FEMS Microbiology Reviews, 2025, 49, fuaf017
DOI: 10.1093/femsre/fuaf017
Advance access publication date: 26 April 2025
Review article

Molecular typing of *Mycobacterium tuberculosis*: a review of current methods, databases, softwares, and analytical tools

David Couvin [1,2,*], Anne-Sophie Allaguy[2], Ayoub Ez-zari[3], Tomasz Jagielski [4,*], Nalin Rastogi [1,*]

[1]WHO Supranational TB Reference Laboratory—TB and Mycobacteria Unit, Institut Pasteur de la Guadeloupe, F-97139, Les Abymes, Guadeloupe, France
[2]Laboratoire de Mathématiques Informatique et Applications (LAMIA), Université des Antilles, F-97154, Pointe-à-Pitre, Guadeloupe, France
[3]Laboratory of Biology and Health (UAE/U06FS), Department of Biology, Faculty of Science, Abdelmalek Essaâdi University, BP 2121, 93002 Tetouan, Morocco
[4]Department of Medical Microbiology, Institute of Microbiology, Faculty of Biology, University of Warsaw, I. Miecznikowa 1, 02-096 Warsaw, Poland
*Corresponding authors. David Couvin and Nalin Rastogi, Tuberculosis & Mycobacteria Unit, Institut Pasteur de la Guadeloupe, Les Abymes, F-97139, Guadeloupe, France. E-mails: dcouvin@pasteur-guadeloupe.fr; rastogi.nalin@gmail.com; Tomasz Jagielski, Department of Medical Microbiology, Institute of Microbiology, Faculty of Biology, University of Warsaw, I. Miecznikowa 1, 02-096 Warsaw, Poland. E-mail: t.jagielski@uw.edu.pl
Editor: [Grzegorz Wegrzyn]

Abstract

Studies on the epidemiology and clinical relevance of *Mycobacterium tuberculosis* complex (MTBC) have immensely benefited from molecular typing methods, associated software applications, and bioinformatics tools. Over the last two decades, the Pasteur Institute of Guadeloupe has developed a range of bioinformatic resources, including databases and software, to advance understanding of TB epidemiology. Traditional methods, such as IS*6110*-RFLP, MIRU-VNTR typing, and spoligotyping, have been instrumental but are increasingly supplanted by more precise and high-throughput techniques. These typing methods offer relatively good discrimination and reproducibility, making them popular choices for epidemiological studies. However, the advent of whole-genome sequencing (WGS) has revolutionized *Mycobacterium tuberculosis* complex (MTBC) typing, providing unparalleled resolution and data analysis depth. WGS enables the identification of single nucleotide polymorphisms and other genetic variations, facilitating robust phylogenetic reconstructions, and detailed outbreak investigations. This review summarizes current molecular typing methods, as well as databases and software tools used for MTBC data analysis. A comprehensive comparison of available tools and databases is provided to guide future research on the epidemiology of TB and pathogen-associated variables (drug resistance or virulence) and public health initiatives.

Keywords: *Mycobacterium tuberculosis*; tuberculosis; epidemiology; software; database; drug resistance; family; genomics; genotyping; lineage

Introduction

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* complex (MTBC), remains a critical global health issue. According to the World Health Organization (WHO) Global TB Report 2024, 10.8 million people developed TB in 2023, and 1.25 million people succumbed to the disease. Encouragingly, 74 million lives have been saved since 2000 through WHO-guided worldwide TB control actions. TB is widespread throughout the world, yet it disproportionately affects low-income regions, including Sub-Saharan Africa and South Asia, in particular 30 high TB burden countries (HBCs), which accounted for 87% of the global TB burden in 2023. Limited healthcare infrastructure certainly exacerbates the epidemic in HBCs, since the disease has been linked to poverty, malnutrition, promiscuity, poor living conditions, and other unfavorable socio-demographic factors (Farmer et al. 2006, Keshavjee et al. 2008). Furthermore, HIV infection and drug resistance, whose prevalence has been particularly high in many HBCs, are major impediments to successful TB treatment, and have considerably contributed to the persistence of the global TB epidemic. Drug-resistant *Mycobacterium tuberculosis* has recently been introduced into the WHO bacterial priority pathogens list, 2024 (https://www.who.int/publications/i/item/9789240093461), further emphasizing the need to improve TB control and prevention, especially in resource-limited settings.

The MTBC includes a group of closely related species (e.g. *Mycobacterium tuberculosis sensu stricto*, *Mycobacterium africanum*, *Mycobacterium bovis*, *Mycobacterium caprae*, *Mycobacterium pinnipedii*, *Mycobacterium suricattae*, *Mycobacterium orygis*, *Mycobacterium microti*, *Mycobacterium mungi*, and probably other species) that are potentially pathogenic for both humans and animals. These species generally belong to one or more phylogenetic lineages (or clades/families). TB-dedicated genotyping databases and software tools developed in the Institut Pasteur de la Guadeloupe are able to provide a holistic view of specific aspects of TB research by aiding the analysis of data collected from numerous TB laboratories worldwide (data available for >128 000 MTBC strains from 160 countries; the most recent versions under development being SITVITEXTEND and SITVITGeno). Figure 1 displays a brief history of evolution of various SpolDB/SITVIT databases over time in our laboratory (Sola et al. 1999, 2001, Filliol et al. 2003, Brudey et al. 2006, Demay et al. 2012, Couvin et al. 2019, 2022).

Received 2 January 2025; **revised** 20 April 2025; **accepted** 25 April 2025

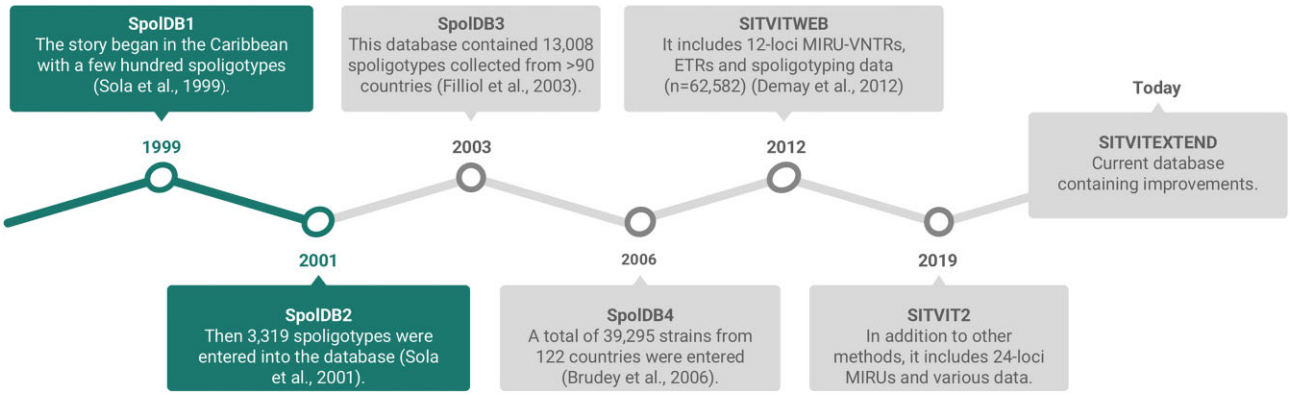

Figure 1. Brief histogram of SpolDB/SITVIT database evolution.

