Genetic diversity, evolution and drug resistance of *Mycobacterium tuberculosis* lineage 2

Sabina Atavliyeva*, Dana Auganova and Pavel Tarlykov*

Genomics and Proteomics Core Facility, National Center for Biotechnology, Astana, Kazakhstan

Mycobacterium tuberculosis causes a chronic infectious disease called tuberculosis. Phylogenetic lineage 2 (L2) of *M. tuberculosis*, also known as the East Asian lineage, is associated with high virulence, increased transmissibility, and the spread of multidrug-resistant strains. This review article examines the genomic characteristics of the *M. tuberculosis* genome and *M. tuberculosis* lineage 2, such as the unique insertion sequence and spoligotype patterns, as well as MIRU-VNTR typing, and SNP-based barcoding. The review describes the geographical distribution of lineage 2 and its history of origin. In addition, the article discusses recent studies on drug resistance and compensatory mechanisms of *M. tuberculosis* lineage 2 and its impact on the pathogen's transmissibility and virulence. This review article discusses the importance of establishing a unified classification for lineage 2 to ensure consistency in terminology and criteria across different studies and settings.

KEYWORDS

Mycobacterium tuberculosis, phylogeny, tuberculosis, multidrug-resistant, virulence, disease outbreaks

1 Introduction

Mycobacterium tuberculosis (MTB) poses a serious threat to human health, with 10 million people becoming infected with tuberculosis every year. Until the coronavirus (COVID-19) pandemic, TB was the leading cause of death from a single infectious agent, ranking above HIV/AIDS. In 2022, according to the World Health Organization (WHO), tuberculosis remains the second leading cause of death from infectious diseases after COVID-19 and caused almost twice as many deaths as HIV/AIDS (World Health Organization, 2023). Patients with comorbidities such as TB and HIV/AIDS are at significantly higher risk of developing severe forms of COVID-19 (Udoakang et al., 2023). Co-infection with COVID-19 and tuberculosis results in higher mortality than in patients with COVID-19 alone (du Bruyn et al., 2023). HIV also increases the risk of latent tuberculosis progressing to its active form by 26 times (Yang et al., 2022). Thus, co-infection provides synergistic effects to both pathogens, resulting in atypical clinical manifestations of tuberculosis.

Among MTB genetic lineages, lineage 2 (L2), also known as East Asian lineage, stands out as one of the most widely found lineages. This lineage, mainly represented by the Beijing family, plays a significant role in the global epidemiology of tuberculosis, accounting for nearly a quarter of all reported TB cases (Liu et al., 2020). Experimental and clinical data indicate high virulence and increased mutation rates in Beijing strains, highlighting their clinical and epidemiological significance (Guerra-Assunção et al., 2015; Rajwani et al., 2018). Understanding the genetic and phenotypic features of L2, characterized by highly transmissible strains and multidrug resistance, is critical for developing effective strategies

for the treatment and control of tuberculosis (Holt et al., 2018; Liu et al., 2018a).

The *Mycobacterium tuberculosis* Beijing genotype was first identified by Van Soolingen et al. (1995) in the mid-1990s in China. The strains differed in the pattern of the IS*6110* insertion sequence with 15–20 copies in the genome and were also distinguished by a peculiar spoligotype, where only spacers from 35 to 43 in the DR locus predominated (Van Soolingen et al., 1995). Strains of this lineage may have originated in East Asia in its northern or southern parts and then spread through China to the rest of the world. It subsequently attracted attention for its ability to cause large epidemics in various parts of the world, especially in Central Asia, Eastern Europe, and Eastern Africa (Pardini et al., 2009; Rutaihwa et al., 2019; Daniyarov et al., 2021). In a short time, both multidrug resistance (MDR) and extensive drug resistance (XDR) were reported to be developed. It is known that mycobacteria have mechanisms of resistance to all anti-tuberculosis drugs that exist today (World Health Organization, 2021).

This article covers current research on the genetic characteristics of MTB lineage 2, highlighting its implications for understanding the epidemiology and diagnosis of tuberculosis. Next, the genomic diversity of lineage 2 is described concerning the phylogeny and classification of sublineages and clusters. In addition, the distribution of these sublineages across geographic regions, with a description of the place and time of origin of lineage 2 is covered. Finally, the latest knowledge on drug resistance of *M. tuberculosis* and the impact of compensatory mutations on the development of high transmissibility and virulence is discussed. This review article aims to provide a state-of-the-art overview of the genomic characterization of lineage 2 strains, identifying key trends, challenges, and prospects for future research.

2 Genomic characteristics of *Mycobacterium tuberculosis* lineage 2

Mycobacterium tuberculosis has several distinctive genetic features compared to other bacterial species, such as the extremely rare events of horizontal gene transfer (HGT), which leads to a clonal and hierarchical population structure (Boritsch et al., 2016). The underlying mechanism of rare HGT events in MTB is still under discussion. However, features of the genome such as the similarity of phylogenetic trees constructed by different genetic markers, as well as the constantly high G + C content in the genome and the rare occurrence of homoplasy, all indicate that MTBC has partially lost the ability for ongoing genetic recombination (Saelens et al., 2019). Given this structure, the evolution of *M. tuberculosis* depends on mutations, such as substitutions, deletions, and duplications of single nucleotides (Karmakar et al., 2019). Analysis of genotypes can provide unique information on the pathogen's distribution dynamics and evolutionary genetics. Assessing the genetic distances between strains is used to complement epidemiological transmission data.

2.1 Genotyping methods

Genetic typing for molecular identification of clinical strains of *M. tuberculosis* has been widely used in recent decades. This molecular identification raised questions about strain-specific differences in clinical presentation and epidemiological characteristics of the infection. All MTBC genotyping methods can be classified based on repetitive sequences (IS*6110*, spoligotyping, and MIRU-VNTR), large sequence polymorphism (LSP) typing, and whole genome sequencing (WGS) for SNP-based analysis (Shevtsov et al., 2013; Jagielski et al., 2014; Żukowska et al., 2023). A schematic representation and timeline of *M. tuberculosis* marker types and genotyping methods are shown in Figures 1, 2.

2.1.1 IS*6110*

The insertion sequence IS*6110* is a hallmark of the mycobacterial genome and became the first gold standard for strain genotyping. The insert is a fragment 1,361 bp long inserted into various parts of the genome and surrounded by inverted repeats 28 bp long. The IS*6110* RFLP detection assay itself is based on differences in copy number and identification of the insertion site in the genome. The introduction of IS*6110* into the genome is a random event that creates differences between tuberculosis strains both within and between families. The Euro-Asian lineage is characterized by a high copy number of this insertion, which may be associated with its influence on the evolution and virulence of strains of this lineage (Alonso et al., 2013).

Strains of the Beijing family showed similar patterns and were considered a homogeneous family based on IS*6110* profiles. Initially, two insertions were found: at the *dnaA-dnaN* origin of replication of the *M. tuberculosis* chromosome and one or two IS*6110* insertions (two for the W strain) in the NTF region (Bifani et al., 2002). Currently, IS*6110*-RFLP typing is not as common as before, however, IS*6110*-RFLP profiles have been accumulated in various databases (RIVM in the Netherlands), or PHRI in the USA and can be used to track strains in historical collections, which is compatible with more modern methods such as WGS (Mokrousov et al., 2021). However, the specific insertion site of this element in the Rv*2180c* gene is used to identify the Beijing family using PCR among clinical strains (Pérez-Lago et al., 2019). Despite its successful application in studying the epidemiology and spread of tuberculosis, this method has limitations, such as the requirement of large quantities of DNA with high concentration and purity. It is worth adding that the method is difficult to reproduce in different laboratories, as it is technically complex and does not discriminate sufficiently for isolates containing less than five copies of the IS*6110* element or none at all (Mestre et al., 2011; Coscolla and Gagneux, 2014).

2.1.2 Spoligotyping

Spoligotyping is a method for typing the DR region, which is structurally composed of 43 spacers, each of which is surrounded by direct repeats. Identification is based on the absence or presence of a particular spacer sequence (Coscolla and Gagneux, 2014). The spoligoprofile of the L2 family lacks spacers 1 to 34, and at least three spacers 35 to 43 are observed (Shitikov et al., 2017). Spoligotyping is a cost-effective method for mycobacterial strains but has limited ability to discriminate sublineages of the L2 family. According to the SITVITWEB database, the spoligotypes of the L2 family are divided into SIT523 (777777777777771) known to be characteristic of proto-Beijing, and SIT1 (000000000003771), attributed to the Beijing family (Demay et al., 2012). However, this method is not always suitable for the other genetic families of MTB, such as the LAM family of Lineage
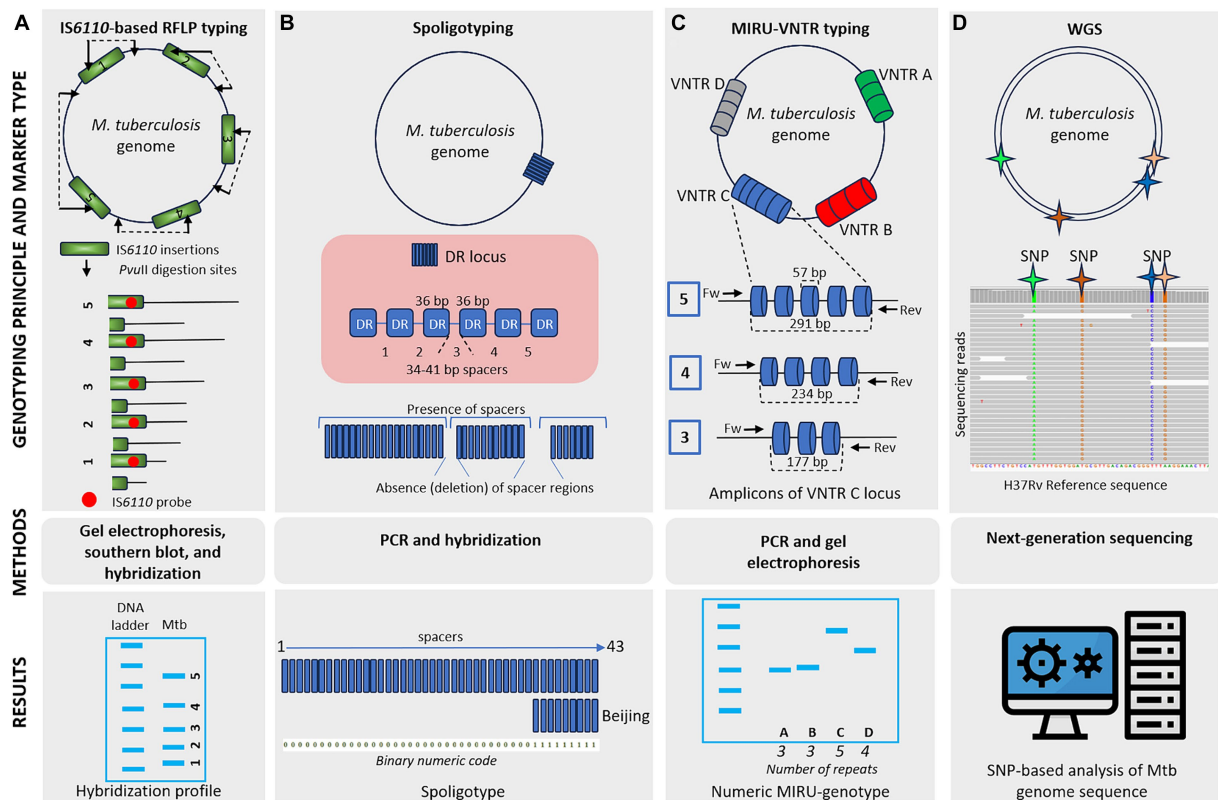
FIGURE 1

