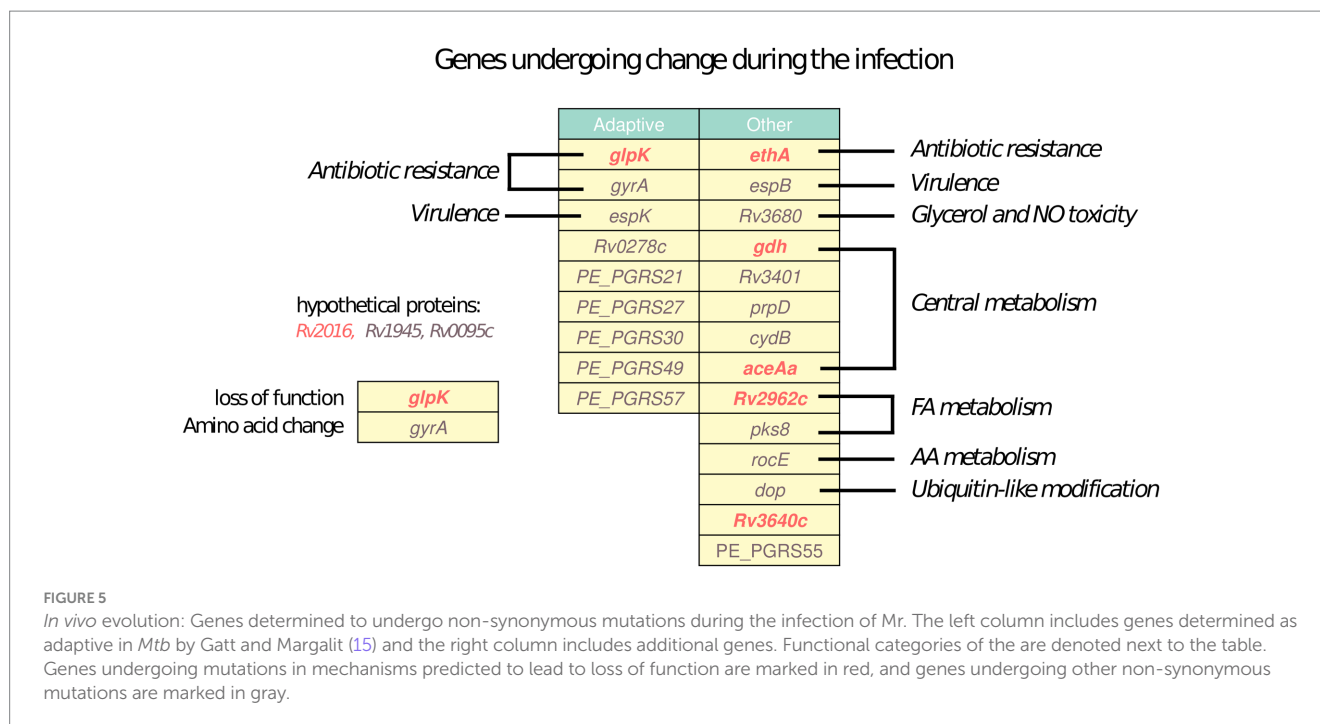
In vivo evolution

All *Mtb* sequenced isolates from Mr. X were clonal. Nevertheless, due to the large deletion, the lack of adherence to therapy, exposure to multiple drugs and the comorbidities, the *Mtb* strain infecting Mr. X

was subjected to unique selective pressures during the infection. We therefore utilized the pipeline developed by Gatt and Margalit (15), and identified 26 *Mtb* genes which underwent non-synonymous mutations during the eight and a half years of active tuberculosis (Figure 5). Nine of these genes have been previously identified by Gatt and Margalit as "adaptive genes," that tend to undergo changes at a relatively high frequency during in-host infection of *Mtb* (15), many of them belong to the PE-PGRS family. Among the 26 genes, seven underwent mutations predicted to lead to loss of function such as large deletions or frameshift mutations, and the remaining 19 genes underwent other non-synonymous mutations. The biological processes these genes are involved in are diverse, and include antibiotic resistance, virulence, fatty acid metabolism, central metabolism and more, reflecting the dynamic nature of bacterial adaptation to the *in vivo* environment.

Discussion

In this study, we present a case of exceptionally long survival in a man over 60 years of age who concurrently had progressive AIDS,

FIGURE 5
In vivo evolution: Genes determined to undergo non-synonymous mutations during the infection of Mr. The left column includes genes determined as adaptive in *Mtb* by Gatt and Margalit (15) and the right column includes additional genes. Functional categories of the are denoted next to the table. Genes undergoing mutations in mechanisms predicted to lead to loss of function are marked in red, and genes undergoing other non-synonymous mutations are marked in gray.

active tuberculosis caused by MDR-TB and alcoholism. This remarkable extended survival is particularly notable, given the patient's failure to complete his anti-TB treatment regimens, and his refusal of anti-HIV treatment as reflected by his very low CD4 cells counts and high viremic state. A previous study on survival rates in patients with drug resistance TB found a median survival of 5.9 years in MDR-TB and XDR-TB patients and 1.9 years in patients co-infected with HIV (8). The advanced age, male gender, and alcoholism of our patient makes this case even more exceptional (8). The severe consequences and poor prognosis for patients co-infected with HIV/MDR-TB, particularly those that do not take ART is well documented in the literature (26–29). Furthermore, prior HIV infection, like in our case, significantly increase the chance of acquiring MDR-TB (1.42 times higher) (30).