In this manuscript, we review several molecular tools, such as IS*6110*-RFLP, spoligotyping, MIRU-VNTR, and whole-genome sequencing (WGS), that have revolutionized TB epidemiology, offering insights into genetic diversity, transmission patterns, and drug resistance mechanisms. This review evaluates current molecular typing methods, WGS advancements, and bioinformatics tools, emphasizing their potential to address critical gaps such as inadequate strain characterization, limited access to drug resistance data, and challenges in tracing transmission dynamics in TB-endemic regions.

Chapter 1: an overview of TB molecular typing techniques

Since the early understanding of TB transmissibility and control, monitoring and surveillance of the disease have traditionally relied on conventional methods, such as contact tracing, which involve human interviews and the collection and analysis of extensive demographic and clinical data (Fox et al. 2013). Over the past three decades, the emergence of molecular epidemiology has significantly enhanced our understanding of TB transmission and evolution, thereby greatly contributing to both public health strategies and clinical management of this devastating disease (Jagielski et al. 2016).

Molecular typing, or genotyping, of *M. tuberculosis* strains is a cornerstone of molecular TB epidemiology. A fundamental assumption is that strains with identical or highly similar genotyping patterns form a "genotypic cluster," which is considered a proxy for cases arising from recent transmission. Despite variations in the criteria used to define genetic identity, genotyping has proven irreplaceable in investigating outbreaks, deciphering chains of transmission or distinguishing between relapses and reinfections.

Furthermore, molecular typing has been crucial in phylogenetic and evolutionary studies of TB, allowing to identify major lineages and emerging clones, which are often associated with specific geographical regions, increased transmission rates, or acquisition of virulence and drug resistance traits (Mathema et al. 2006, Manson et al. 2017, Dookie et al. 2018). Overall, genotyping has become an essential tool for gaining deep insights into the genetic diversity, prevalence patterns, circulation, and evolution of the pathogen.

Both historical and contemporary methods of *M. tuberculosis* genotyping were reviewed earlier (Mathema et al. 2006, Jagielski et al. 2014, 2016). Here, we briefly describe and update the most widely accepted and currently used modalities.

IS*6110*-RFLP typing

The insertion sequence *6110* (IS*6110*) was among the first genetic elements used as a marker for strain typing of *M. tuberculosis* (Thierry et al. 1990). IS*6110* is a 1355 bp-long IS3 family sequence, uniquely found in MTBC. It usually occurs in multiple copies ranging from 0 to 25 (five copies are presented), dispersed across the entire genome. The copy number and their chromosomal location determine the high level of marker's polymorphism (Fig. 2A). The IS*6110*-RFLP was duly standardized, and several local or international databases were constructed (Van Embden et al. 1993, Heersma et al. 1998, Crawford et al. 2002), leading it to be considered as a gold standard for *M. tuberculosis* genotyping, as of the early 2000s (Kremer et al. 1999, Clark et al. 2006, Bifani et al. 2009). The IS*6110*-RFLP technique is highly discriminatory and generates profiles that are stable over time, yet whose rate of change allows detection of ongoing transmission events (Fang et al. 1998, Yeh et al. 1998, de Boer et al. 1999, Niemann et al. 2000). Apart from the identification of TB transmission chains and investigation of TB outbreaks (Valway et al. 1998, Kubín et al. 1999, Diel et al. 2004, Ruddy et al. 2004, Devaux et al. 2009), it was IS*6110*-RFLP that allowed, for the first time, to differentiate relapse from reinfection (Van Embden et al. 1993, van Rie et al. 1999) or trace the source of laboratory cross-contamination (Van Duin et al. 1998).

Despite its utility, the IS*6110*-RFLP typing system has several limitations, including its labor-intensive procedure, culture dependency, requirement for high DNA yield (>1 µg), need for advanced computer software, skilled personnel, and the lack of inter-laboratory reproducibility due to variations in assay conditions and interpretation of banding patterns (Van Soolingen and Arbeit 2001, Braden et al. 2002). Additionally, it has low discriminatory power for isolates with five or fewer IS*6110* copies, common in certain Asian regions where such isolates make up 47%–72% of circulating TB bacilli (Das et al. 2005, Rienthong et al. 2005, Joseph et al. 2013). These drawbacks led to a preference for Polymerase Chain Reaction (PCR)-based techniques, which often targeted IS*6110* (Friedman et al. 1995, Otal et al. 1997, Reisig et al. 2005, Thorne et al. 2007), but none of these methods achieved widespread adoption due to the lack of standardized performance measures and reference databases for cross-laboratory comparisons. Recent improvements in IS6110-based typing, such as the IS6110-5′3′FP (for IS6110 5′ and 3′ fluorescent polymorphisms) and semiautomated IS*6110*-*Pvu*II systems (tailored to the RiboPrinter