Schematic representation of different *M. tuberculosis* marker types and typing methods. **(A)** IS*6110*-based RFLP typing involves the cleaving of mycobacterial DNA with *Pvu*II, generating fragments of varying sizes. These fragments are separated by gel electrophoresis, transferred to a membrane, and hybridized, producing a banding pattern where each band represents a single IS*6110* element with flanking DNA of different lengths. IS*6110*-based RFLP typing offers high discriminatory power, although it is labor-intensive and requires skilled technicians. **(B)** Spoligotyping relies on PCR amplification of a single direct repeat (DR) locus with unique 34−41 bp spacer sequences. Genetic diversity depends on the deletion of these spacer regions. Hybridization of PCR products to a membrane containing oligonucleotides corresponding to spacers generates a pattern of positive or negative signals. Spoligoprofile and the binary code of the Beijing family *M. tuberculosis* strain is shown as an example. Spoligotyping is relatively simple and offers high-throughput capability, but it has limited discriminatory power compared to other methods and may not be suitable for all strain types. **(C)** MIRU-VNTR typing involves the amplification of variable number tandem repeat (VNTR) loci of different repetitive numbers scattered in the genome. PCR products are sized on agarose gels to determine the number of repeats at each locus resulting in a numeric MIRU-genotype. MIRU-VNTR typing provides higher discriminatory power than spoligotyping and allows for simultaneous analysis of multiple loci. **(D)** Whole-genome sequencing (WGS) utilizes information from the entire genome sequence, providing high-resolution data for SNP-based analysis. While WGS offers unprecedented resolution and accuracy, it is costly and requires advanced bioinformatics expertise.
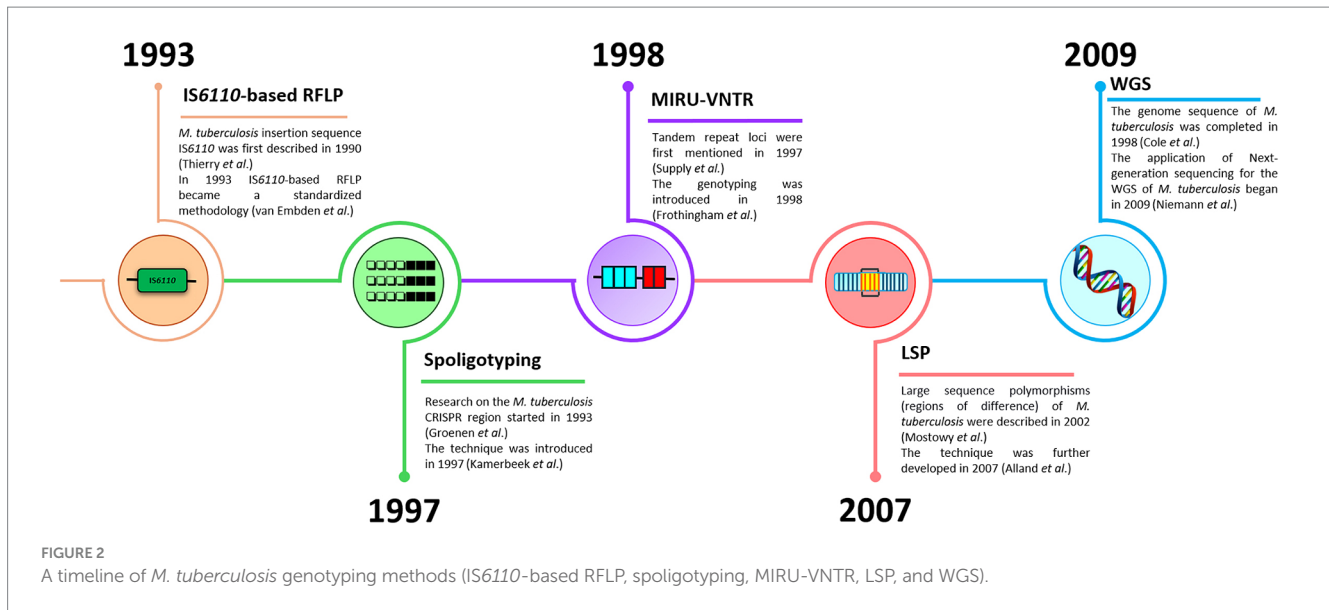
4. For instance, both SIT254 and SIT266 spoligotypes belong to the LAM family. However, in the SITVITWEB database, they are labeled as T5-RUS1 or T1, respectively. Notably, profiles SIT37 and SIT40 (T3 and T4 in SITVITWEB, respectively) have a single deleted spacer, yet they share the same family signature—spacers deleted from 33 to 36. Both of these WGS-based profiles demonstrate potential instances of convergent evolution, where spacer removal independently occurred in different isolates. Thus, it is recommended to be cautious with the interpretation of LAM spoligotypes and to complement spoligotyping with other genotyping methods (Mokrousov et al., 2017; Mokrousov, 2018).

2.1.3 MIRU-VNTR

MIRU profiles identify polymorphic tandem repeat variations in *M. tuberculosis* chromosomes, which classify strains based on the number of repeats at different VNTR loci (Weniger et al., 2010). There are studies based on 12, 15, and 24 loci; therefore, typing at 15 and 24 loci has a higher discriminatory power compared to the classical

method at 12 loci, especially for the Beijing family (Mokrousov, 2015; Shi et al., 2018). East Asian lineage genotypes can also be identified based on analysis of 24 loci using the online tool MIRUVNTRplus[1] (Shitikov et al., 2017). In addition, according to the provided protocol, MLVA (Multiple Locus) VNTR subtyping is carried out based on the variable copy number of tandem repeats. Accordingly, a specific genotype is assigned to facilitate data classification. For example, the L2 family following MIRU-VNTR exhibits a wide copy number range, and MLVA analysis provides insight into clusters within the family at the microevolutionary level. Thus, for each pattern, a unique haplotype MtbC 15–9 is assigned and subsequently divided into genotypes/clusters and look like a set of numbers, so the Beijing family is characterized by the following clusters 94–32, 100–32, etc. (Weniger et al., 2010; Engström et al., 2019).

---

1   http://www.miru-vntrplus.org

FIGURE 2
A timeline of *M. tuberculosis* genotyping methods (IS*6110*-based RFLP, spoligotyping, MIRU-VNTR, LSP, and WGS).

2.1.4 LSP

Large sequence polymorphisms (LSPs) or regions of differences (RDs) are considered ideal phylogenetic markers since horizontal gene transfer rarely occurs in mycobacteria, making them irreversible (Coscolla and Gagneux, 2014). The classification is based on the presence or absence of 15 standard LSPs, which range in size from approximately 200 to 11,000 base pairs. For example, L2 is defined as the presence/deletion of five large genomic deletions (RD105, RD207, RD181, RD150, and RD142). Thus, deletions RD105 and RD207 are markers for the East Asian family, the latter leading to the absence of the first 34 spacers in the DR region. Further, proto-Beijing (RD105 deletion) and two Beijing lineages are divided into sublineages: 1 is Ancestral Beijing (RD105 and RD207 deleted) and 2 is Modern Beijing (RD105, RD207, RD181 deleted) (Cerezo-Cortés et al., 2019). It should be noted that RD markers initially do not differentiate between ancient and modern Beijing, this was achieved using whole genome sequencing (WGS) (Shitikov et al., 2017).

2.1.5 SNP-based classifications

With the development of WGS methods, comparative analysis of genomes has identified sets of phylogenetically significant SNPs, which have been applied to the development of various strain typing methods. WGS has significant discriminatory power because it examines large regions of the genome (or whole genome) and SNPs exhibit very low degrees of homoplasy. The main types of SNP-based classification will be discussed in the next part of the review article. However, before the widespread use of this method, individual regions of the genome were studied and primers were selected to classify the main lineages of *M. tuberculosis*; for example, for L2, an SNP was identified at position Rv3304966 GA (position in reference genome H37Rv) (Fenner et al., 2011; Stucki et al., 2012). Another global classification of *M. tuberculosis* based on SNP divided them into six SCGs (SNP cluster groups), the Beijing family belonged to the SCG2 group (Filliol et al., 2006).

The division of isolates into Ancestral and Modern Beijing is significant for L2, this SNP is located in the MUT genes encoding DNA repair enzymes called ogt, mutT2, and mutT4. Thus, isolates of the ancestral family have *mutT2* mutation in codon 58 (GGA to

CGA), causing the substitution of glycine by arginine, and the *mutT4* mutation in codon 48, causing the substitution of arginine by glycine (Rad et al., 2003). A more refined classification has been proposed based on polymorphisms of genes involved in DNA replication, recombination, and repair (3R). The complete differentiation scheme of the Beijing family divided them into 26 types of Bmyc sequences (Mestre et al., 2011).

Importantly, all proposed SNP-based classifications were consistent with the evolutionary pathways proposed by the other genetic markers described above. Despite this, these methods have limitations in distinguishing genotypes that do not always reflect actual phylogenetic groups, since deep genetic differentiation is only accessible and resolvable with WGS (Thawornwattana et al., 2021). However, it should be mentioned that classical methods are widely used in modern epidemiology, in particular, in the study of localized outbreaks of tuberculosis infection.