Over the years, since his first TB diagnosis, Mr. X started treatment several times, but never completed it. He was treated with 13 different antibiotics in total. His treatment ultimately failed to cure his TB, presumably due to the combination of his co-morbidities, the pathogen's multiple drug resistance and his low compliance (Figure 1; Supplementary Table S1). Initially the standard combination of RIF, INH, EMB, and PZA was administered, but when DST results from the NMRL indicated that the strain was resistant to RIF, INH, and PZA, these drugs were replaced with second line drugs. His treatment later included EMB again, despite most of his *Mtb* isolates exhibiting intermediate resistance to this anti-mycobacterial agent, and two isolate being fully resistant.

The isolation and genome sequencing of the *Mtb* isolates from this patient revealed an unusual genomic deletion, termed Del-X. This deletion spans from bases 2,274,648 to 2,290,386 in the H37Rv reference genome (RefSeq id NC_000962.3), causing the loss of 18 genes (Rv2028c to Rv2045c in the standard *Mtb* loci numbering). What effect could Del-X have on the bacterium's physiology, and the nature of the disease? Many of the genes inside Del-X are involved in dormancy (Figure 2; Supplementary Table S3). The *hspX* gene

(Rv2031c, also known as *acr*) encodes alpha-Crystallin-like heat shock protein (Hsp16.3), which plays a pivotal role in *Mtb* dormancy response (23, 31, 32). DosR, a transcription factor that activates the transcription of *Mtb* dormancy regulon, was shown to induce the expression of *hspX* and *acg* (33). In turn, Rv2028 and Rv2034 positively regulate the expression of the *dosR* gene (32, 34). Many of the genes within Del-X are induced by the conditions typical of macrophages phagosomes, such as hypoxia, starvation, low levels of nitric oxide, and acidic pH (32, 35–41). *hspX* actively contributes to slowing the growth of *Mtb* immediately following infection and is essential for its resistance to low oxygen stress (31, 32). Therefore, it is reasonable to speculate that the strain infecting Mr. X was unable to enter a dormant state. Studies have shown that a knockout strain of *Mtb* lacking *hspX* exhibits reduced growth in macrophages (23), and that depletion of HspX protein deteriorates *Mtb* tolerance to anaerobic conditions (32). Rv2037c too was shown to enhance *in vivo* and *ex vivo* intracellular survival (39). Thus, we suggest that this genomic deletion not only prevented the pathogen from entering a dormant state, but might have also abolished its ability to survive inside macrophages' phagosomes, imposing an escape from the phagosomes into cells' cytosol (42).

Several of the genes in Del-X region play an important role in the host immune response to *Mtb*. *hspX*, Rv2029c, Rv2034, Rv2041c, and possibly Rv2028c encode strong immunogenic antigens (43, 44) (Figure 2; Supplementary Table S3). HspX in particular plays a major role in activation of dendritic cells and induces the production of TNF-α in lymphocytes (44, 45). *acg* is important for development of the typical tuberculosis granulomas (46). In a study on MDR-TB patients, Rv2037c was shown to induce the production of antibodies and of pro-inflammatory cytokines (39). Rv2041c increases the production of cytokines, which are important for macrophage maturation (M1 polarization) (47). Therefore, Del-X is predicted to weaken the immune response against the pathogen, compared to the response against wild type *Mtb*. The patient's immune system was severely compromised by his HIV infection, as evidenced by the low

CD4 counts (about 80 per μL), low CD4/CD8 ratio (10%), and high viral load (near 350,000 and 900,000 per mL) measured in 2017 and in 2018, respectively. The combined effect of impaired immune system and lack of *Mtb* determinants important for immunogenic recognition and activation, likely diminished the host ability to neutralize the infection, resulting in a constantly active TB.

On the other hand, Del-X may have severely affected the pathogen's proliferation inside the host. Two of the deleted genes, Rv2028 and Rv2040c, have been previously identified in a comprehensive screen for genes that are important for in-host survival (48). *acg* has also been shown to be an essential virulence factor, as its deletion attenuated the bacterial growth in acute and persistent mice infections (46). The Rv2038c-Rv2041c operon encodes a sugar importer of the ABC transporter family, which is thought to play an important role in virulence or in-host survival (49), and its absence may limit the bacterial energy source. Furthermore, Del-X may impair lipid metabolism, due to the deletion of Rv2034 and Rv2037c (34, 39), and nucleotide metabolism, due to the deletion of Rv2030c (32). These deficiencies could further attenuate the growth of this strain. We hypothesize that Del-X genomic deletion resulted in an attenuated *Mtb* strain, unable to enter dormancy, yet with poor proliferation rate. The reduced virulence of this strain may have been a major factor contributing to the long survival of the patient. However, in contrast to this hypothesis, it is worth noting that some evidences have shown that deletion of *hspX* increases bacterial growth *in vivo* (31). Notably, the *Mtb* mutant used in the paper of Hu et al. (31), was deleted only in *hspX* gene, on the other hand, the del-X mutants in our study, contain an unusually large genomic deletion that covers 18 genes involved in multiple cellular functions, as described. Therefore, these strains cannot be compared. An obvious effect of Del-X was PZA resistance, due to the deletion of *pncA*. Treatment with PZA likely selected for the strain with this deletion, allowing it to take over the population, despite its defects. Indeed, although deletions in this genomic region are rare, almost all known *Mtb* genomes with a deletion of 5 kb or more in this region also exhibit a deletion of *pncA* (Figure 4; Supplementary File S5). In line with this, the work of Godfroid et al. (50), focusing on analysis of multidrug resistant MTB outbreak isolates, discovered that genomic insertions and deletions are significantly enriched in genes conferring antibiotic resistance.