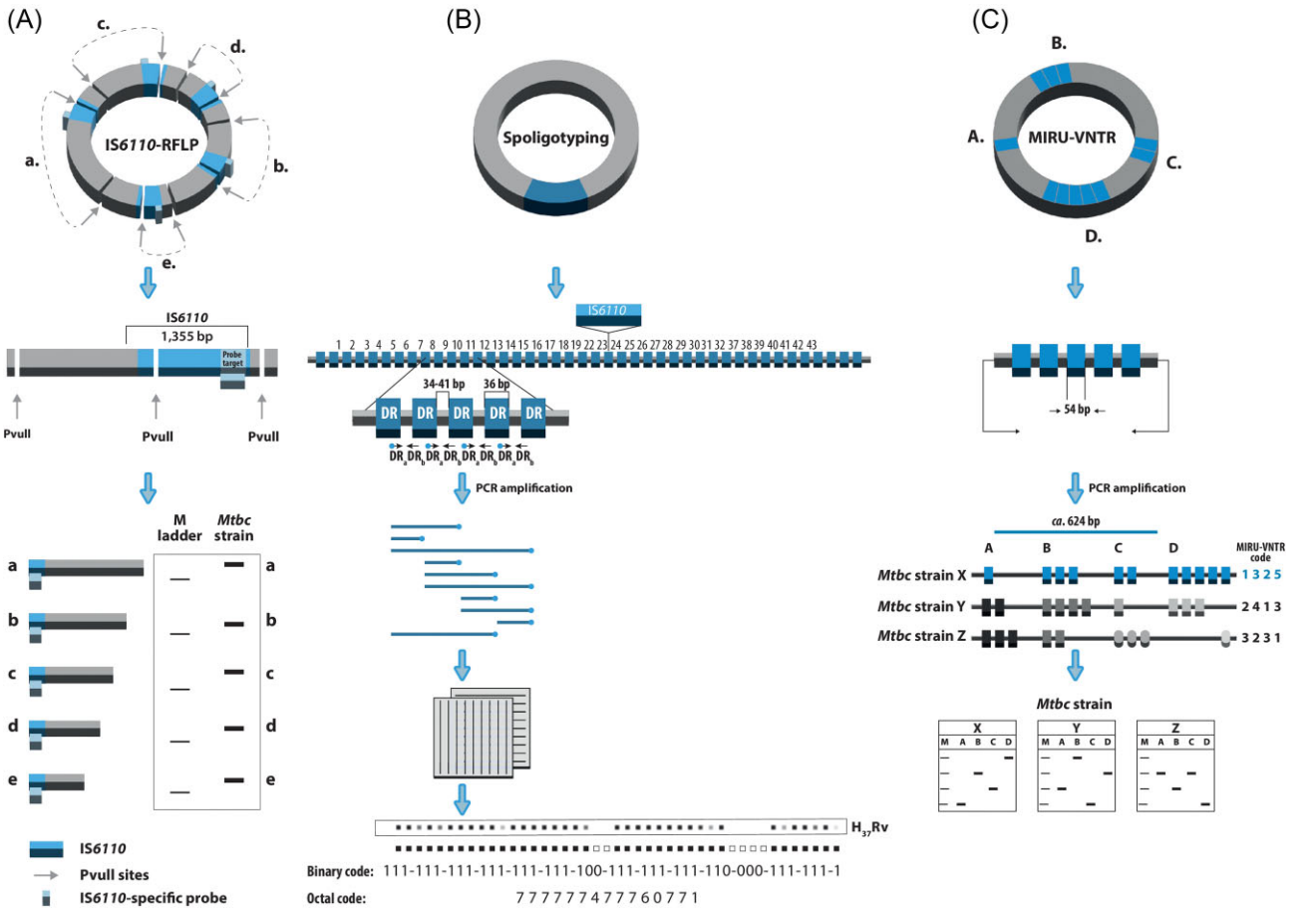

Figure 2. Schematic representation of three typing schemes used for *M. tuberculosis* strains. (A) Technically, the IS6110 genotyping is a Restriction Fragment Length Polymorphism (RFLP)-based method and involves genomic DNA digestion with PvuII endonuclease, electrophoretic separation of the fragments thus produced, and their hybridization with a peroxidase-labeled probe complementary to the 3′ end of the IS6110, allowing each copy of the sequence to be visualized as a separate band on an autoradiogram. (B) The standard spoligotyping procedure begins with PCR amplification of the entire DR region using two inversely oriented primers, complementary to short DR sequences, with one primer being biotinylated to make all PCR products labeled. These products are then hybridized to a membrane with a set of 43 immobilized, covalently bound synthetic oligonucleotides, corresponding to unique spacer sequences identified in either *M. tuberculosis* H37Rv or *M. bovis* BCG strains. Afterward, the membrane is incubated with a streptavidin-peroxidase or streptavidin-alkalic phosphatase conjugate, and the hybridization signals are detected by chemiluminescence. Strain-specific patterns (spoligotypes) are visualized autoradiographically by exposing the membrane to X-ray film. The presence or absence of a given spacer is represented by black squares ("1") or blank spaces ("0"), respectively. Thus, a spoligotype is expressed as a 43-digit binary code, which can further be converted into octal code. (C) The MIRU-VNTR typing technique involves two major steps: PCR amplification of each MIRU-VNTR locus, with primers complementary to their flanking regions and analysis of thus produced amplicons, resolved electrophoretically. The number of tandem repeat units, at each locus, is deduced from the amplicon size, in relation to the known size of the repeat unit within the specific locus. The final result is a multidigit numerical code (MIRU-VNTR code), corresponding to the repeat number at each locus.

microbial characterization system, DuPont Molecular Diagnostics, USA), offer better discriminatory power and technical flexibility, as well as improved throughput, reproducibility, and data portability (Thabet et al. 2014, Said et al. 2016, Dekhil et al. 2018). Nonetheless, high setup and maintenance costs may still deter most potential users.

To sum up, despite its high discriminatory power, limitations such as labor-intensive protocols and low reproducibility have led to decline of IS6110-RFLP in favor of PCR-based methods. Yet, regardless of the rise of WGS as the current gold standard for *M. tuberculosis* genotyping, IS6110-RFLP continues to be used for local TB investigations, either alone (Diaz et al. 2001, Razanamparany et al. 2002, Chauhan et al. 2007, Pescarini et al. 2018, Essahale et al. 2024), or in combination with other methods, like spoligotyping and MIRU-VNTR typing, particularly in low-income countries (Groenheit et al. 2011, Peres et al. 2018, Ei et al. 2019, Pokam et al. 2019, Chisompola et al. 2021).

Spoligotyping

Spacer oligonucleotide typing, or spoligotyping (illustrated in Fig. 2B), is a widely used method for *M. tuberculosis* genotyping. It targets polymorphisms in the Clustered Regularly Interspersed Short Palindromic Repeats (CRISPR) in the Direct Repeat (DR) locus, consisting of 36-bp repeats interspersed with non-repetitive, 35–41-bp spacers, whose variability provides discriminatory power (Hermans et al. 1991, Groenen et al. 1993). Spoligotyping detects the presence or absence of 43 spacers selected from the *M. tuberculosis* H37Rv (spacers 1–19, 22–32, and 37–43) and *M. bovis* BCG vaccine strain P3 (spacers 20–21 and 33–36), yielding binary results suited for database portability and inter-laboratory comparisons (Groenen et al. 1993, Kamerbeek et al. 1997), as well as identification of members of the MTBC at both species and subspecies levels (Plikaytis et al. 1993, Kremer et al. 2004). Major databases include SpolDB4 (Brudey et al. 2006), SITVITWEB (Demay et al. 2012), and its 2019 update (Couvin et al. 2019), along

with online tools like SPOTCLUST (Vitol et al. 2006), SpolLineages, and SpolSimilaritySearch (Couvin et al. 2017, 2020). These enable global tracking of TB genotypes (Eldholm et al. 2006, Ani et al. 2010, Dong et al. 2010, Tilahun et al. 2018).

Spoligotyping is fast, cost-effective, and highly sensitive, requiring only 10 fg of DNA, equivalent to the quantity from 2 to 3 bacterial cells (Jagielski et al. 2016). As a culture-independent method, it can be performed on diverse sample types, including TB-positive smears, paraffin-embedded tissue sections, or paleopathological specimens (Van Der Zanden et al. 1998, Zink et al. 2003, Schewe et al. 2005, Molina-Moya et al. 2018). Though initially used alone (De La Salmonière et al. 1997, Heyderman et al. 1998, Niang et al. 1999, Soini et al. 2000, Mistry et al. 2002, Puustinen et al. 2003, Augustynowicz-Kopeć et al. 2008), spoligotyping was unable to accurately assess the epidemiological links between TB cases (De La Salmonière et al. 1997, Goyal et al. 1997, Cronin et al. 2001). It is therefore combined with higher-resolution methods like IS*6110*-RFLP or MIRU-VNTR typing for enhanced epidemiological insights (Diaz et al. 1998, Cowan et al. 2005, Clark et al. 2006, Joseph et al. 2013, Bouklata et al. 2015, Jagielski et al. 2015, Ribeiro et al. 2015, Bakuła et al. 2019). Nevertheless, due to financial and/or organizational reasons, some laboratories, especially in developing countries, still rely on spoligotyping only (Zewdie et al. 2016, Elegail et al. 2018, Ramazanzadeh et al. 2020, Bellad et al. 2022, Hussien et al. 2022). Attempts to improve its discriminatory power, such as second-generation spacers (Van Embden et al. 2000, Van der Zanden et al. 2002), did not significantly increase the discriminatory resolution for *M. tuberculosis* (Van der Zanden et al. 2002, Kremer et al. 2005).

Efforts to improve technical aspects of spoligotyping involved (i) attempts to increase its high-throughput capacity, (ii) mitigate interpretative ambiguities with manual reading of the membrane, and (iii) expedite the turnaround time for obtaining results. Advanced detection techniques, such as Luminex technology, where the spacer probes are immobilized on microspheres and detected, upon hybridization, with PCR products and fluorochrome-mediated binding, by laser-based flow cytometry (Cowan et al. 2004, Zhang et al. 2010), were attempted. Additional developments included (i) advanced Luminex analyzer (MAGPIX) based on the use of magnetic beads and light-emitting-diode/charge-coupled-device image-based detection system (Ocheretina et al. 2013); (ii) Matrix Assisted Laser Desorption Ionization - Time of Flight Mass Spectrometry (MALDI-TOF MS) for spoligotype detection, with the hybridization step replaced with a multiplexed primer extension assay (Honisch et al. 2010); and (iii) variety of microarray platforms designed for spoligotyping to optimize its performance and efficiency (Song et al. 2007, Gomgnimbou et al. 2012, Bespyatykh et al. 2014). Though promising with respect to working time and data processing, these assays remain limited by high costs (Honisch et al. 2010, Ocheretina et al. 2013, Bespyatykh et al. 2014), especially for middle- and low-income countries. Newer innovations, such as a new, three-reaction, one-step real-time PCR-based McSpoligotyping and its refined, single-tube version (MeltArray-based spoligotyping), proposed as a rapid and reliable alternative for the conventional spoligotyping protocol, might have the potential to be implemented in resource-limited settings (Zeng et al. 2018, Xia et al. 2024).