Portable long-read sequencing technology, such as the Nanopore platforms from Oxford Nanopore Technologies (ONT), provide an advantage over short-read sequencing methods, particularly in the case of the *M. tuberculosis* genome. This genome is characterized by extensive repeating elements, which pose challenges for short-read sequencing (Dippenaar et al., 2022). Furthermore, the TB Oxford Nanopore Diagnostics test has been developed to effectively detect drug resistance following tuberculosis diagnosis, facilitating clinical decisions regarding the treatment of drug-resistant tuberculosis. The analysis includes the simultaneous sequencing of amplicons from 27 targets, including 24 genes associated with drug resistance, the *hsp65* gene for identifying non-tuberculous mycobacteria, the DR genotyping target, and an internal control (World Health Organization, 2024).

3 Genomic diversity

The study of the genomic diversity of *M. tuberculosis* is characterized by the absence of horizontal gene transfer and is characterized by a strictly clonal population structure. Whole-genome

sequencing methods are often used in studies of genomic diversity. To date, accumulating mycobacterial WGS data provides a phylogenetically robust basis for strain differentiation, coupled with resistance studies, providing information to quantify genomic diversity within or between groups of strains (Coscolla and Gagneux, 2014).

3.1 Lineage 2 classification

Based on techniques such as spoligotyping and RFLP, the Beijing family has been shown to lack genetic diversity and form a relatively recent clonal population. However, WGS analysis revealed that the family has evolved into several distinct sublineages over the past 1800 years (Luo et al., 2015; Liu et al., 2018a). Thus, based on the data, we know that *M. tuberculosis* is divided into 7 lineages, and the Transposon-derived region 1 (TbD1) deletion divides the species into modern and ancient representatives. This deletion is associated with modern strains of MTBC including L2 (East Asian), L3 (East African-Indian), and L4 (Euro-American), while evolutionarily "ancient" strains do not have this deletion (Brosch et al., 2002). The ancient lineages are L1 (Indo-Oceanic), L5 (West African 1), L6 (West African 2), and L7 (Ethiopian) (Comas et al., 2013).

The L2 lineage is in turn subdivided into two large-scale phylogenetic sublineages: proto-Beijing (L2.1) and Beijing (L2.2) (Shitikov et al., 2017; Ajawatanawong et al., 2019). The subsequent division of the L2 lineage is controversial in the scientific community. The main difficulty is the hierarchical population structure of the mycobacteria, which consists of large phylogenetic lineages, smaller sublineages, and, finally, clonal clusters. Thus, in this review, we will focus on the main proposed classifications.

The classification proposed by Coll et al. based on SNPs and consistent with the RD differentiation system, divides *M. tuberculosis* into 7 lineages and 55 sublineages. Interestingly, the phylogenetic analysis in this study revealed the presence of new clades when RD was not able to differentiate them. Thus, in addition to the L2-specific deletion RD105, other deletions (RD207, RD181, RD150, and RD142) were discovered within the L2.2 lineage at the microevolutionary level (Coll et al., 2014). Despite the low Hunter Gaston discriminatory index (0.18), the proposed classification was adapted to changes and the addition of new genotypes (Shitikov et al., 2017).

Further, Merker et al. classified the L2.2 lineage (Beijing) into eight main family sublineages, the scheme included three ancient groups, known as Asia Ancestral 1, 2, 3 and five modern lineages: Asian African 1, 2; Pacific RD150; Europe-Russia B0/ W148 and Central Asia. These phylogenetic branches of the L2 family were constructed based on SNPs. Six sublineages or clonal complexes (CC) based on 24 MIRU-VNTR markers were also established: CC1, CC2, CC3, CC4, CC5, and CC6, and the basal subline BL7 was isolated separately. A subsequent study of the NTF region for the IS*6110* insertion among the selected 337 strains divided CC into two groups, so CC1-CC5 belong to modern isolates, and the remaining CC6 and BL7 include atypical ancestral Beijing variants (Merker et al., 2015). The proposed classification is considered incomplete for the L2 lineage since it does not include the evolutionary subgroup L2.1. However, the Hunter Gaston discrimination index (0.77) has good discriminatory ability (Shitikov et al., 2017).

Another classification was proposed by Shitikov et al., which had higher discriminatory power (HGDI 0.79), as it considers two sub-lineages: L2.1 and L2.2. In turn, the Beijing group (L2.2) consists of 10 groups, three of them belong to the "ancestral" Beijing group (ancestral Asia 1, ancestral Asia 2, ancestral Asia 3), and seven belong to the "modern" Beijing group (Asiatic-African 1, Asia-Africa 2, Asia-Africa 2/RD142, Asia-Africa 3, Pacific RD150, Europe/Russia B0/W148 outbreak and Central Asia). The proposed classification system summarized the SNPs of several studies, and sets of markers were studied to construct a phylogenetic tree of L2 strains (Shitikov et al., 2017). Recently Thawornwattana et al. proposed a classification scheme complementing Shitikov et al., and expanding the L2.2 sublineage. Currently, it is the most complete classification in terms of genetic diversity covering various endemic regions. A classification with a total of 15 genotypes and three levels was proposed, with genotype designations for strains following a scheme similar to Rutaihwa et al. (2019). In addition to numbers, letters were added to the classification; for Ancestral the letters from "A" to "E" were used and for Modern Beijing, it was designated "M." In the proposed classification, it is worth noting the revised Ancestral Beijing genotypes and the proposed six Modern Beijing genotypes (L2.2.M1–L2.2.M6). The revised nomenclature of L2 proposed by Thawornwattana et al. is currently the most comprehensive in comparison with the other known classifications (Figure 3). Figure 3 is based on the Thawornwattana et al. classification scheme of L2, each branch in the nomenclature corresponds to specific SNPs, which are utilized as SNP barcodes. In summary, the proposed classification is easy to interpret, compatible with the known genotypes, and expandable if novel genotypes are discovered (Thawornwattana et al., 2021).

The genomic diversity of *M. tuberculosis*, characterized by a strict clonal population structure and the absence of horizontal gene transfer, plays a key role in the formation of phylogenetic relationships. However, in the case of modern Beijing tuberculosis strains, phylogenetic relationships pose a challenge because the star-shaped phylogeny leads to the independent evolution of new groups. This situation causes controversy and inconsistency between different authors, leading to the proposal of different genotyping schemes in attempts to elucidate evolution, and the diversity of groups depends on the collection of strains. Therefore, newly developed classifications should not ignore previous ones, since this issue is extremely important for public health.

3.2 Beijing B0/W148 and Central Asia outbreak

The phylogenetic division of modern Beijing family strains into clonal clusters has clinical and epidemiological significance. These strains were attributed to be associated with outbreaks of multidrug-resistant tuberculosis (MDR-TB) and characterized by high mutagenic potential, as well as increased transmissibility and hypervirulence (Merker et al., 2018). Distinctive outbreaks circulating in Russia and Central Asia include the following strains: B0/W148, and Central Asia outbreak (CAO) (Shitikov et al., 2017; Merker et al., 2018; Akhmetova et al., 2023). According to the nomenclature of Thawornwattana et al., the outbreaks are characteristic of the L2.2.M4 lineage, and in particular L2.2.M4.5 for B0/W148 and L2.2.M4.9.1 for CAO (Thawornwattana et al., 2021).

The origin of these strains has always been of interest, so initially there was a hypothesis about the emergence of Beijing B0/W148 in
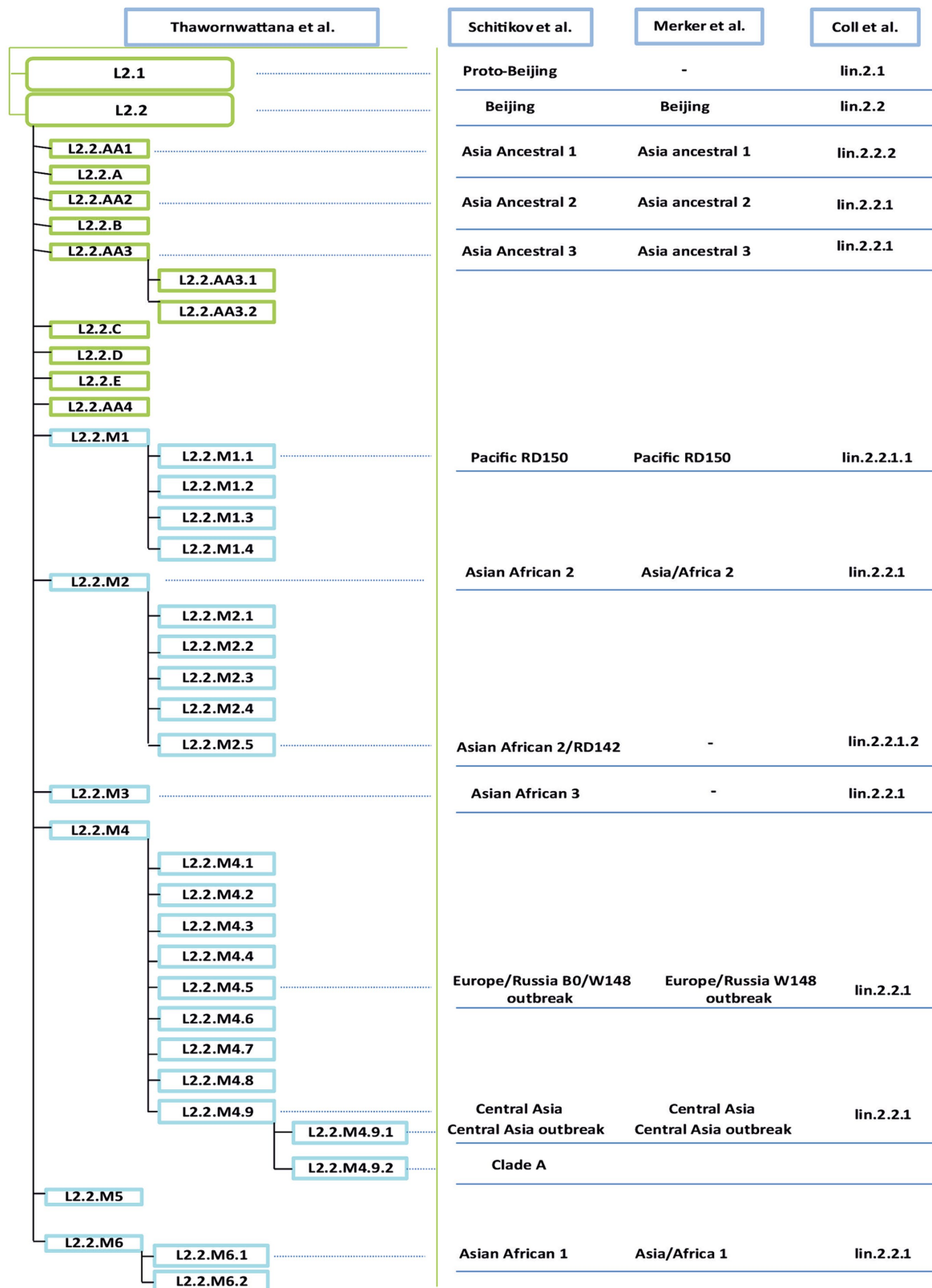
FIGURE 3

Comparison of four phylogenetic classifications of *M. tuberculosis* lineage 2. The authors of the proposed nomenclatures are indicated at the top of the table, from left to right (Coll et al., 2014; Merker et al., 2015, 2018, 2022; Shitikov et al., 2017, 2019; Thawornwattana et al., 2021). The most comprehensive classification is proposed by Thawornwattana et al. is divided into several levels, the level-1 is L2, the level-2 is L2.1, and L2.2. The level-3 divides L2.2 into ancient (A) and modern (M), and level-4 and level-5 represent subsequent genetic clusters. The other classifications of the sublineages and clusters from Shitikov et al. (2017, 2019), Merker et al. (2015, 2018, 2022), and Coll et al. (2014) are connected by lines dash-dotted line with labels from the classification of Thawornwattana et al. (2021).