The discovery of Del-X underscores the importance of detecting large genomic deletions in routine analysis of *Mtb* whole genome sequencing. Standard pipelines based on mapping to reference genomes and variant call may miss such deletions. Indeed, following this case, our laboratory has incorporated a step of identifying deletions based on uncovered genomic regions to its routine analysis.

We found a few *Mtb* genomes with genomic deletions in Del-X, from different parts of the world, mainly belonging to the Beijing or LAM lineages. As mentioned, the isolates from Mr. X also belong to the Beijing lineage that is known to be hyper-mutagenic, virulent, that rapidly acquires mutations for resistance to anti-tuberculosis drugs (51–54).

Interestingly, two of these genomes were isolated from a parent and their child, and are genetically clonal. Hence, one of these family members probably infected the other (25). This finding suggests that the Del-X genomic deletion does not prevent transmission from one patient to another and may have already been present in the strain when Mr. X was infected. However, none of this strain's known closely

related isolates had a deletion in the Del-X region. Thus, we cannot exclude the possibility that the deletion arose within this patient (Figure 3; Supplementary Table S4).

The isolates obtained from Mr. X at different time points during his nearly 9 years of infection belong to the same clone, yet there are several variations differing between them. These variations may originate from co-existing sub populations randomly sampled, or reflect changes that accumulated over time during the infection. Our phylogenetic analysis also attempts to separate these cases and include only mutations likely to have emerged during the infection. The later possibility is generally the case in *Mtb* infections (17). Many of the genes mutated during the infection were previously identified as "hot spots" for *in vivo* evolution during TB infection (Figure 5) (15, 16). For example, the *gyrA* A90V mutation discussed above, which explains the emergence of FQ resistance in isolates obtained between August 2016 and November 2017, and was probably selected for by the treatment with MFX and LFX (55). Furthermore, the fact that the patient did not receive continuous treatment is known to be a major risk factor for the appearance of resistance (5). After the exposure to these drugs was removed, an isolate taken in February 2019 lacked this mutation, probably representing a sub-population that evaded FQs and became dominant once they were no longer used. Another interesting variation is the insertion of C in the homopolymeric tract of 7 cytosine in the *glpK* gene, relative to the TB reference H37Rv genome. This mutation is common to all isolates except for the one from August 2016 and the last isolate from February 2019. The *glpK* gene encodes a glycerol kinase involved in glycerol metabolism. This mutation is known to serve as a mechanism for drug tolerance and growth slow-down, which occurs and reverses at high frequency (56). Notably, among the 26 genes mutated in-host there were six PE-PGRS genes, the functions of which are largely unknown (Figure 5).

The medical records of Mr. X indicate that he started TB treatment at least six different times, probably reflecting worsening of the symptoms (Figure 1). In all of these episodes, the patient stopped the treatment before completing it, at times by simply disappearing. This demonstrates the challenges faced by the health authorities in treating TB, which often involve social and personal difficulties. The case of Mr. X underscores the importance of patient adherence to appropriate medical treatment and highlights the significant challenges in achieving this goal due to personal and social issues. It is probable that with a combined therapy, including ART and suitable anti-TB drugs, this patient could have successfully overcome the infection with the del-X mutant. Treatment of low compliant and unmotivated patients is often unsuccessful and requires a tailored plan based on establishing a long-term trust relationship. Directly observed therapy (DOT) is an important part of such a plan (57). Mr. X's case raises complex ethical, legal, and personal dilemmas. Israeli law grants district doctors the authority to forcibly hospitalize individuals suffering from diseases that jeopardize public health, provided they have obtained judicial approval. The issue of forced medical treatment is even more intricate, and, on rare occasions, the law allows for the treatment of such patients. However, it is important to note that the successful treatment of drug-resistant tuberculosis hinges on the patient's cooperation, which, unfortunately, was unattainable in this particular case.

Taken together, we hypothesize that the unusual prolonged TB illness of Mr. X was a result of a unique intricate interplay between

the host and the pathogen. The potentially low immunogenic signature of the pathogen together with the severe immune deficiency of the host prevented the elimination of the pathogen. The Del-X genomic deletion prevented latency, resulting in constant active TB. At the same time, Del-X probably resulted in an attenuated pathogen, allowing the long-term survival of the patient.

The importance of this report stems from the unusual long survival of Mr. X, considering the circumstances, as well as the uniqueness of the genomic deletion described here. This case raises important questions about the influence of *Mtb* genetic factors on disease progression and the potential implications for prognosis and treatment. It is interesting to find out whether in additional cases of patients who survive active tuberculosis for many years without treatment, mutated *Mtb* with genomic mutations in important genes are involved. Further investigations may elucidate the precise mechanisms by which the Del-X deletion affects the pathogenicity and virulence of *Mtb*, as well as the host immune response to the infection. In addition, this special isolate, can be used as a model bacterium for a better understanding of host-pathogen fundamental interactions, such as *Mtb* escape from the phagosome to the cytosol, dormancy and more.

Materials and methods

Data classification and identification of *Mycobacterium tuberculosis complex*

All cultures processed in Israel are sent to the National Mycobacterium Reference Laboratory (NMRL), in Tel Aviv, Israel. At the NMRL, smears from samples and TB cultures are stained using Ziehl-Neelsen staining and cultured on Lowenstein-Jensen (LJ) media using standard methods (58), and in addition cultured on BACTECe MGITe 960 (BD, Sparks, MD, United States) system. The species are identified using conventional biochemical methods (59, 60) and a commercially available strip DNA probe test (Hain Lifesciences, Nehren, Germany).