In silico spoligotyping tools such as SpoTyping (Xia et al. 2016), SpolPred (Coll et al. 2012), SpolPred2 (Napier et al. 2023), lorikeet (Cohen et al. 2015), and TGS-TB (Sekizuka et al. 2015) offer simplified workflows and compatibility with WGS. Note that the latter provides *in silico* genotyping for spoligotyping as well as other typing formats, including the analysis of IS*6110* insertion sites and customized VNTR loci. *In silico* spoligotyping has been employed in several WGS-based studies, allowing backward compatibility of WGS with molecular spoligotyping (Coll et al. 2012, Hijikata et al. 2017, Gautam et al. 2018, Bogaerts et al. 2021, Genestet et al. 2022, Bakuła et al. 2023a, Napier et al. 2023).

Direct comparisons between conventional and *in silico* spoligotyping are discouraged due to factors like sequence read quality and bioinformatic criteria. *In silico* methods may miss changes in the DR locus, such as IS*6110* insertions, but improve accuracy in depicting strain relatedness (Bakuła et al. 2023a).

Despite limitations, including homoplasy and being less discriminatory for closely related strains (Reyes and Tanaka 2010, Reyes et al. 2012), spoligotyping remains valuable for assessing genetic diversity and phylogenetic relationships among *M. tuberculosis* strains (Liang et al. 2020, Bakuła et al. 2023b, Yin et al. 2023). Even after three decades, spoligotyping remains a largely used method in the investigation of genetic diversity and transmission dynamics of TB bacilli circulating within specific populations and settings (Razo et al. 2018, Shi et al. 2018, Ramazanzadeh et al. 2020, Hussien et al. 2022, Yin et al. 2023, Rudeeaneksin et al. 2024, Valencia-Trujillo et al. 2024). With nearly 1800 PubMed articles referencing it as of December 2024, it continues to feature prominently in TB molecular epidemiology.

MIRU-VNTR genotyping

Minisatellite-like VNTR loci were identified in *M. tuberculosis* genomes in the late 1990s (Supply et al. 1997, Frothingham and Meeker-O'Connell 1998). These 40–100-bp mycobacterial interspersed repetitive units (MIRUs) are scattered across 41 chromosomal locations (Supply et al. 2000). A 12-locus MIRU-VNTR typing scheme (shown in Fig. 2C) was developed for genotyping (Mazars et al. 2001), offering high-throughput analysis through PCR and gel or capillary electrophoresis (Supply et al. 2001, Nikolayevskyy et al. 2016b, Tafaj et al. 2020). Further, its digitized results allowed easy global database integration, aiding researchers. Remarkably, the 12-locus MIRU-VNTR typing is more discriminatory than spoligotyping and IS*6110*-RFLP for IS*6110* low-copy strains (Mazars et al. 2001, Cowan et al. 2002, Lee et al. 2002); however, it is less effective for high-copy strains unless combined with another typing method (Blackwood et al. 2004, Cowan et al. 2005, Gopaul et al. 2006).

To enhance resolution, a standardized 24-locus format, including a subset of 15 discriminatory loci, was proposed, suitable for epidemiology and phylogenetic studies (Supply et al. 2006, Oelemann et al. 2007, Allix-Béguec et al. 2008). However, homoplasy issues necessitate lineage-specific locus sets (Comas et al. 2009, Maghradze et al. 2022), particularly for Beijing lineage strains necessitating hypervariable loci (Iwamoto et al. 2007, Mokrousov et al. 2008, Comas et al. 2009, Velji et al. 2009, Allix-Béguec et al. 2014), different from the standard 15- or 24-loci formats. Thus, a consensus set of 4 hypervariable loci was proposed as an adjunct to standard typing for Beijing clonal clusters (Allix-Béguec et al. 2014).

MIRU-VNTR typing has largely replaced IS*6110*-RFLP as the gold standard due to its technical advantages (Merker et al. 2017), and has been applied to studies on TB transmission (Van Deutekom et al. 2005, Oelemann et al. 2007, Maes et al. 2008, Bidovec-Stojkovic et al. 2011, Mansoori et al. 2018, Chen et al. 2022, Maghradze et al. 2022), discriminate relapses from reinfections (Afshar et al. 2019, Maghradze et al. 2019, Shao et al. 2021), identify mixed infections (Wang et al. 2015, Kargarpour Kamakoli et al. 2020, Micheni et al. 2022), and laboratory cross-contaminations (Martín et al. 2008).

Used alongside spoligotyping, it aids in unraveling genetic diversity and evolutionary relationships (Sola et al. 2003, Chaoui et al. 2014, Bouklata et al. 2015, Shi et al. 2018), leveraging databases like MIRU-VNTRplus that allow phylogenetic comparisons between worldwide samples of TB bacilli populations (Allix-Be´guec et al. 2008, Weniger et al. 2010). Despite advancements in sequencing, MIRU-VNTR remains an efficient tool for TB epidemiology, with emerging *in silico* approaches promising to replace conventional methods (Rajwani et al. 2018, Maeda et al. 2020). Digital MIRU-VNTR typing, performed on complete or draft genome sequences, is expected to eventually replace the conventional procedure in future, similar to spoligotyping.

Whole-genome sequencing

The advent of WGS has transformed the study of pathogen genetics, including tubercle bacilli, delivering significant advancements in TB epidemiology over the past two decades (Box 1). WGS surpasses spoligotyping, MIRU-VNTR, and other methods in determining genetic relatedness among *M. tuberculosis* strains (Nikolayevskyy et al. 2019). Two key WGS approaches are widely used: single nucleotide polymorphism (SNP) variant calling, which identifies single nucleotide differences with a reference genome (e.g. *M. tuberculosis* H37Rv, GCF_000195955.2) and provides robust phylogenetic markers due to the rarity of SNP events and low homoplasy (Stucki and Gagneux 2013, Gagneux 2018), and gene-by-gene typing, which detects allelic variations in core or accessory genes, extending the multi-locus sequence typing (MLST; Maiden et al. 2013).

Early studies demonstrated WGS's superiority over IS*6110*-RFLP and MIRU-VNTR for inferring epidemiological links and genetic relatedness among *M. tuberculosis* isolates (Schürch et al. 2010, Gardy et al. 2011). A systematic review confirmed WGS's higher discriminatory power compared to classical genotyping (Nikolayevskyy et al. 2016a). WGS frequently subdivides MIRU-VNTR clusters, ruling out false transmission events. MIRU-VNTR clustering rates were overestimated by 7%–92%, particularly for monomorphic lineages like the Beijing family (Gurjav et al. 2016, Stucki et al. 2016, Meehan et al. 2018, Wyllie et al. 2018, Alaridah et al. 2019). Additionally, WGS identified transmission events missed by conventional epidemiological methods (Nikolayevskyy et al. 2016a, 2019).