Siberia, with subsequent spread in 1960–1980 throughout Russia, mainly to the European part of the country (Mokrousov, 2013). Later, Merker correlated the strains with clonal clusters according to MIRU-VNTR. Accordingly, the 94–32 Central Asian (or Central Asian/Russian) cluster including CAO was assigned to the CC1 complex, while the 100–32 cluster including B0/W148 genotype was assigned to the CC2 complex, dominant in Russia and Eastern Europe. Presumably, the most recent expansion of these genetic lineages occurred around the beginning of the 19th century, which may be associated with the national uprising in China from 1861 to 1877, which caused population migration to the Russian Empire, especially to the countries of Central Asia and the spread of CC1 strains and CC2 in this region (Merker et al., 2015). The time of origin was confirmed in a study by Shitikov et al. when they analyzed clades B (part of the B0/W148 outbreak), A and CAO and concluded that these lineages formed approximately 180 years ago (Shitikov et al., 2017). It is worth noting that it is the MDR strains of these clusters that have spread recently to the territories of Russia, Kazakhstan, Kyrgyzstan, and Uzbekistan. The distribution of clonal complexes was accelerated by the disruption of the public health system during the collapse of the USSR (Merker et al., 2015).

Recent studies suggested that the B0/W148 strain appeared in Central Asia in the early 60s of the 20th century and subsequently spread westward, with epidemic waves continuing until the end of the 20th century. Analysis of genetic diversity revealed a similar genetic profile and little variation with a mean of 32 SNPs between strain samples, supporting the idea of their close relationship and therefore outbreak clone status. Thus, the successful spread of this clone was provoked by two factors, namely, the genetic characteristics of the mycobacteria and the impact of historical and political events. Thus, the spread began in the late 70s, followed by a 20-fold increase in the population over 10 years (Merker et al., 2022).

The CAO outbreak is similar in timing to B0/W148, with a lineage beginning in 1974, with the highest probability (95% confidence interval) occurring between 1969 and 1982 (Merker et al., 2018). This clade dominates the territory of Central Asian countries, and also became widespread and acquired MDR in the 90s during the collapse of the USSR (Merker et al., 2018; Klotoe et al., 2019). According to the phylogenetic relationship, it descends from the Central Asian (CA) sublineage L2. Since these clades are of epidemic importance in healthcare, PCR test kits were developed for their rapid detection (Mokrousov et al., 2012; Shitikov et al., 2019).

The genetic diversity of *M. tuberculosis* is represented by 7 lineages. L2 shows a complex phylogenetic substructure and forms star-shaped structures on the tree clusters. This structure causes many controversial issues in the classification of isolates, and it is important for determining epidemiological studies, controlling the spread of the disease and developing effective treatment strategies. In particular, *M. tuberculosis* strains belonging to the B0/W148 and Central Asia outbreak (CAO) clades are associated with outbreaks of multidrug-resistant tuberculosis. Similarities in the evolution and genetic characteristics of these strains highlight the importance of historical events and genetic features in their distribution and epidemiological significance.

4 Geographic distribution

A study of population genomics and genotyping data identifies several hypotheses for the geographical origin of lineage 2, the first

suggests that it has originated in the northeastern region of Asia, and the other suggests its emergence in the southern territory of East Asia with subsequent spread to the South (Mokrousov, 2008; Merker et al., 2015; Yin et al., 2016). It is important to note that there is a growing body of research supporting the assumption of a Southeast origin for the Beijing family (Luo et al., 2015; Guyeux et al., 2022). This conclusion is justified by the high genetic diversity and wide distribution of this family in this region. The case for a southern Chinese origin is strengthened by recent observations of the L2 sublineage described as "Asian Ancestral 4" or L2.2.AA4. This study was conducted in northern Thailand, Chiang Rai, where ethnic minorities have lived since the 7th century. The tribes migrated from southern China to this province about 100 years ago, so this sub-lineage has connections to this region, highlighting a possible southern Chinese origin. It has also been hypothesized that the evolutionary transition from ancestral to modern Beijing sublineages occurred in Southern China. This connection is indicated by the presence of a mutation in Ancestral 4 or L2.2.AA4, which is characteristic of modern Beijing samples (Ajawatanawong et al., 2019). It is worth noting that the origin of the ancestral lineage known as Proto-Beijing also extends to southern China (Luo et al., 2015). Evidence for a northern origin of the family is reflected in a recent study, where strains of the L2.3 (Modern Beijing) or L2.2.M lineage were shown to originate from northern East Asia (Zhu et al., 2023).

The temporal origin of the L2 lineage has been widely speculated; the age of the mycobacteria is 70,000 years, and the L2 lineage is 30,000 years old (Comas et al., 2013; Luo et al., 2015). Sublineage analysis revealed that the time of fusion of two genetic lineages, namely proto-Beijing and ancestral Beijing, occurred approximately 2,200 years ago. The time of separation of all of Beijing's ancestral lineages was also estimated to be approximately 1,300 years ago. Within these estimates, the expansion of Proto-Beijing was estimated to have occurred approximately 900 years ago, while the emergence of modern Beijing was dated to approximately 500 years ago (Liu et al., 2018a). However, analysis of only the L2.2 sublineage revealed a time of 6,000 years before the common ancestor of L2 (Merker et al., 2015). The estimated time of appearance of the most recent common ancestor of modern Beijing specimens, or L2.2.M, occurred 591 years ago. It is assumed that the sublineage arose approximately 1,200 years after the appearance of its ancestor (Zhu et al., 2023). These time estimates help to better understand the evolutionary chronology and genetic dynamics of L2 lineages, as well as their relationship to the history of migrations and interactions of human populations.
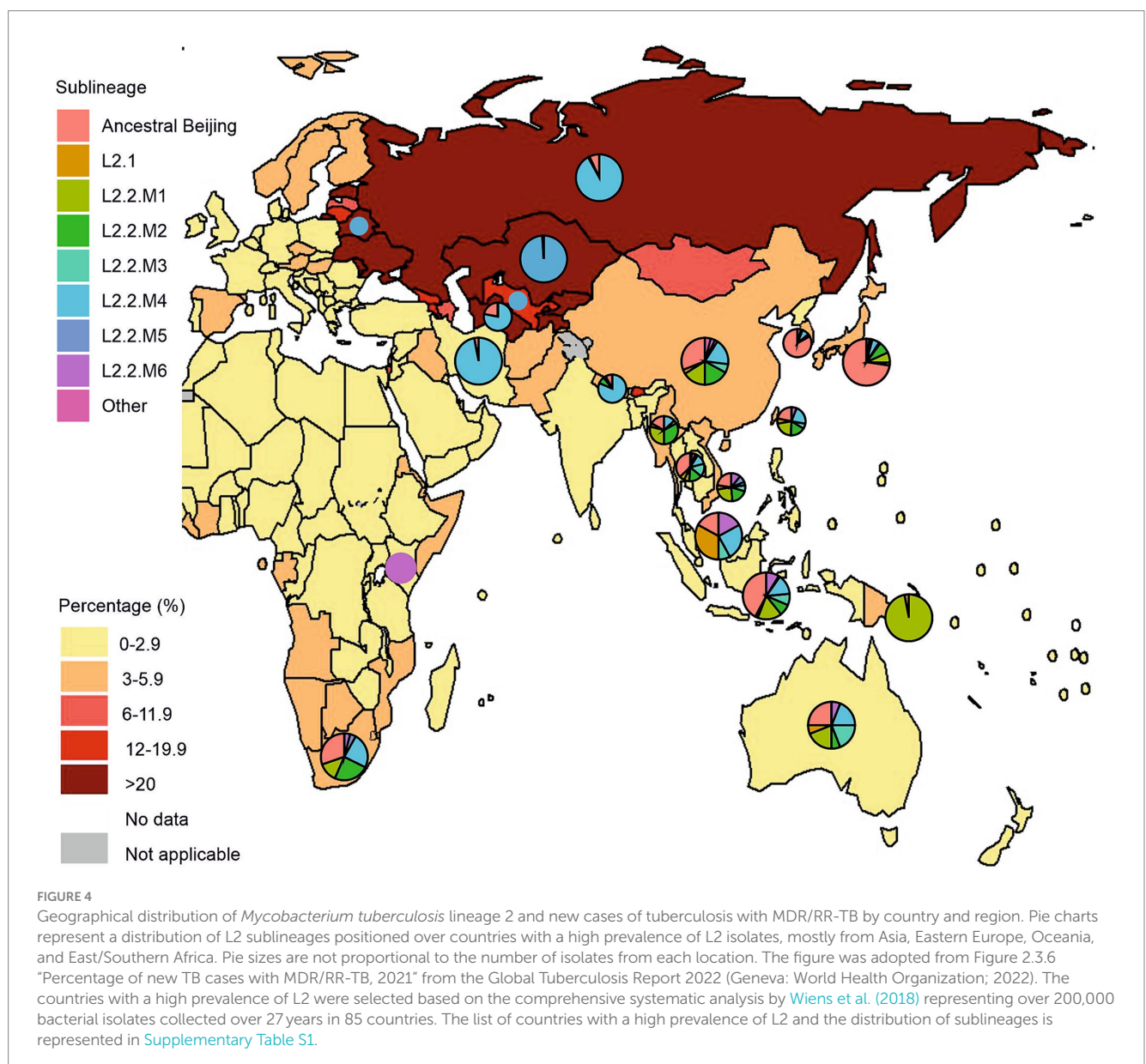
Human-adapted *M. tuberculosis* complex (MTBC) lineages exhibit geographic specialization. Lineages L2 and L4 are widely distributed throughout the world, representing "widespread" lineages, while L5, L6, and L7 are geographically restricted and are "regionally restricted" lineages (Coscolla and Gagneux, 2014; Tarlykov et al., 2020). There has been geographic variability in the distribution of L2 strains to date. In some countries, certain sublineages or even genetic clusters predominate (Figure 4). For example, the overall prevalence of lineage 2 in the East Asia region reaches 60%. This geographic dispersion of L2 sublineages is aggravated by the high prevalence of MDR tuberculosis. Figure 4 visually supports a strong positive correlation between a prevalence of the L2.2.M4 sublineage and a higher proportion of new TB cases with MDR/RR-TB in Russia, Belarus, and Central Asian countries (Kazakhstan, Uzbekistan, and Turkmenistan). Moreover, L2.2.M4 sublineage is known for two MDR-TB outbreak strains of the L2.2.M4.9 Central Asian clade: Clade A and Central Asia Outbreak