Drug susceptibility testing for MTBC strains

Drug susceptibility testing for all the drugs was done using resistance ratio method (61). Pyrazinamide (PZA) was tested using Mark's stepped pH method (62).

Drugs threshold levels used to determine resistance are described in Supplementary Table S6. Importantly, for the purpose of the article, the categories Borderline and RR4 were redefined as Intermediate.

Whole genome sequencing and bioinformatics

Genomic DNA from inactivated bacterial cultures was extracted according to the manufacturer's protocol using Maxwell RSC cultured cells DNA kit and Maxwell® 16 System (Promega) (63). DNA Paired-end libraries were prepared using the Illumina Nextera XT

DNA Library Preparation Kit according to Illumina protocols. For sequencing, we utilized the Illumina MiSeq platform using a MiSeq Reagent Kit v2 (500-cycles; catalog MS-102-2003) or a MiSeq Reagent Kit v3 (600-cycle), (catalog MS-102-3003).

Raw reads fastq files were quality analyzed by FastQC (64), Kraken2 taxonomic classification (65) was used to identify mixed cultures. MTBseq, a comprehensive pipeline for whole genome sequence analysis of MTBC isolates, was used as the main tool for lineage annotation, identification of genomic variations, and clusters analysis (66). Lineages were identified as part of the MTBseq pipeline, based on the SNP scheme developed by Coll et al. (67). Genomic variations were compared to the WHO 2021 Catalog of drug resistance associated mutations in MTBC and used to identify resistant strains to antibiotics (20).

Searching for deletions among assemblies from the NCBI

A search for deletions was carried out on 7,298 *Mtb* (taxid1773) assemblies available in the NCBI RefSeq database (68) (accessed March 15 2023). After excluding replaced and anomalous assemblies, 7,028 valid assemblies remained. Assembly metadata were retrieved using the Entrez Direct,[1] and sequences were downloaded using the ncbi-genome-download v0.3.1 tool.[2] All 7,028 sequences were then aligned to the reference sequence of *Mtb* H37Rv (NC_000962.3) using the NUCmer program from the MUMmer v4.0.0rc1 package (69) with the default options. Assembly containing uncovered regions of any length that overlap with H37Rv region 2,274,648–2,290,368 was considered to have a deletion (Supplementary File S5).

Deletions verification with sequence read archive data

Samples with assembly-predicted deletions longer than 1 kb were additionally evaluated using sequencing data from the SRA database (70). 32 samples out of 51 with total deletion sizes of more than 1 kb have sequencing data in the SRA. The GRIDSS v2.13.2 software for the detection of genomic rearrangements (71, 72) was used to examine these samples. Additionally, Samtools v1.16.1 (73) was used to calculate per position read depth for all samples. The deletion is considered to be correctly verified if: (1) The deletion predicted by GRIDSS overlaps with the deletion predicted from the assembly by more than 99% of the length and vice versa; (2) More than 90% of the bases in the given region have a read depth of less than 5% of the mean genome depth. Assembly GCF_000679915.1 had four SRA runs, three of them passed, therefore we considered this assembly as verified. After applying these filters, seven samples were discarded. Samples that lack sequencing data in the SRA database were used in the further analysis without verification (Supplementary File S5).

---

1 https://www.ncbi.nlm.nih.gov/books/NBK179288/

2 https://github.com/kblin/ncbi-genome-download

Long reads

For Oxford Nanopore sequencing, DNA from Mr. X 10 isolates (except 02–19) were converted into a sequencing library using Rapid Barcoding Kit 96 (SQK-RBK110.96) chemistry according to the manufacturer's protocol. Sequencing was performed on a MinION sequencer with an R9.4.1 (FLO-MIN106D) flow cell, and basecalling was carried out by Guppy version 6.0.1. Sequences of all 11 isolates were used to create a single assembly.

MST construction

13 samples from the Pathogen detection cluster PDS000010075.9 and 11 Mr. X's isolates were used to construct the minimum spanning tree. There are three samples from Pathogen detection (BioSample accession numbers SAMN23441947, SAMN15963001, and SAMN12325235) for which there are no sequencing data in the SRA database (70). Assembled sequences of these samples were artificially converted to paired-end reads using a custom python script. The forward and reverse read lengths were both set to 150 bp with a minimal coverage of 50. Prefetch and fasterq-dump from the SRA Toolkit v3.0.1[3] were used to obtain FASTQ files from the remaining 10 samples. BWA-MEM v0.7.17-r1188 (21) was used to map the reads to the *Mtb* H37Rv (NC_000962.3) reference sequence. Viewing, sorting, and duplication removal in BAM files were done using Samtools v1.16.1 (73). FreeBayes v1.3.6 (74) was used to call SNPs with the $-p$ $1-q$ three parameters. SNPs were filtered using the following criteria: Site quality ≥ 30, variant read depth ≥ 5, distance to nearest SNP ≥ 3 bp, $\geq 5\%$ read mapping up-or downstream, $\geq 5\%$ read mapping to least-covered strand, variant allele frequency $\geq 90\%$. In addition, SNPs in genes associated with antibiotic resistance as well as those located in PE/PPE were excluded in the subsequent analysis (75). Sites with more than 5% of missing variants were filtered out as the final stage of filtering. SNP filtering was carried out using VCFtools v 0.1.16 (76), bcftools v1.16 (73), vcffilter from vcflib v 1.0.3 (77), and bedtools v2.30.0 (78). The GrapeTree v 1.5.0 (79) with MSTree method was used to create the minimum spanning tree.