The SNP threshold for defining transmission clusters is critical. A 5-SNP cut-off, based on within-strain divergence over three years, is often used to indicate recent transmission, while >12 SNPs suggest no direct link (Walker et al. 2013). Subsequent studies validated the 5-SNP threshold for epidemiologically linked cases (Casali et al. 2016, Norheim et al. 2017, Iwamoto et al. 2023; Zhang et al. 2023). However, appropriate SNP thresholds vary depending on factors like strain diversity, read quality, within-host diversity, and amplification steps (Hatherell et al. 2016). Some studies identified links at 2–3 SNPs (Roetzer et al. 2013, Walker et al. 2014, 2018), while others found connections even at >12 SNPs, which is defined as the upper limit of genomic relatedness between epidemiologically related individuals (Luo et al. 2014, Nikolayevskyy et al. 2016a, Jajou et al. 2018, Cancino-Muñoz et al. 2022, Xiao et al. 2024).

WGS and SNP differences are used to distinguish relapse from reinfection, with SNP distances varying significantly between original and new infections (0–8 vs. >1000 SNPs) (Bryant et al. 2013, Witney et al. 2017). SNP-based WGS protocols face challenges such as sequencing errors, base-calling inaccuracies, and incomplete genome assembly due to repetitive elements (Meacham et al. 2011, Ahmad et al. 2021). Additionally, the lack of global data standardization and diverse analysis pipelines hinder interlaboratory comparisons (Merker et al. 2015, Kohl et al. 2018a, Meehan et al. 2019). To address these limitations, core genome multilocus sequence typing (cgMLST) offers a standardized approach, using uniform allele numbering to describe genetic variation based on a defined scheme of loci and alleles (Maiden et al. 2013, Kohl et al. 2014, 2018a).

The initial cgMLST scheme included 3257 loci shared among reference genomes of Lineages 4 and 6 of *M. tuberculosis*, and *M. bovis*, covering 80% of the coding capacity of *M. tuberculosis* H37Rv (Kohl et al. 2014, Merker et al. 2017, Kohl et al. 2018a). A refined scheme with 2891 core genes was later developed using the cgMLST definer tool of "SeqSphere+" software (Kohl et al. 2018a) with a broader set of genomes, including all MTBC lineages and isolates from animal-adapted species, such as *M. bovis*, *M. caprae*, *M. microti*, and *M. pinnipedii*, achieving 97.4% coverage compared to 94.2% in the initial scheme (Kohl et al. 2018a, O'Toole 2018). This updated scheme has been widely applied in several studies (Peker et al. 2021, Leong et al. 2022, Mekonnen et al. 2023, Quan et al. 2024; Song et al. 2024). In cgMLST, each locus is assigned a unique allele number, forming a sequence type (ST) for strain identification (Maiden et al. 2013, Jajou et al. 2019a). A threshold of more than 12 allele differences is recommended to exclude recent transmission, with allele change rates within the 2891 loci set, being comparable to SNPs (~0.5 changes/year) (Gagneux 2017, Kohl et al. 2018a). While reliable, cgMLST showed slightly lower discriminatory power in regions with low genetic diversity (Peker et al. 2021). Software tools like RIdom SeqSphere (Jünemann et al. 2013) and databases like TB Portals and GenTB facilitate standardized data analysis, enhancing TB outbreak investigations and epidemiological studies (Rosenthal et al. 2017, Gröschel et al. 2021). This makes cgMLST a powerful tool for understanding TB epidemiology, including outbreak investigations, disease control, and assessing risk factors (Maiden et al. 2013, Kohl et al. 2014, Satta et al. 2017, Jajou et al. 2019a, Jones et al. 2019, Merker et al. 2021, Mudliar et al. 2022).

WGS technology has significantly advanced the diagnosis and monitoring of drug-resistant TB, offering rapid and accurate detection of resistance-associated mutations (Papaventsis et al. 2017, Veziris et al. 2017, Walker et al. 2017, Acharya et al. 2020, Ramirez et al. 2020). Compared to traditional phenotypic tests, WGS demonstrates high sensitivity in predicting drug resistance, particularly for first-line anti-TB drugs, often exceeding 90% accuracy (Shea et al. 2017, The CRyPTIC Consortium and the 100 000 Genomes Project 2018, Jajou et al. 2019b, Wu et al. 2020). Under continuous selective antibiotic pressure on TB bacilli, the mutation rate can increase significantly, from 0.5 up to 4.3 SNPs per genome per year (Walker et al. 2013). WGS has been crucial in tracking resistance evolution (Eldholm et al. 2014, Manson et al. 2017, Jajou et al. 2019b, Li et al. 2022), predicting treatment outcomes (He et al. 2020, Katale et al. 2020), and deciphering resistance mechanisms, including for newer drugs like bedaquiline and Delamanid (Ramirez et al. 2020, Chesov et al. 2022). Supported by an ever-expanding battery of bioinformatics tools (e.g. TBprofiler, Mykrobe Predictor, CASTB, and Resistance Sniffer), WGS is a suitable method for investigating TB drug resistance (Iwai et al. 2015, Hunt et al. 2019, Phelan et al. 2019, Muzondiwa et al. 2020, Lam et al. 2021).

WGS has also markedly improved the resolution of *M. tuberculosis* strain genotyping, enhancing the ability to detect transmission clusters accurately (Meehan et al. 2019). While its wider routine use is still limited by cost, turnaround time, and re-

Box 1.

The first generation of WGS is mostly represented by Sanger's chain termination-based sequencing method (Sanger et al. 1977). This method uses dideoxynucleotides, which interrupt the elongation of DNA strands during replication, making it possible to produce reading sequences with a maximum length of a few hundred nucleotides. The ABI 370 was the first commercially available automated sequencer by Applied Biosystems Co. It used fluorescently labeled dideoxynucleotides and capillary electrophoresis to perform the sequencing automatically, as designed by Sanger.

Soon after the first application of WGS for MTB with the Sanger method (Cole et al. 1998), a need for faster and more cost-effective alternatives emerged. This led to the development of next-generation sequencing (NGS) technologies, which have surpassed Sanger sequencing by enabling rapid and simultaneous sequencing of thousands or millions of DNA fragments. These technologies differ on various parameters, such as DNA extraction methods, library preparation (including fragment size), sequencing strategy, and base-pair detection system. Based on these variables, NGS techniques are classified as second- and third-generation (Tyler et al. 2016).

Unlike traditional Sanger approach, second-generation sequencing (SGS) methods have the ability to perform massive parallel sequencing of multiple DNA fragments (Tucker et al. 2009). Numerous SGS platforms have become available, including (i) Roche's 454 sequencing method, where the sequence is determined by detecting pyrophosphate release upon nucleotide addition to the DNA template (Margulies et al. 2005), (ii) Ion Torrent sequencing by identifying hydrogen ion release during DNA synthesis (Parson et al. 2013), (iii) Illumina sequencing uses reversible dye terminators in a sequencing-by-synthesis method, with repeatedly added fluorescently labeled nucleotides to build up the reads, (iv) ABI SOLiD sequencing (sequencing by oligonucleotide ligation and detection), which employs a ligation-based approach with reversible terminators for DNA sequence determination. However, SGS methods have several drawbacks, such as the highly fragmented reads, which make reconstructing the genome difficult to perform, especially for genomes with a wide range of repeated regions, and with the GC-rich fragments poorly amplified and under-represented (Niedringhaus et al. 2011, Liu et al. 2012). Third-generation sequencing (TGS) technologies represent the latest advancements in DNA sequencing, overcoming the limitations of the previous generations. These technologies provide long-read sequencing capabilities, enabling the sequencing of much larger DNA fragments without the need of upstream PCR amplification. Among TGS platforms are PacBio sequencing, which uses a single-molecule, real-time approach with fluorescently labeled nucleotides, enabling long-read sequencing of DNA fragments of up to tens of kb in length (Rhoads and Au 2015) or Oxford Nanopore sequencing, based on nanopore technology, where a single-stranded DNA molecule passes through a nanopore, and changes in electrical current are measured to determine the DNA sequence (Lu et al. 2016). This technology is currently the most advanced in the field of sequencing and genotyping of MTBC. All WGS approaches follow the same general path depicted in Fig. 3.