(CAO), as well as L2.2.M4.5 (Europe/Russia B0/W148) and L2.2.M4.1 (Bmyc22) outbreak, making it a significant public health concern.

L2.1 (proto-Beijing) is primarily composed of isolates originating from China, Thailand, and Vietnam, with a smaller number from Japan, Malaysia, and Indonesia. Notably, there is not a clear delineation by country, although a distinct cluster from Vietnam suggests the emergence of a local strain. Generally, isolates from China, Thailand, and Vietnam are distributed relatively evenly throughout the phylogeny, spanning across most L2 sublineages. This suggests their longstanding presence and hints at the potential origin and diversification of L2 strains within the region."(Thawornwattana et al., 2021). L2.1 strains do not exhibit distinct clinically significant features and are characterized by moderate transmission capacity, as is the case in Japan (Guyeux et al., 2022). However, a study by Mokrousov et al. found two clusters of the ancient Beijing sublineage in the Russian part of Asia, and, importantly, these strains were multidrug-resistant (Mokrousov et al., 2019, 2023). In contrast, modern sublineages of the Beijing family show a global distribution in various regions of the world (Figure 4). This indicates the high adaptability of modern Beijing sublineages in the context of global mycobacterial dynamics (Luo et al., 2015; Liu et al., 2018a).

According to a study by Merker et al. (2015), the following Beijing CCs exhibit different geographic distribution patterns: CC1 (Central Asia), CC2 (Russia and Eastern Europe), CC3 (East Asia, the Pacific, and the Americas), CC4 (East and Southern Africa and Pacific), CC5 (Pacific, Micronesia and Polynesia), CC6 and BL7 (East Asia, North America and Mexico). The geographic distribution of CCs has been confirmed by several studies. Thus, in East and Southern Africa there are "modern" Asian-African Beijing sublineages (L2.2.M2, L2.2.M3, L2.2.M6.1) and "ancient" Beijing strains (L2.2.AA1, L2.2.AA2, L2.2.AA3). However, strains CC1 and CC2, characteristic of Eastern Europe and Central Asia, are completely absent in African populations. This indicates their relatively recent origin and their likely formation outside East Asia (Rutaihwa et al., 2019).

FIGURE 4
Geographical distribution of *Mycobacterium tuberculosis* lineage 2 and new cases of tuberculosis with MDR/RR-TB by country and region. Pie charts represent a distribution of L2 sublineages positioned over countries with a high prevalence of L2 isolates, mostly from Asia, Eastern Europe, Oceania, and East/Southern Africa. Pie sizes are not proportional to the number of isolates from each location. The figure was adopted from Figure 2.3.6 "Percentage of new TB cases with MDR/RR-TB, 2021" from the Global Tuberculosis Report 2022 (Geneva: World Health Organization; 2022). The countries with a high prevalence of L2 were selected based on the comprehensive systematic analysis by Wiens et al. (2018) representing over 200,000 bacterial isolates collected over 27 years in 85 countries. The list of countries with a high prevalence of L2 and the distribution of sublineages is represented in Supplementary Table S1.

Overall, the L2.2.M1 clade is represented widely across geographic regions, evenly distributed across the countries of China, Japan, South Korea, Taiwan, Vietnam, Thailand, Myanmar, Indonesia, and Australia. This clade is also characterized by local outbreaks in Papua New Guinea and South Africa. L2.2.M2 consists of five subclades designated L2.2.M2.1 – L2.2.M2.5, L2.2.M2.4 is a small group consisting of isolates originating from China, and the remaining four clades are widespread in China, Vietnam, Thailand, and South Africa. It is worth noting that this clade is dominant in Myanmar. The L2.2.M3 clade is distributed evenly in the countries of China, Taiwan, Vietnam, Malaysia, and Indonesia. Interestingly, isolates from Thailand and Australia dominate the number of samples in this cluster. Thus, the Thai samples form a separate subcluster, which is associated with the recent outbreak of MDR in Thailand (Thawornwattana et al., 2021).

The L2.2.M4 clade consists of nine subclades and is expressed by the distribution of clusters across geographic regions, so L2.2.M4.1 and L2.2.M4.2 are found mainly in Thailand. In China, L2.2.M4.3, L2.2.M4.6, and L2.2.M4.8 predominate; L2.2.M4.4 is mainly from South Africa, and L2.2.M4.7 is from Nepalese isolates (Thawornwattana et al., 2021).

In the countries of Central Asia, Eastern Europe, and Russia, L2 strains dominate, mainly the Modern Beijing sublineage, such as L2.2.M.4.5 (B0W/148, Clade B), L2.2.M4.9 (Central Asian), L2.2.M4.9.1 (CAO) and L2.2.M4.9.2 (Clade A) (Ibrayeva et al., 2014; Shitikov et al., 2017; Auganova et al., 2023). At the country level, the highest proportions of new TB cases with MDR/RR-TB were found in the Russian Federation, Central Asia, and several countries in Eastern Europe. Russia was the transboundary area for these clades to Europe and Central Asia, with Eastern Europe a possible center of dispersal (Zhu et al., 2023). At last, L2.2.M5 and L2.2.M6 are small clades formed recently and represent samples from Vietnam, Thailand, and China, in addition, L2.2.M6 are found in Kenya and South Africa (Thawornwattana et al., 2021).

4.1 Transmission dynamics

Although MTBC lineage 2 is represented by sublineages predominantly in certain geographic regions, the global distribution of these strains suggests its high transmission potential in different parts of the world. Previously thought to be a family-wide occurrence of increased transmissibility, such transmission is now known to be limited to the L2.2.M sublineage, and in particular to newly emerging clades within the sublineage (Liu et al., 2016; Zhu et al., 2023).

This increased transmissibility is partly the basis for the "MDR outbreak hypothesis" proposed by Merker et al. Thus, clusters 94–32 (CC1) and 100–32 (CC2) in a study of average pairwise genetic distances between genomes showed low values and fewer mutations, which confirms their relatively recent distribution in Russia and Central Asia. However, the spread of the two clusters occurred before the antibiotic era, which may indicate that the spread of drug resistance in these countries with shared genetic characteristics was not the underlying cause but was likely a consequence of broader problems in public health systems and clinical practice (Merker et al., 2015).

Intriguing data was shown by Zhu et al., when analyzing modern L2.2.M (L2.3) lineages from 51 countries, so according to their classification, 84.8% of strains belong to clades from L2.3.4 to L2.3.6. It is important to note that the "MDR outbreak hypothesis" clades belong to L2.3.6 and even after excluding them from the analysis, the percentage was higher than that of L2.3.1-L2.3.3 (Zhu et al., 2023). This shows that the spread of modern lineages occurred not only in this region but in connection with the B0/W148 and CAO outbreaks throughout the world.

High virulence in modern lineages is associated with positive selection in several genes, such as *ftsK, fadE17*, and, *Rv2209*, which led to their fixation in the population and likely provides adaptation benefits to lineage 2 (Zhu et al., 2023). Merker et al. also isolated the *kdpDE* operon, which is located inside the *Rv0176* gene, encoding a protein associated with MCE1 (molecular complex of extracellular pathogen 1). Thus, in strains of the B0/W148 (CC2) subline, a frameshift mutation was detected in this operon, which presumably may lead to increased virulence due to the formation of a fusion protein with a change in its functions (Merker et al., 2015). Subsequent study of the genetic characteristics of modern Beijing strains can serve as a model for understanding the dynamics of the pathogenicity of this family as a whole.

4.2 Drug resistance and compensatory mutations

Transmission-driven drug-resistant (MDR/XDR) tuberculosis stands as the foremost contributor to human mortality arising from antimicrobial resistance. For example, specific lineage 2 clades, such as B0/W148 or CAO, currently play a pivotal role in the elevated prevalence of MDR-TB in the Eurasian region (Merker et al., 2022).

Shortly after Isoniazid (INH) began to be used against TB in the 1950s, Middlebrook found that most INH-resistant MTB strains were less virulent in Guinea pigs (Middlebrook, 1954; Middlebrook et al., 1954). Based on these observations, it was believed that drug-resistant strains were inherently weaker, or less fit than drug-sensitive strains. In contrast, the situation with the drug-resistant strains was observed to be quite the opposite. One plausible explanation is the acquisition of compensatory mutations after the emergence of resistance mutations. Initial findings by Gagneux et al. emphasized the impact of drug resistance-associated mutations on the bacterial fitness of MTB strains (Gagneux, 2009; Emane et al., 2021). Over time, resistant mycobacteria accumulate additional mutations to offset the initially observed fitness deficits, thereby mitigating the reduction in overall fitness (Maisnier-Patin and Andersson, 2004; Bhatter et al., 2012; Abeldenov et al., 2015; Kodio et al., 2019; Daniyarov et al., 2023). Importantly, the acquisition of compensatory mutations following the introduction of resistance-conferring mutations, such as *rpoB* S450L, plays a crucial role in elevating transmission rates. It has been suggested that compensatory mutations are associated with structural changes in RNA-polymerase *rpoA, rpoB*, and *rpoC*, which increases the transcriptional activity, following the growth of MTB. The *rpoB* gene codes for the RNA polymerase β subunit, which is the target of Rifampicin (RIF), an important anti-tuberculosis drug.