Lineage assignment

Genome assemblies were mapped to the reference *Mtb* H37Rv genome (RefSeq id NC_000962.3), and variants were detected as previously described (80). These variants were compared to a SNP barcode of *Mtb* lineages to assign a lineage (67).

Determination of within-host evolution

Detection of mutations was performed as in Gatt and Margalit (15). Briefly, assemblies were constructed for all paired-end libraries using kSNP v.3.1 (81, 82), and phylogenetic structure was determined using the TRACE algorithm with high confidence progenitor-progeny isolate pairs ascertained from the different isolates of Mr. X. For each progenitor-progeny pair, *breseq* v.0.32 (83, 84) was then used to

determine differential variations from the reference genome *Mtb* H37Rv (RefSeq id NC_000962.3) appearing in the progeny isolate and not in the progenitor isolate. SnpEff v.4.1 (85) was used to predict the effect of the different mutations, with high impact mutations predicted to lead to loss of function.

Data availability statement

The raw genomic sequences from this work have been deposited to NCBI and are available in BioProject PRJNA834625.

Ethics statement

This study was approved by the National Helsinki Committee for Medical Experiments on Humans, of the Israeli Ministry of Health (# MOH-081-2021). The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from a by-product of routine care or industry. Written informed consent for participation was not required from the participant or the participant's legal guardians/next of kin in accordance with the national legislation and institutional requirements. Written informed consent was not obtained from the individual for the publication of any potentially identifiable images or data included in this article because the patient this article is about died several years before publication. Therefore, a written informed consent was obtained from the participant's next of kin, for the publication of this case report, and any potentially identifiable images or data included in this article.

---

3 https://github.com/ncbi/sra-tools/

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2023.1292665/full#supplementary-material

References

1. World Health Organization (2022). Global tuberculosis report 2022. Licence: CC BY-NC-SA 3.0 IGO. 2022. Geneva: World Health Organization.

2. World Health Organization. *WHO Operational Handbook on Tuberculosis. Module 4: Treatment—Drug-Resistant Tuberculosis Treatment, 2022 update*. Geneva: World Health Organization (2022).

3. Sotgiu G, Sulis G, Matteelli A. Tuberculosis-a World Health Organization perspective. *Microbiol Spectr*. (2017) 5. doi: 10.1128/microbiolspec.TNMI7-0036-2016

4. Linh NN, Viney K, Gegia M, Falzon D, Glaziou P, Floyd K, et al. World Health Organization treatment outcome definitions for tuberculosis: 2021 update. *Eur Respir J*. (2021) 58:2100804. doi: 10.1183/13993003.00804-2021

5. Gillespie SH. Evolution of drug resistance in *Mycobacterium tuberculosis*: clinical and molecular perspective. *Antimicrob Agents Chemother*. (2002) 46:267–74. doi: 10.1128/AAC.46.2.267-274.2002

6. Bell LCK, Noursadeghi M. Pathogenesis of HIV-1 and *Mycobacterium tuberculosis* co-infection. *Nat Rev Microbiol*. (2018) 16:80–90. doi: 10.1038/nrmicro.2017.128

7. Nordholm AC, Andersen AB, Wejse C, Norman A, Ekstrom CT, Andersen PH, et al. Mortality, risk factors, and causes of death among people with tuberculosis in Denmark, 1990-2018. *Int J Infect Dis*. (2023) 130:76–82. doi: 10.1016/j.ijid.2023.02.024

8. Balabanova Y, Ignatyeva O, Fiebig L, Riekstina V, Danilovits M, Jaama K, et al. Survival of patients with multidrug-resistant TB in Eastern Europe: what makes a difference? *Thorax*. (2016) 71:854–61. doi: 10.1136/thoraxjnl-2015-207638

9. Wigger GW, Bouton TC, Jacobson KR, Auld SC, Yeligar SM, Staitieh BS. The impact of alcohol use disorder on tuberculosis: a review of the epidemiology and potential immunologic mechanisms. *Front Immunol*. (2022) 13:864817. doi: 10.3389/fimmu.2022.864817

10. Singh PR, Vijjamarri AK, Sarkar D. Metabolic switching of *Mycobacterium tuberculosis* during hypoxia is controlled by the virulence regulator PhoP. *J Bacteriol*. (2020) 202. doi: 10.1128/JB.00705-19

11. Leistikow RL, Morton RA, Bartek IL, Frimpong I, Wagner K, Voskuil MI. The *Mycobacterium tuberculosis* DosR regulon assists in metabolic homeostasis and enables rapid recovery from nonrespiring dormancy. *J Bacteriol*. (2010) 192:1662–70. doi: 10.1128/JB.00926-09

12. Peddireddy V, Doddam SN, Ahmed N. Mycobacterial dormancy systems and host responses in tuberculosis. *Front Immunol*. (2017) 8:84. doi: 10.3389/fimmu.2017.00084

13. Delogu G, Sali M, Fadda G. The biology of *mycobacterium tuberculosis* infection. *Mediterr J Hematol Infect Dis*. (2013) 5:e2013070. doi: 10.4084/MJHID.2013.070

14. Verma A, Ghoshal A, Dwivedi VP, Bhaskar A. Tuberculosis: the success tale of less explored dormant *Mycobacterium tuberculosis*. *Front Cell Infect Microbiol*. (2022) 12:1079569. doi: 10.3389/fcimb.2022.1079569

15. Gatt YE, Margalit H. Common adaptive strategies underlie within-host evolution of bacterial pathogens. *Mol Biol Evol*. (2021) 38:1101–21. doi: 10.1093/molbev/msaa278