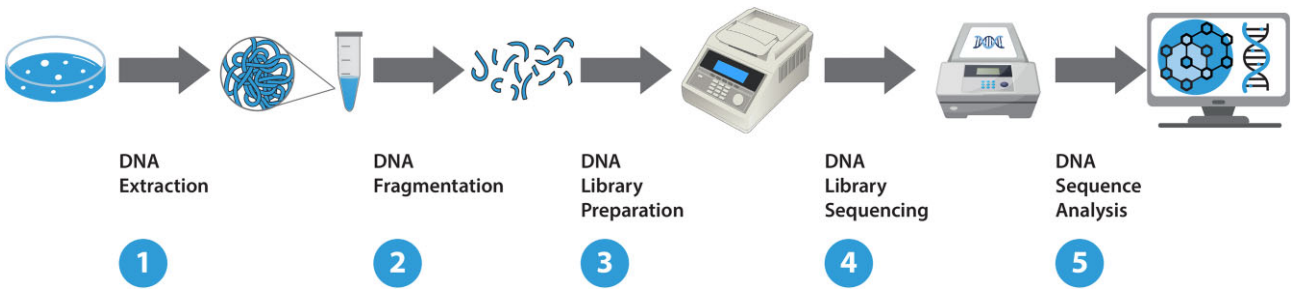

Figure 3. Key steps of the whole-genome sequencing process.

Table 1. Non-exhaustive list of software tools used for *in silico* spoligotyping and/or MIRU-VNTR typing from WGS TB.

Software tool name	Short description	Link or reference
SpolPred	Software tool used for prediction of spoligotypes from short genomic sequences	(Coll et al. 2012)
SpolPred2	An updated version of SpolPred that has been integrated into TB-Profiler	https://github.com/GaryNapier/spolpred; https://github.com/jodyphelan/TBProfiler (Napier et al. 2023)
SpoTyping	Fast and accurate *in silico Mycobacterium tuberculosis* spoligotyping from sequence reads	https://github.com/xiaeryu/SpoTyping-v2.0 (Xia et al. 2016)
Miru-Hero	Mycobacterial interspersed repetitive unit heuristics for evaluation of repeats and their ordinal	https://gitlab.com/LPCDRP/miru-hero
Galru	Long read spoligotyping for *Mycobacterium tuberculosis*	https://github.com/quadram-institute-bioscience/galru (Page et al. 2020)
MIRUReader	In-silico MIRU-VNTR typing using long reads	https://github.com/phglab/MIRUReader (Tang and Ong 2020)
MIRU-profiler	Performing digital 24-loci MIRU-VNTR typing for Mycobacterium tuberculosis	https://github.com/rahimrajwani/MIRU-profiler (Rajwani et al. 2018)
lorikeet	Digital spoligotyping of MTB strains from Illumina read data	https://github.com/AbeelLab/lorikeet (Cohen et al. 2015)
CRISPRbuilder-TB	CRISPR reconstruction based directly on short read sequences in *M. tuberculosis*	https://github.com/cguyeux/CRISPRbuilder-TB (Guyeux et al. 2021)

Table 2. Non-exhaustive list of commonly used TB databases.

Databases	Description	Reference/Link
CPLP-TB	Database aiming to to facilitate exchange of molecular epidemiological data and thus enable the tracking of important MTB clones across the Lusophone space	http://cplp-tb.ff.ulisboa.pt./ (Perdigão et al. 2019)
GMTV	Database integrating clinical, epidemiological and microbiological description with genome variations based on WGS data	(Chernyaeva et al. 2014)
Mbovis.org	Database containing *M. bovis* Spoligotyping data	https://www.mbovis.org/ (Smith and Upton 2012)
Mycobrowser	Comprehensive genomic and proteomic data repository for pathogenic mycobacteria	https://mycobrowser.epfl.ch/ (Kapopoulou et al. 2011)
MycoDB.es	Spanish Database of Animal Mycobacterosis	http://www.vigilanciasanitaria.es/mycodb/ (Rodriguez-Campos et al. 2012)
MIRU-VNTRplus	Web tool for polyphasic genotyping of MTBC bacteria	http://www.miru-vntrplus.org/ (Weniger et al. 2010)
ReSeqTB	Collaborative effort for a centralized worldwide TB relational sequencing data platform	https://www.reseqtb.org/ (Starks et al. 2015)
SITVIT2	the sixth international multimarker database for studying MTBC genetic diversity and molecular epidemiology	http://www.pasteur-guadeloupe.fr:8081/SITVIT2/ (Couvin et al. 2019)
SITVITBovis	Database and mapping tool to get an improved overview of animal and human cases caused by *Mycobacterium bovis*	http://www.pasteur-guadeloupe.fr:8081/SITVIT_Bovis/ (Couvin et al. 2022)
TB-Annotator	Pipeline and database used for the reconstruction of a global TB history	(Senelle et al. 2023)
TBDB	Repository containing the scripts and data to generate all files required to run TBProfiler	https://github.com/jodyphelan/tbdb (Phelan et al. 2019)
tbvar	*Mycobacterium tuberculosis* variome resource	(Joshi et al. 2014)
TB Portals	Web-based platform for global drug-resistant-tuberculosis data sharing and analysis	https://tbportals.niaid.nih.gov/ (Rosenthal et al. 2017)