Song et al. extensively documented the occurrence of compensatory mutations strongly linked to the *rpoB* S450L mutation in a population of MDR strains. Compensatory variants were identified in 54% of strains harboring S450L, in stark contrast to only

10% of isolates carrying other mutations in the *rpoB* gene ($p < 0.0001$) (Song et al., 2014). Numerous studies further reinforce a strong association between the RIF resistance-associated mutation S450L and compensatory mutations in *rpoABC* genes (Comas et al., 2011; Kozhamkulov et al., 2011; Li et al., 2016; Liu et al., 2018b; Bainomugisa et al., 2018). This correlation has been firmly established through diverse experimental approaches, including the generation of mutant MTB clones, assessment of growth rates in comparison to the drug-sensitive parent strain, and evaluation of their relative capacity to induce TB disease or mortality in animal models (Brandis and Hughes, 2013; Vos et al., 2013; Boritsch et al., 2016).

Merker et al. (2018) examined the evolutionary history, resistance, and transmission of MTB isolates on the territory of Uzbekistan in Central Asia. The CAO cluster accounted for about three-quarters of all MDR-TB isolates in the studied population (Merker et al., 2018). The existence of mutations that should compensate for fitness deficits was related to the transfer of drug-resistant TB. The average number of drug-resistance mutations was higher among the isolates with compensatory mutations. Moreover, the isolates with compensatory mutations demonstrated a higher ratio of transmission in comparison to the isolates with no compensatory mutations.

Usually drug-resistant strains are detected in a handful of patients and may be less fit than most drug-sensitive strains, however, regions with high rates of DR-TB show MDR burden caused by a few dominant, highly transmitted genetic clusters of closely related strains (Casali et al., 2014; Disratthakit et al., 2015). Globally, the estimated transmission of MDR-TB strains is responsible for over 70% of MDR-TB cases (Kendall et al., 2015). There is a possibility that differences in strain transmissibility and the equilibrium between the fitness costs associated with specific resistance-conferring mutations and the capacity of acquired compensatory mutations to reinstate the fitness of the resistant strains.

A recent investigation has illuminated the link between resistance mutations and their associated fitness costs with the strain genotype (Castro et al., 2020). *In vitro* studies reveal distinct frequencies of fluoroquinolone-resistant gyrase mutations in different strains of *M. tuberculosis*, each exhibiting preferences for specific mutations and varying fitness costs. Lineage-specific mutation preferences were observed in clinical strains, with fluoroquinolone-resistant strains of lineage 2 predominantly featuring the *gyrA* Asp94Gly substitution, while lineage 4 strains more frequently displayed the *gyrA* Ala90Val substitution. Notably, the complexity and variability of the acquisition and effects of these mutations are expected to be more intricate when they manifest in the diverse array of *M. tuberculosis* strains causing tuberculosis.

5 Conclusion

This review article concludes by emphasizing that *M. tuberculosis*, particularly, lineage 2, poses a serious public health threat. Experimental and clinical data confirm the high virulence and high transmissibility of these strains, especially as represented by the L2/Beijing family.

The article also examines the genomic diversity of *M. tuberculosis* and the unique genetic features of lineage 2. Currently, the WGS provides the most comprehensive approach for strain differentiation and assessment of genomic diversity. The

complex phylogenetic structure and star-shaped structures (distribution of genotypes within a cluster where there is one main "center" of high frequency surrounded by less frequent variants) observed in L2, particularly the L2.2.M lineage, pose a challenge to accurately classifying isolates, which highlight the critical role of this lineage in epidemiological studies, especially in the context of multidrug resistance outbreaks caused by the B0/W148 or CAO clade. Drug resistance and the impact of compensatory mutations on the high transmissibility and virulence of lineage 2 are discussed.

The origin of lineage 2 in East Asia, as well as its distribution across various geographical regions, remains controversial and may be an important direction of future research. Another issue under consideration is a unified classification of lineage 2 that should be accepted globally to avoid misinterpretation of MTBC clusters and genotypes. Overall, the article aims to review the genomic characterization of lineage 2 strains, identifying key trends and prospects for future research.

Author contributions

SA: Conceptualization, Visualization, Writing – original draft, Writing – review & editing. DA: Writing – review & editing. PT: Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing.

Funding

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2024.1384791/full#supplementary-material

References

Abeldenov, S., Talhaoui, I, Zharkov, D. O., Ishchenko, A. A., Ramanculov, E., Saparbaev, M., et al. (2015). Characterization of DNA substrate specificities of Apurinic/Apyrimidinic endonucleases from *Mycobacterium Tuberculosis*. *DNA Repair* 33, 1–16. doi: 10.1016/j.dnarep.2015.05.007

Ajawatanawong, P., Yanai, H., Smittipat, N., Disratthakit, A., Yamada, N., Miyahara, R., et al. (2019). A novel ancestral Beijing sublineage of *Mycobacterium Tuberculosis* suggests the transition site to modern Beijing sublineages. *Sci. Rep.* 9:13718. doi: 10.1038/s41598-019-50078-3

Akhmetova, A., Bismilda, V., Chingissova, L., Filipenko, M., Akilzhanova, A., and Kozhamkulov, U. (2023). Prevalence of Beijing central Asian/Russian cluster 94-32 among multidrug-resistant *M. tuberculosis* in Kazakhstan. *Antibiotics* 13:9. doi: 10.3390/antibiotics13010009

Alonso, H., Samper, S., Martín, C., and 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

Auganova, D., Atavliyeva, S., Amirgazin, A., Akisheva, A., Tsepke, A., and Tarlykov, P. (2023). Genomic characterization of drug-resistant *Mycobacterium Tuberculosis* L2/Beijing isolates from Astana, Kazakhstan. *Antibiotics (Basel, Switzerland)* 12:1523. doi: 10.3390/antibiotics12101523

Bainomugisa, A., Lavu, E., Hiashiri, S., Majumdar, S., Honjepari, A., Moke, R., et al. (2018). Multi-clonal evolution of multi-drug-resistant/extensively drug-resistant *Mycobacterium Tuberculosis* in a high-prevalence setting of Papua New Guinea for over three decades. *Microbial Genomics* 4:e000147. doi: 10.1099/mgen.0.000147

Bhatter, P., Chatterjee, A., D'souza, D., Tolani, M., and Mistry, N. (2012). Estimating fitness by competition assays between drug susceptible and resistant *Mycobacterium Tuberculosis* of predominant lineages in Mumbai, India. *PLoS One* 7:e33507. doi: 10.1371/journal.pone.0033507

Bifani, P. J., Mathema, B., Kurepina, N. E., and Kreiswirth, B. N. (2002). Global dissemination of the *Mycobacterium Tuberculosis* W-Beijing family strains. *Trends Microbiol.* 10, 45–52. doi: 10.1016/S0966-842X(01)02277-6

Boritsch, E. C., Khanna, V., Pawlik, A., and Honoré, N. (2016). Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria. *PNAS* 113, 9876–9881. doi: 10.1073/pnas.1604921113

Brandis, G., and Hughes, D. (2013). Genetic characterization of compensatory evolution in strains carrying rpoB Ser531Leu, the rifampicin resistance mutation Most frequently found in clinical isolates. *J. Antimicrob. Chemother.* 68, 2493–2497. doi: 10.1093/jac/dkt224

Brosch, R., Gordon, S. V., 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, 3684–89. doi: 10.1073/pnas.052548299

du Bruyn, E, Cari, S., Remi, D., Qonita, S.-H., Marvin, H., Georgia, S., et al. (2023). Effects of tuberculosis and/or HIV-1 infection on COVID-19 presentation and immune response in Africa. *Nat. Commun.* 14:188. doi: 10.1038/s41467-022-35689-1

Casali, N., Nikolayevskyy, V., Balabanova, Y., Harris, S. R., Ignatyeva, O., Kontsevaya, I., et al. (2014). Evolution and transmission of drug resistant tuberculosis in a Russian population. *Nat. Genet.* 46, 279–286. doi: 10.1038/ng.2878

Castro, R. A. D., Ross, A., Kamwela, L., Reinhard, M., Loiseau, C., Feldmann, J., et al. (2020). The genetic background modulates the evolution of fluoroquinolone-resistance in *Mycobacterium Tuberculosis*. *Mol. Biol. Evol.* 37, 195–207. doi: 10.1093/molbev/msz214

Cerezo-Cortés, M., Rodríguez-Castillo, J. G., Hernández-Pando, R., and Murcia, M. I. (2019). Circulation of *M. tuberculosis* Beijing genotype in Latin America and the Caribbean. *Pathog. Glob. Health* 113, 336–351. doi: 10.1080/20477724.2019.1710066

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

Comas, I., Borrell, S., Roetzer, A., Rose, G., Malla, B., Kato-Maeda, M., et al. (2011). Whole-genome sequencing of rifampicin-resistant *M. tuberculosis* strains identifies compensatory mutations in RNA polymerase. *Nat. Genet.* 44, 106–110. doi: 10.1038/ng.1038

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., et al. (2013). Out-of-Africa migration and Neolithic Coexpansion of *Mycobacterium Tuberculosis* with modern humans. *Nat. Genet.* 45, 1176–1182. doi: 10.1038/ng.2744

Coscolla, M., and Gagneux, S. (2014). Consequences of genomic diversity in *Mycobacterium Tuberculosis*. *Semin. Immunol.* 26, 431–444. doi: 10.1016/j.smim.2014.09.012

Daniyarov, A., Akhmetova, A., Rakhimova, S., Abilova, Z., Yerezhepov, D., Chingissova, L., et al. (2023). Whole-genome sequence-based characterization of pre-XDR *M. tuberculosis* clinical isolates collected in Kazakhstan. *Diagnostics (Basel, Switzerland)* 13:2005. doi: 10.3390/diagnostics13122005

Daniyarov, A., Molkenov, A., Rakhimova, S., Akhmetova, A., Yerezhepov, D., Chingissova, L., et al. (2021). Genomic analysis of multidrug-resistant *Mycobacterium Tuberculosis* strains from patients in Kazakhstan. *Front. Genet.* 12:683515. doi: 10.3389/fgene.2021.683515