16. Ley SD, de Vos M, Van Rie A, Warren RM. Deciphering within-host microevolution of *Mycobacterium tuberculosis* through whole-genome sequencing: the phenotypic impact and way forward. *Microbiol Mol Biol Rev*. (2019) 83. doi: 10.1128/MMBR.00062-18

17. Gagneux S. Ecology and evolution of *Mycobacterium tuberculosis*. *Nat Rev Microbiol*. (2018) 16:202–13. doi: 10.1038/nrmicro.2018.8

18. Dos Vultos T, Mestre O, Rauzier J, Golec M, Rastogi N, Rasolofo V, et al. Evolution and diversity of clonal bacteria: the paradigm of *Mycobacterium tuberculosis*. *PLoS One*. (2008) 3:e1538. doi: 10.1371/journal.pone.0001538

19. Baena A, Cabarcas F, Ocampo JC, Barrera LF, Alzate JF. Large genomic deletions delineate *Mycobacterium tuberculosis* L4 sublineages in south American countries. *PLoS One*. (2023) 18:e0285417. doi: 10.1371/journal.pone.0285417

20. World Health Organization (2013). Catalogue of mutations in *Mycobacterium tuberculosis* complex and their association with drug resistance. Licence: CC BY-NC-SA 3.0 IGO. Geneva: World Health Organization.

21. Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [Preprint]. doi: 10.48550/arXiv.1303.3997

22. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. (2018) 34:3094–100. doi: 10.1093/bioinformatics/bty191

23. Yuan Y, Crane DD, Simpson RM, Zhu YQ, Hickey MJ, Sherman DR, et al. The 16-kDa alpha-crystallin (Acr) protein of *Mycobacterium tuberculosis* is required for growth in macrophages. *Proc Natl Acad Sci U S A*. (1998) 95:9578–83. doi: 10.1073/pnas.95.16.9578

24. NCBI (2016). The NCBI pathogen detection project [internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. Available at: https://www.ncbi.nlm.nih.gov/pathogens/

25. Dixit A, Freschi L, Vargas R, Calderon R, Sacchettini J, Drobniewski F, et al. Whole genome sequencing identifies bacterial factors affecting transmission of multidrug-resistant tuberculosis in a high-prevalence setting. *Sci Rep*. (2019) 9:5602. doi: 10.1038/s41598-019-41967-8

26. Bayowa JR, Kalyango JN, Baluku JB, Katuramu R, Ssendikwanawa E, Zalwango JF, et al. Mortality rate and associated factors among patients co-infected with drug resistant tuberculosis/HIV at Mulago National Referral Hospital, Uganda, a retrospective cohort study. *PLOS Glob Public Health*. (2023) 3:e0001020. doi: 10.1371/journal.pgph.0001020

27. Chem ED, Van Hout MC, Hope V. Treatment outcomes and antiretroviral uptake in multidrug-resistant tuberculosis and HIV co-infected patients in sub Saharan Africa: a systematic review and meta-analysis. *BMC Infect Dis*. (2019) 19:723. doi: 10.1186/s12879-019-4317-4

28. Isaakidis P, Casas EC, Das M, Tseretopoulou X, Ntzani EE, Ford N. Treatment outcomes for HIV and MDR-TB co-infected adults and children: systematic review and meta-analysis. *Int J Tuberc Lung Dis*. (2015) 19:969–78. doi: 10.5588/ijtld.15.0123

29. Lelisho ME, Wotale TW, Tareke SA, Alemu BD, Hassen SS, Yemane DM, et al. Survival rate and predictors of mortality among TB/HIV co-infected adult patients: retrospective cohort study. *Sci Rep*. (2022) 12:18360. doi: 10.1038/s41598-022-23316-4

30. Sultana ZZ, Hoque FU, Beyene J, Akhlak-Ul-Islam M, Khan MHR, Ahmed S, et al. HIV infection and multidrug resistant tuberculosis: a systematic review and meta-analysis. *BMC Infect Dis*. (2021) 21:51. doi: 10.1186/s12879-020-05749-2

31. Hu Y, Movahedzadeh F, Stoker NG, Coates AR. Deletion of the *Mycobacterium tuberculosis* alpha-crystallin-like hspX gene causes increased bacterial growth in vivo. *Infect Immun*. (2006) 74:861–8. doi: 10.1128/IAI.74.2.861-868.2006

32. Mushtaq K, Sheikh JA, Amir M, Khan N, Singh B, Agrewala JN. Rv2031c of *Mycobacterium tuberculosis*: a master regulator of Rv2028-Rv2031 (HspX) operon. *Front Microbiol*. (2015) 6:351. doi: 10.3389/fmicb.2015.00351

33. Park HD, Guinn KM, Harrell MI, Liao R, Voskuil MI, Tompa M, et al. Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. *Mol Microbiol*. (2003) 48:833–43. doi: 10.1046/j.1365-2958.2003.03474.x

34. Gao CH, Yang M, He ZG. Characterization of a novel ArsR-like regulator encoded by Rv2034 in *Mycobacterium tuberculosis*. *PLoS One*. (2012) 7:e36255. doi: 10.1371/journal.pone.0036255

35. Betts JC, Lukey PT, Robb LC, McAdam RA, Duncan K. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol Microbiol*. (2002) 43:717–31. doi: 10.1046/j.1365-2958.2002.02779.x

36. Barreto C, de Oliveira M, Balan A. The ATP-binding cassette (ABC) transport Systems in *Mycobacterium tuberculosis*: structure, function, and possible targets for therapeutics. *Biology*. (2020) 9:443. doi: 10.3390/biology9120443