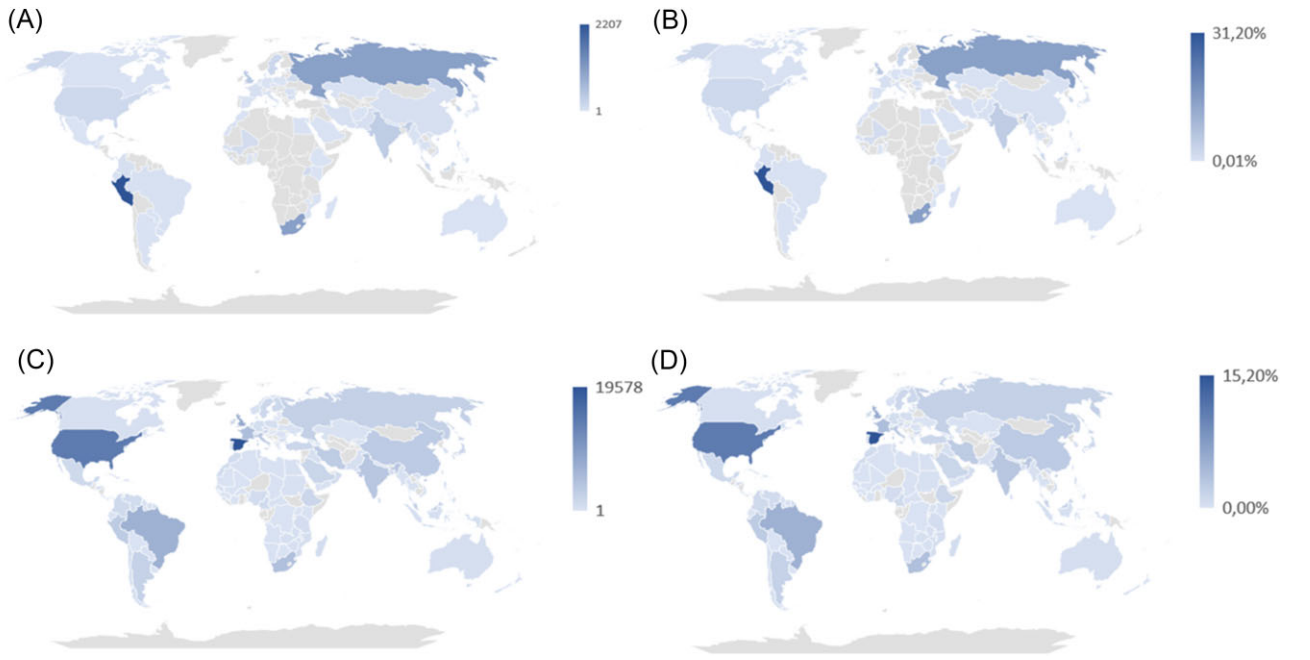

Figure 4. Intensity maps showing the distribution of RefSeq genome assemblies in terms of number (A) and percentage by country (B) contained in RefSeq repository (data collected in November 2023); and intensity maps showing the distribution of isolates contained in SITVITEXTEND database, in terms of number (C) and percentage by country (D).

quired expertise, WGS is expected to become the new gold standard for studying TB transmissions and surveillance. Ongoing technological advancements aim to increase throughput capacity while reducing complexity, potentially making WGS more accessible to smaller laboratories. However, global implementation may take longer for TB due to its concentration in resource-limited settings.

Chapter 2: an overview of software tools and databases for analyzing TB molecular data, and examples of their use

Molecular typing techniques are essential for developing TB-specific software tools and databases, as they provide detailed genetic insights into M. tuberculosis strains, including transmission

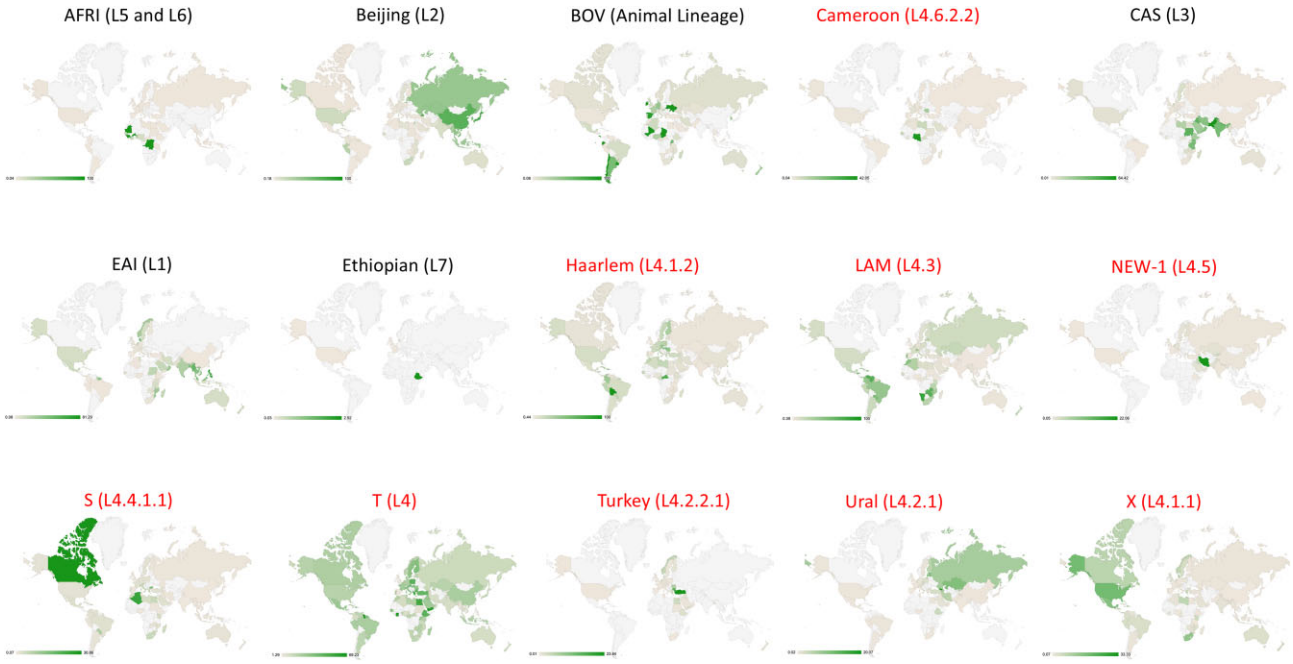

Figure 5. Distribution of main TB families (associated with SNP-based lineages/sublineages) contained in SITVITEXTEND. Families and lineages written in red represent Euro-American lineage isolates (i.e. Cameroon, Haarlem, LAM, NEW-1, S, T, Turkey, Ural and X).

Table 3. Correspondence between TB SNP-based lineages and spoligotyping families.

SNP-based lineage	Spoligotyping families
Lineage 1 (Indo-Oceanic)	East-African-Indian (EAI)
Lineage 2 (East-Asian)	Beijing
Lineage 3 (East-African-Indian)	Central Asian (CAS)
Lineage 4 (Euro-American)	Cameroon, Haarlem (H), Latin-American-Mediterranean (LAM), NEW-1, S, T, Turkey, Ural, and X
Lineage 5 (West-Africa 1)	AFRI 2 and AFRI 3
Lineage 6 (West-Africa 2)	AFRI 1
Lineage 7 (Ethiopian)	Ethiopian
Lineage 8 (African Great Lakes)	Not Defined
Lineage 9 (East Africa)	Not Defined
Lineage 10 (Central Africa)	Not Defined

routes, evolutionary patterns, and drug resistance mechanisms. These data support monitoring and surveillance efforts, requiring regular updates to software and databases to keep pace with advances in WGS methods. This section briefly highlights current tools and databases used to elucidate TB molecular epidemiology, with examples demonstrating their applications.

Software tools for analyzing TB WGS data

Various software tools and bioinformatics workflows have been developed to analyze TB molecular and WGS data, including raw sequencing reads and assembled genomes. Variant calling, a key method for SNP-based comparative genomics, helps differentiate isolates using reference genomes. These workflows aim to elucidate TB transmission, drug resistance mechanisms, and lineage prediction. Genomic data also enable broader analyses. A recent preliminary list of TB-specific tools includes MTBseq, PhyResSE, SAM-TB, TB-Profiler, TransFlow, and Mykrobe predictor TB (Couvin et al. 2021).

(i) MTBseq is an automated pipeline for mapping, variant calling, and detecting drug resistance determinants, enabling detailed phylogenetic classification of MTBC isolates from Illumina WGS data (Kohl et al. 2018b).
(ii) PhyResSE is a web tool that identifies *M. tuberculosis* lineage and drug resistance from WGS data, integrating tools like FastQC, BWA, QualiMap, SAMtools, and others for quality checks before reporting lineage and resistance patterns (Feuerriegel et al. 2015).
(iii) SAM-TB predicts MTBC drug resistance, identifies species, and assesses inter-strain genetic relatedness, including mixed samples with NTM and MTBC. It offers a user-friendly online platform (Yang et al. 2022).
(iv) TB-Profiler aligns reads to the H37Rv genome using bowtie2, BWA, or minimap2, calls variants using bcftools, and compares them to a drug-resistance database (tbdb). It supports lineage detection, spoligotyping, SNP distance computation, and more (Phelan et al. 2019).
(v) TransFlow is a modular TB transmission analysis workflow that processes raw sequencing data to infer transmission clusters, networks, and risk factors, generating summary reports with visualization (Pan et al. 2023).
(vi) Mykrobe predictor TB rapidly analyzes bacterial WGS data to predict drug resistance, requiring no expertise and operating offline on standard devices. It has been extensively validated on thousands of samples (Hunt et al. 2019).