Demay, C., Liens, B., Burguière, T., Hill, V., Couvin, D., Millet, J., et al. (2012). SITVITWEB – A publicly available international multimarker database for studying *Mycobacterium Tuberculosis* genetic diversity and molecular epidemiology. *Infect. Genet. Evol.* 12, 755–766. doi: 10.1016/j.meegid.2012.02.004

Dippenaar, A., Goossens, S. N., Grobbelaar, M., Oostvogels, S., Cuypers, B., Laukens, K., et al. (2022). Nanopore sequencing for *Mycobacterium Tuberculosis*: A critical review of the literature, new developments, and future opportunities. *J. Clin. Microbiol.* 60:e0064621. doi: 10.1128/JCM.00646-21

Disratthakit, A., Meada, S., Prammananan, T., Thaipisuttikul, I., Doi, N., and Chaiprasert, A. (2015). Genotypic diversity of multidrug-, quinolone-and extensively drug-resistant *Mycobacterium Tuberculosis* isolates in Thailand. *Infect. Genet. Evol.* 32, 432–439. doi: 10.1016/j.meegid.2015.03.038

Emane, A., Kevin, A., Guo, X., Takiff, H. E., and Liu, S. (2021). Highly transmitted *M. tuberculosis* strains are more likely to evolve MDR/XDR and cause outbreaks, but what makes them highly transmitted? *Tuberculosis* 129:102092. doi: 10.1016/j.tube.2021.102092

Engström, A., Antonenka, U., Kadyrov, A., Kalmambetova, G., Kranzer, K., Merker, M., et al. (2019). Population structure of drug-resistant *Mycobacterium Tuberculosis* in Central Asia. *BMC Infect. Dis.* 19:908. doi: 10.1186/s12879-019-4480-7

Fenner, L., Malla, B., Ninet, B., Dubuis, O., Stucki, D., Borrell, S., et al. (2011). Pseudo-Beijing: evidence for convergent evolution in the direct repeat region of *Mycobacterium Tuberculosis*. *PLoS One* 6:e24737. doi: 10.1371/journal.pone.0024737

Filliol, I., Motiwala, A. S., Cavatore, M., Qi, W., Hazbón, M. H., Bobadilla, M., et al. (2006). Global phylogeny of *Mycobacterium Tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J. Bacteriol.* 188, 759–772. doi: 10.1128/JB.188.2.759-772.2006

Gagneux, S. (2009). Fitness cost of drug resistance in *Mycobacterium Tuberculosis*. *Clin. Microbiol. Infect.* 15, 66–68. doi: 10.1111/j.1469-0691.2008.02685.x

Guerra-Assunção, J. A., Crampin, A. C., Houben, R. M. G. J., Mzembe, T., Mallard, K., Coll, F., et al. (2015). Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. *eLife* 4:e05166. doi: 10.7554/eLife.05166

Guyeux, C., Senelle, G., Refrégier, G., Bretelle-Establet, F., Cambau, E., and Sola, C. (2022). Connection between two historical tuberculosis outbreak sites in Japan, Honshu, by a new *ancestral Mycobacterium tuberculosis* L2 sublineage. *Epidemiol. Infect.* 150:e56. doi: 10.1017/S0950268822000048

Holt, K. E., McAdam, P., Thai, P. V. K., Thuong, N. T. T., Ha, D. T. M., Lan, N. N., et al. (2018). Frequent transmission of the *Mycobacterium Tuberculosis* Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. *Nat. Genet.* 50, 849–856. doi: 10.1038/s41588-018-0117-9

Ibrayeva, A., Kozhamkulov, U., Raiymbek, D., Alenova, A., Igilikova, S., Zholdybayeva, E., et al. (2014). Molecular epidemiology of *Mycobacterium Tuberculosis* strains circulating in the penitentiary system of Kazakhstan. *Int. J. Tuberc. Lung Dis.* 18, 298–301. doi: 10.5588/ijtld.13.0558

Jagielski, T., van Ingen, J., Rastogi, N., Dziadek, J., Mazur, P. K., and Bielecki, J. (2014). Current methods in the molecular typing of mycobacterium tuberculosis and other mycobacteria. *Biomed. Res. Int.* 2014:645802. doi: 10.1155/2014/645802

Karmakar, M., Trauer, J. M., Ascher, D. B., and Denholm, J. T. (2019). Hyper transmission of Beijing lineage *Mycobacterium Tuberculosis*: systematic review and Meta-analysis. *J. Infect.* 79, 572–581. doi: 10.1016/j.jinf.2019.09.016

Kendall, E. A., Fofana, M. O., and Dowdy, D. W. (2015). Burden of transmitted multidrug resistance in epidemics of tuberculosis: A transmission modelling analysis. *Lancet Respir. Med.* 3, 963–972. doi: 10.1016/S2213-2600(15)00458-0

Klotoe, B. J., Kacimi, S., Costa-Conceição, E., Gomes, H. M., Barcellos, R. B., Panaiotov, S., et al. (2019). Genomic characterization of MDR/XDR-TB in Kazakhstan by a combination of high-throughput methods predominantly shows the ongoing transmission of L2/Beijing 94–32 central Asian/Russian clusters. *BMC Infect. Dis.* 19:553. doi: 10.1186/s12879-019-4201-2

Kodio, O., Georges Togo, A. C., Sadio Sarro, Y. D., Fane, B., Diallo, F., Somboro, A., et al. (2019). Competitive fitness of *Mycobacterium Tuberculosis* in vitro. *Int. J. Mycobacteriol.* 8, 287–291. doi: 10.4103/ijmy.ijmy_97_19

Kozhamkulov, U., Akhmetova, A., Rakhimova, S., Belova, E., Alenova, A., Bismilda, V., et al. (2011). Molecular characterization of rifampicin-and isoniazid-resistant *Mycobacterium Tuberculosis* strains isolated in Kazakhstan. *Jpn. J. Infect. Dis.* 64, 253–255. doi: 10.7883/yoken.64.253

Li, Q.-j., Jiao, W.-w., Yin, Q.-q., Fang, X., Li, J.-q., Sun, L., et al. (2016). Compensatory mutations of rifampin resistance are associated with transmission of multidrug-resistant *Mycobacterium Tuberculosis* Beijing genotype strains in China. *Antimicrob. Agents Chemother.* 60, 2807–2812. doi: 10.1128/AAC.02358-15

Liu, Q., Luo, T., Dong, X., Sun, G., Liu, Z., Gan, M., et al. (2016). Genetic features of *Mycobacterium Tuberculosis* modern Beijing sublineage. *Emerg. Microbes Infect.* 5, 1–8. doi: 10.1038/emi.2016.14

Liu, Q., Ma, A., Lanhai Wei, Y., Pang, B. W., Luo, T., Zhou, Y., et al. (2018a). China's tuberculosis epidemic stems from historical expansion of four strains of

Mycobacterium Tuberculosis. *Nat. Ecol. Evol.* 2, 1982–1992. doi: 10.1038/s41559-018-0680-6

Liu, Q., Wang, D., Martinez, L., Lu, P., Zhu, L., Lu, W., et al. (2020). *Mycobacterium Tuberculosis* Beijing genotype strains and Unfavourable treatment outcomes: A systematic review and Meta-analysis. *Clin. Microbiol. Infect.* 26, 180–188. doi: 10.1016/j.cmi.2019.07.016

Liu, Q., Zuo, T., Peng, X., Jiang, Q., Jie, W., Gan, M., et al. (2018b). Have compensatory mutations facilitated the current epidemic of multidrug-resistant tuberculosis? *Emerg. Microbes Infect.* 7:98. doi: 10.1038/s41426-018-0101-6

Luo, T., Comas, I., Luo, D., Bing, L., Jie, W., Wei, L., et al. (2015). Southern east Asian origin and Coexpansion of *Mycobacterium Tuberculosis* Beijing family with Han Chinese. *Proc. Natl. Acad. Sci. USA* 112, 8136–8141. doi: 10.1073/pnas.1424063112

Maisnier-Patin, S., and Andersson, D. I. (2004). Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. *Res. Microbiol.* 155, 360–369. doi: 10.1016/j.resmic.2004.01.019

Merker, M., Barbier, M., Cox, H., Rasigade, J.-P., Feuerriegel, S., Kohl, T. A., et al. (2018). Compensatory evolution drives multidrug-resistant tuberculosis in Central Asia. *eLife* 7:e38200. doi: 10.7554/eLife.38200

Merker, M., Blin, C., Mona, S., Duforet-Frebourg, N., Lecher, S., Willery, E., et al. (2015). Evolutionary history and global spread of the *Mycobacterium Tuberculosis* Beijing lineage. *Nat. Genet.* 47, 242–249. doi: 10.1038/ng.3195

Merker, M., Rasigade, J.-P., Barbier, M., Cox, H., Feuerriegel, S., Kohl, T. A., et al. (2022). Transcontinental spread and evolution of *Mycobacterium Tuberculosis* W148 European/Russian clade toward extensively drug resistant tuberculosis. *Nat. Commun.* 13:5105. doi: 10.1038/s41467-022-32455-1

Mestre, O., Luo, T., Dos Vultos, T., Kremer, K., Murray, A., Namouchi, A., et al. (2011). Phylogeny of *Mycobacterium Tuberculosis* Beijing strains constructed from polymorphisms in genes involved in DNA replication, recombination and repair. *PLoS One* 6:e16020. doi: 10.1371/journal.pone.0016020

Middlebrook, G. (1954). Isoniazid-resistance and catalase activity of tubercle Bacilli; a preliminary report. *Am. Rev. Tuberc.* 69, 471–472. doi: 10.1164/art.1954.69.3.471

Middlebrook, G., Cohn, M. L., and Schaefer, W. B. (1954). Studies on isoniazid and tubercle Bacilli. III. The isolation, drug-susceptibility, and catalase-testing of tubercle Bacilli from isoniazid-treated patients. *Am. Rev. Tuberc.* 70, 852–872. doi: 10.1164/art.1954.70.5.852

Mokrousov, I. (2008). Genetic geography of *Mycobacterium Tuberculosis* Beijing genotype: A multifacet Mirror of human history? *Infect. Genet. Evol.* 8, 777–785. doi: 10.1016/j.meegid.2008.07.003

Mokrousov, I. (2013). Insights into the origin, emergence, and current spread of a successful Russian clone of *Mycobacterium Tuberculosis*. *Clin. Microbiol. Rev.* 26, 342–360. doi: 10.1128/CMR.00087-12

Mokrousov, I. (2015). *Mycobacterium Tuberculosis* Phylogeography in the context of human migration and Pathogen's pathobiology: insights from Beijing and Ural families. *Tuberculosis* 95, S167–S176. doi: 10.1016/j.tube.2015.02.031