37. Garbe TR, Hibler NS, Deretic V. Response to reactive nitrogen intermediates in *Mycobacterium tuberculosis*: induction of the 16-kilodalton alpha-crystallin homolog by exposure to nitric oxide donors. *Infect Immun*. (1999) 67:460–5. doi: 10.1128/IAI.67.1.460-465.1999

38. Kim SY, Lee BS, Shin SJ, Kim HJ, Park JK. Differentially expressed genes in *Mycobacterium tuberculosis* H37Rv under mild acidic and hypoxic conditions. *J Med Microbiol*. (2008) 57:1473–80. doi: 10.1099/jmm.0.2008/001545-0

39. Kumari B, Saini V, Kaur J, Kaur J. Rv2037c, a stress induced conserved hypothetical protein of *Mycobacterium tuberculosis*, is a phospholipase: role in cell wall modulation and intracellular survival. *Int J Biol Macromol*. (2020) 153:817–35. doi: 10.1016/j.ijbiomac.2020.03.037

40. Saxena A, Srivastava V, Srivastava R, Srivastava BS. Identification of genes of *Mycobacterium tuberculosis* upregulated during anaerobic persistence by fluorescence and kanamycin resistance selection. *Tuberculosis*. (2008) 88:518–25. doi: 10.1016/j.tube.2008.01.003

41. Snasel J, Machova I, Solinova V, Kasicka V, Krecmerova M, Pichova I. Phosphofructokinases a and B from *Mycobacterium tuberculosis* display different catalytic properties and allosteric regulation. *Int J Mol Sci*. (2021) 22. doi: 10.3390/ijms22031483

42. Jamwal SV, Mehrotra P, Singh A, Siddiqui Z, Basu A, Rao KV. Mycobacterial escape from macrophage phagosomes to the cytoplasm represents an alternate adaptation mechanism. *Sci Rep*. (2016) 6:23089. doi: 10.1038/srep23089

43. Legesse M, Ameni G, Medhin G, Mamo G, Franken KL, Ottenhoff TH, et al. IgA response to ESAT-6/CFP-10 and Rv2031 antigens varies in patients with culture-confirmed pulmonary tuberculosis, healthy *Mycobacterium tuberculosis*-infected and non-infected individuals in a tuberculosis endemic setting. *Ethiopia Scand J Immunol*. (2013) 78:266–74. doi: 10.1111/sji.12080

44. Meier NR, Battegay M, Ottenhoff THM, Furrer H, Nemeth J, Ritz N. HIV-infected patients developing tuberculosis disease show early changes in the immune response to novel *Mycobacterium tuberculosis* antigens. *Front Immunol*. (2021) 12:620622. doi: 10.3389/fimmu.2021.620622

45. Amir M, Aqdas M, Nadeem S, Siddiqui KF, Khan N, Sheikh JA, et al. Diametric role of the latency-associated protein Acr1 of *Mycobacterium tuberculosis* in modulating the functionality of pre-and post-maturational stages of dendritic cells. *Front Immunol*. (2017) 8:624. doi: 10.3389/fimmu.2017.00624

46. Hu Y, Coates AR. *Mycobacterium tuberculosis* acg gene is required for growth and virulence in vivo. *PLoS One*. (2011) 6:e20958. doi: 10.1371/journal.pone.0020958

47. Kim S-Y. A-rum Shin, Byung-Soo Lee, Hwa-Jung Kim2 Bo young Jeon, sang-Nae Cho, Jeong-Kyu Park, sung Jae Shin characterization of immune responses to *Mycobacterium tuberculosis* Rv2041c protein. *J Bacteriol Virol*. (2009) 39:183–93. doi: 10.4167/jbv.2009.39.3.183

48. Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol*. (2003) 48:77–84. doi: 10.1046/j.1365-2958.2003.03425.x

49. De la Torre LI, Vergara Meza JG, Cabarca S, Costa-Martins AG, Balan A. Comparison of carbohydrate ABC importers from *Mycobacterium tuberculosis*. *BMC Genomics*. (2021) 22:841. doi: 10.1186/s12864-021-07972-w

50. Godfroid M, Dagan T, Merker M, Kohl TA, Diel R, Maurer FP, et al. Insertion and deletion evolution reflects antibiotics selection pressure in a *Mycobacterium tuberculosis* outbreak. *PLoS Pathog*. (2020) 16:e1008357. doi: 10.1371/journal.ppat.1008357

51. Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, et al. *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat Genet*. (2013) 45:784–90. doi: 10.1038/ng.2656

52. Hakamata M, Takihara H, Iwamoto T, Tamaru A, Hashimoto A, Tanaka T, et al. Higher genome mutation rates of Beijing lineage of *Mycobacterium tuberculosis* during human infection. *Sci Rep*. (2020) 10:17997. doi: 10.1038/s41598-020-75028-2

53. Liu Q, Wang D, Martinez L, Lu P, Zhu L, Lu W, et al. *Mycobacterium tuberculosis* Beijing genotype strains and unfavourable treatment outcomes: a systematic review and meta-analysis. *Clin Microbiol Infect*. (2020) 26:180–8. doi: 10.1016/j.cmi.2019.07.016

54. Ribeiro SC, Gomes LL, Amaral EP, Andrade MR, Almeida FM, Rezende AL, et al. *Mycobacterium tuberculosis* strains of the modern sublineage of the Beijing family are more likely to display increased virulence than strains of the ancient sublineage. *J Clin Microbiol*. (2014) 52:2615–24. doi: 10.1128/JCM.00498-14