Mokrousov, I. (2018). On sunspots, click science and molecular iconography. *Tuberculosis* 110, 91–95. doi: 10.1016/j.tube.2018.04.004

Mokrousov, I., Narvskaya, O., Vyazovaya, A., Otten, T., Jiao, W.-W., Gomes, L. L., et al. (2012). Russian 'successful' clone B0/W148 of *Mycobacterium Tuberculosis* Beijing genotype: A multiplex PCR assay for rapid detection and global screening. *J. Clin. Microbiol.* 50, 3757–3759. doi: 10.1128/JCM.02001-12

Mokrousov, I., Shitikov, E., Skiba, Y., Kolchenko, S., Chernyaeva, E., and Vyazovaya, A. (2017). Emerging peak on the Phylogeographic landscape of *Mycobacterium Tuberculosis* in West Asia: definitely smoke, likely fire. *Mol. Phylogenet. Evol.* 116, 202–212. doi: 10.1016/j.ympev.2017.09.002

Mokrousov, I., Vyazovaya, A., Pasechnik, O., Gerasimova, A., Dymova, M., Chernyaeva, E., et al. (2019). Early ancient sublineages of *Mycobacterium Tuberculosis* Beijing genotype: unexpected clues from Phylogenomics of the pathogen and human history. *Clin. Microbiol. Infect.* 25, 1039.e1–1039.e6. doi: 10.1016/j.cmi.2018.11.024

Mokrousov, I., Vyazovaya, A., Shitikov, E., Badleeva, M., Belopolskaya, O., Bespiatykh, D., et al. (2023). Insight into Pathogenomics and Phylogeography of Hypervirulent and highly-lethal *Mycobacterium Tuberculosis* strain cluster. *BMC Infect. Dis.* 23:426. doi: 10.1186/s12879-023-08413-7

Mokrousov, I., Vyazovaya, A., Sinkov, V., Gerasimova, A., Ioannidis, P., Jiao, W., et al. (2021). Practical approach to detection and surveillance of emerging highly resistant *Mycobacterium Tuberculosis* Beijing 1071-32-cluster. *Sci. Rep.* 11:21392. doi: 10.1038/s41598-021-00890-7

Pardini, M., Niemann, S., Varaine, F., Iona, E., Meacci, F., Orrù, G., et al. (2009). Characteristics of drug-resistant tuberculosis in Abkhazia (Georgia), a high-prevalence area in Eastern Europe. *Tuberculosis (Edinb.)* 89, 317–324. doi: 10.1016/j.tube.2009.04.002

Pérez-Lago, L., Campos-Herrero, M. I., Cañas, F., Copado, R., Sante, L., Pino, B., et al. (2019). A *Mycobacterium Tuberculosis* Beijing strain persists at high rates and extends its geographic boundaries 20 years after importation. *Sci. Rep.* 9:4687. doi: 10.1038/s41598-019-40525-6

Rad, M. E., Bifani, P., Martin, C., Kremer, K., Samper, S., Rauzier, J., et al. (2003). Mutations in putative Mutator genes of *Mycobacterium Tuberculosis* strains of the W-Beijing family. *Emerg. Infect. Dis.* 9, 838–845. doi: 10.3201/eid0907.020803

Rajwani, R., Yam, W. C., Ying Zhang, Y., Kang, B. K., Wong, C., Leung, K. S. S., et al. (2018). Comparative whole-genomic analysis of an ancient L2 lineage *Mycobacterium Tuberculosis* reveals a novel phylogenetic clade and common genetic determinants of Hypervirulent strains. *Front. Cell. Infect. Microbiol.* 7:539. doi: 10.3389/fcimb.2017.00539

Rutaihwa, L. K., Menardo, F., Stucki, D., Gygli, S. M., Ley, S. D., Malla, B., et al. (2019). Multiple introductions of *Mycobacterium Tuberculosis* lineage 2–Beijing into Africa over centuries. *Front. Ecol. Evol.* 7:112. doi: 10.3389/fevo.2019.00112

Saelens, J. W., Viswanathan, G., and Tobin, D. M. (2019). Mycobacterial evolution intersects with host tolerance. *Front. Immunol.* 10:528. doi: 10.3389/fimmu.2019.00528

Shevtsov, T. P., Elena, Z., Dauren, M., Ainur, S., Nazira, M., and Kuvat, M. (2013). Draft genome sequence of *Rhodococcus Erythropolis* DN1, a crude oil biodegrader. *Genome Announc.* 1:e00846-13. doi: 10.1128/genomea.00846-13

Shi, J., Zheng, D., Zhu, Y., Ma, X., Wang, S., Li, H., et al. (2018). Role of MIRU-VNTR and Spoligotyping in assessing the genetic diversity of *Mycobacterium Tuberculosis* in Henan Province, China. *BMC Infect. Dis.* 18:447. doi: 10.1186/s12879-018-3351-y

Shitikov, E., Kolchenko, S., Mokrousov, I., Bespyatykh, J., Ischenko, D., Ilina, E., et al. (2017). Evolutionary pathway analysis and unified classification of east Asian lineage of *Mycobacterium Tuberculosis*. *Sci. Rep.* 7:9227. doi: 10.1038/s41598-017-10018-5

Shitikov, E., Vyazovaya, A., Malakhova, M., Guliaev, A., Bespyatykh, J., Proshina, E., et al. (2019). Simple assay for detection of the Central Asia outbreak clade of the *Mycobacterium tuberculosis* Beijing genotype. *J. Clin. Microbiol.* 57, e00215–e00219. doi: 10.1128/JCM.00215-19

Song, T., Park, Y., Shamputa, I. C., Seo, S., Lee, S. Y., Jeon, H.-S., et al. (2014). Fitness costs of rifampicin resistance in *Mycobacterium Tuberculosis* are amplified under conditions of nutrient starvation and compensated by mutation in the β' subunit of RNA polymerase. *Mol. Microbiol.* 91, 1106–1119. doi: 10.1111/mmi.12520

Stucki, D., Malla, B., Hostettler, S., Huna, T., Feldmann, J., Yeboah-Manu, D., et al. (2012). Two new rapid SNP-typing methods for classifying *Mycobacterium Tuberculosis* complex into the Main phylogenetic lineages. *PLoS One* 7:e41253. doi: 10.1371/journal.pone.0041253

Tarlykov, P., Atavliyeva, S., Alenova, A., and Ramankulov, Y. (2020). Draft genome sequence of an extensively drug-resistant *Mycobacterium Tuberculosis* clinical isolate, 3485_MTB, from Nur-sultan, Kazakhstan. *Microbiol. Resour. Announc.* 9, e00025–e00020. doi: 10.1128/MRA.00025-20

Thawornwattana, Y., Mahasirimongkol, S., Yanai, H., Maung, H. M. W., Cui, Z., Chongsuvivatwong, V., et al. (2021). Revised nomenclature and SNP barcode for *Mycobacterium tuberculosis* lineage 2. *Microbial Genomics* 7:697. doi: 10.1099/mgen.000697

Udoakang, A. J., Alexandra, L. D. Z., Kesego, T., Nora, N. N., Frances, N. O., Claudia, A. A., et al. (2023). The COVID-19, tuberculosis and HIV/AIDS: ménage à trois. *Front. immunol.* 14. doi: 10.3389/fimmu.2023.1104828

Van Soolingen, D., Qian, L., de Haas, P. E., Douglas, J. T., Traore, H., Portaels, F., et al. (1995). Predominance of a single genotype of *Mycobacterium Tuberculosis* in countries of East Asia. *J. Clin. Microbiol.* 33, 3234–3238. doi: 10.1128/jcm.33.12.3234-3238.1995

Vos, M.De, Müller, B., Borrell, S., Black, P. A., van Helden, P. D., Warren, R. M., et al. (2013). "Putative compensatory mutations in the rpoC gene of rifampin-resistant *Mycobacterium Tuberculosis* are associated with ongoing transmission." *Antimicrob. Agents Chemother.* 57. 827–832. doi: 10.1128/AAC.01541-12

Weniger, T., Krawczyk, J., Supply, P., Niemann, S., and Harmsen, D. (2010). MIRU-VNTRplus: A web tool for Polyphasic genotyping of *Mycobacterium Tuberculosis* complex Bacteria. *Nucleic Acids Res.* 38, W326–W331. doi: 10.1093/nar/gkq351

Wiens, K. E., Lauren, P. W., Jorge, R. L., Jennifer, M. R., Roberto, Z.-C., Amador, G., et al. (2018). Global variation in bacterial strains that cause tuberculosis disease: a systematic review and meta-analysis. *BMC Med.* 16:196. doi: 10.1186/s12916-018-1180-x

World Health Organization. (2021). *Catalogue of mutations in mycobacterium tuberculosis complex and their association with drug resistance*. Geneva: World Health Organization.

World Health Organization. (2023). *Global tuberculosis report 2023*. Geneva: World Health Organization. Available at: https://apps.who.int/iris/bitstream/handle/10665/365208/9789240024563-eng.pdf

World Health Organization. (2024). "Consolidated guidelines on tuberculosis. Module 3: diagnosis – rapid diagnostics for tuberculosis detection, third edition." Geneva: World Health Organization.

Yang, Q., Jinjin, H., Jingjing, S., Xinsen, P., Lurong, Z., and Xuejing, Y. (2022). Diagnosis and treatment of tuberculosis in adults with HIV. *Medicine.* 101:e30405. doi: 10.1097/MD.0000000000030405

Yin, Q.-q., Liu, H., Jiao, W.-w., Li, Q.-j., Han, R., Tian, J.-l., et al. (2016). Evolutionary history and ongoing transmission of phylogenetic sublineages of *Mycobacterium Tuberculosis* Beijing genotype in China. *Sci. Rep.* 6:34353. doi: 10.1038/srep34353

Zhu, C., Yang, T., Yin, J., Jiang, H., Takiff, H. E., Gao, Q., et al. (2023). The global success of *Mycobacterium Tuberculosis* modern Beijing family is driven by a few recently emerged strains. *Microbiol. Spectr.* 11, e03339–e03322. doi: 10.1128/spectrum.03339-22

Żukowska, L., Zygała-Pytlos, D., Struś, K., Zabost, A., Kozińska, M., Augustynowicz-Kopeć, E., et al. (2023). An overview of tuberculosis outbreaks reported in the years 2011–2020. *BMC Infect. Dis.* 23:253. doi: 10.1186/s12879-023-08197-w