55. Singh R, Dwivedi SP, Gaharwar US, Meena R, Rajamani P, Prasad T. Recent updates on drug resistance in *Mycobacterium tuberculosis*. *J Appl Microbiol*. (2020) 128:1547–67. doi: 10.1111/jam.14478

56. Safi H, Gopal P, Lingaraju S, Ma S, Levine C, Dartois V, et al. Phase variation in *Mycobacterium tuberculosis* glpK produces transiently heritable drug tolerance. *Proc Natl Acad Sci U S A*. (2019) 116:19665–74. doi: 10.1073/pnas.1907631116

57. World Health Organization. Communicable Diseases Cluster (1999). What is DOTS?: A guide to understanding the WHO-recommended TB control strategy known as DOTS. World Health Organization. Available at: https://apps.who.int/iris/handle/10665/65979

58. Kent PT, Kubica GP. *Public Health Mycobacteriology: A Guide for the Level III Laboratory*. Atlanta: Centers for Disease Control (1985).

59. Marks J. A system for the examination of tubercle bacilli and other mycobacteria. *Tubercle*. (1976) 57:207–25. doi: 10.1016/0041-3879(76)90030-1

60. Collins C. H., Grange J. M., Yates M.D. (1985). Organization and Practice in Tuberculosis Bacteriology.

61. European Centre for Disease Prevention and Control. *Handbook on Tuberculosis Laboratory Diagnostic Methods in the European Union—Updated 2018*. Stockholm: ECDC (2018).

62. Marks J. A 'Stepped Ph' technique for the estimation of pyrazinamide sensitivity. *Tubercle*. (1964) 45:47–50. doi: 10.1016/s0041-3879(64)80087-8

63. Cabibbe AM, Walker TM, Niemann S, Cirillo DM. Whole genome sequencing of *Mycobacterium tuberculosis*. *Eur Respir J*. (2018) 52:1801163. doi: 10.1183/13993003.01163-2018

64. Andrews S. (2010). FASTQC. A quality control tool for high throughput sequence data.

65. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with kraken 2. *Genome Biol*. (2019) 20:257. doi: 10.1186/s13059-019-1891-0

66. Kohl TA, Utpatel C, Schleusener V, De Filippo MR, Beckert P, Cirillo DM, et al. MTBseq: a comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex isolates. *PeerJ*. (2018) 6:e5895. doi: 10.7717/peerj.5895

67. Coll F, McNerney R, Guerra-Assuncao JA, Glynn JR, Perdigao J, Viveiros M, et al. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun*. (2014) 5:4812. doi: 10.1038/ncomms5812

68. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. (2016) 44:D733–45. doi: 10.1093/nar/gkv1189

69. Marcais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput Biol*. (2018) 14:e1005944. doi: 10.1371/journal.pcbi.1005944

70. Leinonen R, Sugawara H, Shumway M. International nucleotide sequence database, collaboration the sequence read archive. *Nucleic Acids Res*. (2011) 39:D19–21. doi: 10.1093/nar/gkq1019

71. Cameron DL, Baber J, Shale C, Valle-Inclan JE, Besselink N, van Hoeck A, et al. GRIDSS2: comprehensive characterisation of somatic structural variation using single breakend variants and structural variant phasing. *Genome Biol*. (2021) 22:202. doi: 10.1186/s13059-021-02423-x

72. Cameron DL, Schroder J, Penington JS, Do H, Molania R, Dobrovic A, et al. GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res*. (2017) 27:2050–60. doi: 10.1101/gr.222109.117

73. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. (2021) 10. doi: 10.1093/gigascience/giab008

74. Garrison E., Marth G. T. (2012). Haplotype-based variant detection from short-read sequencing. arXiv [Preprint]. doi: 10.48550/arXiv.1207.3907

75. Kohl TA, Diel R, Harmsen D, Rothganger J, Walter KM, Merker M, et al. Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. *J Clin Microbiol*. (2014) 52:2479–86. doi: 10.1128/JCM.00567-14

76. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. Genomes project analysis, group the variant call format and VCFtools. *Bioinformatics*. (2011) 27:2156–8. doi: 10.1093/bioinformatics/btr330

77. Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P. A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-VCF, cyvcf2, hts-nim and slivar. *PLoS Comput Biol*. (2022) 18:e1009123. doi: 10.1371/journal.pcbi.1009123

78. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. (2010) 26:841–2. doi: 10.1093/bioinformatics/btq033

79. Zhou Z, Alikhan NF, Sergeant MJ, Luhmann N, Vaz C, Francisco AP, et al. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res*. (2018) 28:1395–404. doi: 10.1101/gr.232397.117

80. Rubinstein M, Grossman R, Nissan I, Schwaber MJ, Carmeli Y, Kaidar-Shwartz H, et al. *Mycobacterium intracellulare* subsp. chimaera from cardio surgery heating-cooling units and from clinical samples in Israel are genetically unrelated. *Pathogens*. (2021) 10:1392. doi: 10.3390/pathogens10111392

81. Gardner SN, Hall BG. When whole-genome alignments just won't work: kSNP v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes. *PLoS One*. (2013) 8:e81760. doi: 10.1371/journal.pone.0081760

82. Gardner SN, Slezak T, Hall BG. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics*. (2015) 31:2877–8. doi: 10.1093/bioinformatics/btv271

83. Barrick JE, Colburn G, Deatherage DE, Traverse CC, Strand MD, Borges JJ, et al. Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. *BMC Genomics*. (2014) 15:1039. doi: 10.1186/1471-2164-15-1039

84. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol*. (2014) 1151:165–88. doi: 10.1007/978-1-4939-0554-6_12

85. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. (2012) 6:80–92. doi: 10.4161/fly.19695