These tools offer fast and accurate WGS-based TB analysis, as evaluated by regular assessment of performance (Morey-León et al. 2023). With the advent of artificial intelligence (AI), new software tools, such as GenTB, tend to integrate AI approaches into their algorithms to enhance prediction accuracy (Gröschel et al. 2021). As summarized in Table 1, WGS-based software tools such as *in silico* platforms for spoligotyping and/or MIRU-VNTR typing

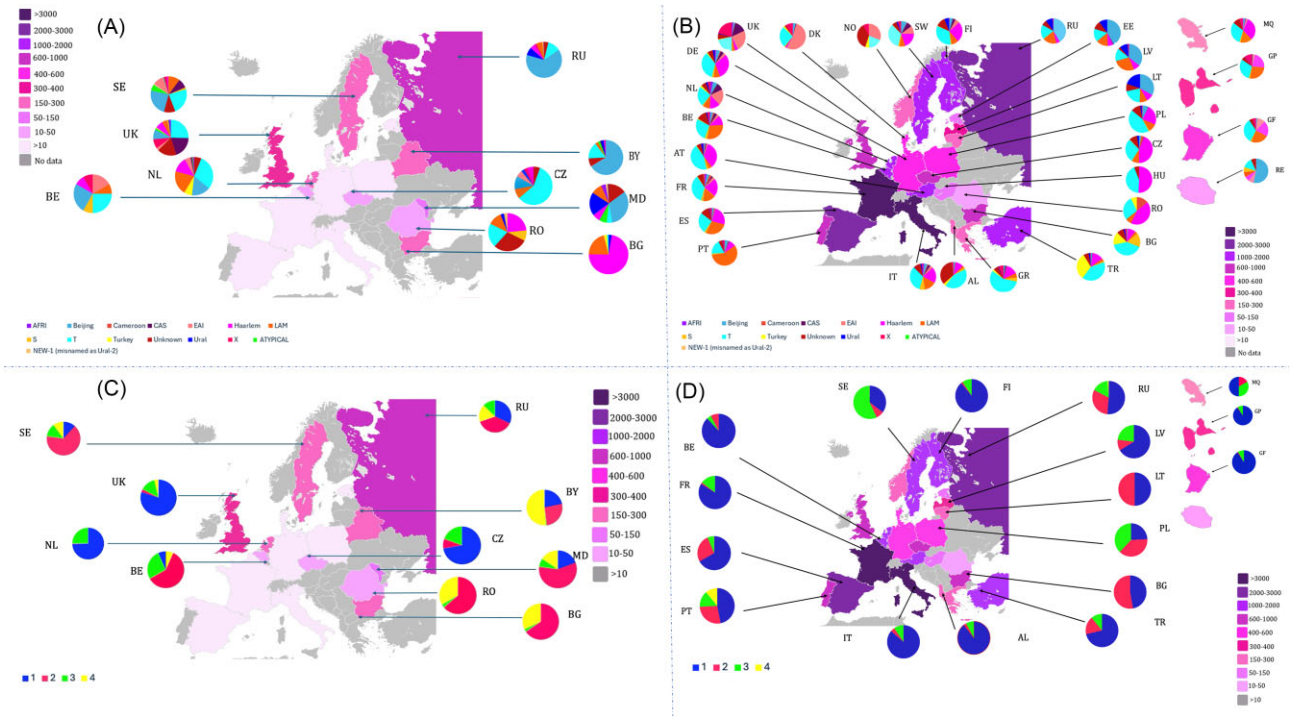

Figure 6. Maps showing TB families distribution in Europe for RefSeq repository (A) and SITVITEXTEND database (B); and for drug resistance distribution in RefSeq and SITVITEXTEND, respectively (C and D). 1–4 code numbers used in these maps (C and D) represent drug resistance profiles used in SITVIT databases. Note that two-letter country codes were used to identify countries based on ISO 3166-1 standard (https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2). Countries have been colored according to the number of isolates (the darker the color, the higher the number of isolates).

are now designed to replace classical typing methods (Morey-León et al. 2023). The selection criteria for these tools were accessibility and the ability to predict spoligotyping and/or MIRU-VNTR patterns.

In conclusion, software tools are invaluable for analyzing WGS data to identify and detect specific patterns. These tools can facilitate the creation of dedicated databases by processing output/result files or utilizing internal scripts. Additionally, specialized programs can be developed to establish automated routines that regularly update databases based on predefined criteria or rules.

Databases for analyzing TB WGS data

Databases are essential for efficiently managing and sharing large-scale data in scientific fields, including TB research. Over the past two decades, numerous TB databases have been developed, offering multi-level information on strains from global, national, or regional studies. Key databases for classical TB genotyping include MIRU-VNTRplus and SpolDB/SITVIT, the latter maintained by the Institut Pasteur de la Guadeloupe. Data entry is automated through scripts, but manual curation ensures accuracy. These databases integrate diverse data (e.g. patient demographics, drug resistance, phylogeographic, epidemiologic, genetic, and available socio-demographic data), enabling large-scale comparisons and benchmarking. Table 2 highlights examples of databases available for TB data analysis, based on their accessibility and their relationship with MTBC data. These databases are valuable for resolving TB phylogeographic diversity at local and global levels, integrating epidemiological and demographic data. Geo-mapping combined with phylogeographic analysis enables effective monitoring of clonal transmission patterns, enhancing TB surveillance and control efforts.

Mapping global circulation, transmission patterns, and TB surveillance

Global mapping data provide valuable insights for TB surveillance and analysis across diverse contexts. Figure 4 illustrates the geographical distribution of TB strains based on NCBI's RefSeq repository (Haft et al. 2024) and the SITVITEXTEND database. As of November 2023, RefSeq included 7057 genome assemblies, with distributions shown by isolate counts and percentages per country (Fig. 4A, B). Similar data from SITVITEXTEND (Fig. 4C, D) revealed disparities in genome availability, particularly in resource-limited regions such as Africa, the Caribbean, and Southeast Asia, where WGS adoption is limited. Note that the SITVIT database, with 128 000 isolates collected over 20 years through extensive collaborations, contains a significantly larger dataset.

Databases have enabled numerous studies on the distribution of TB phylogenetic lineages and families by country, region, or continent (Fig. 5). These studies highlight the geographical specificity of TB strains, with spoligotyping families in the SITVIT databases often used alongside SNP-based lineages to assess phylogeographic patterns. Table 3 shows the correspondence between spoligotyping families and SNP-based lineages (SNP barcode nomenclature). Notably, recently discovered lineages (Lineage 8, Lineage 9, and Lineage 10), clearly appear to be restricted to Africa (Ngabonziza et al. 2020, Coscolla et al. 2021, Guyeux et al. 2024).

A phylogeographic study of 21 574 TB strains (excluding *M. bovis*) from SITVITEXTEND revealed disparities across European countries. These strains were isolated from year 1890 to 2021 from

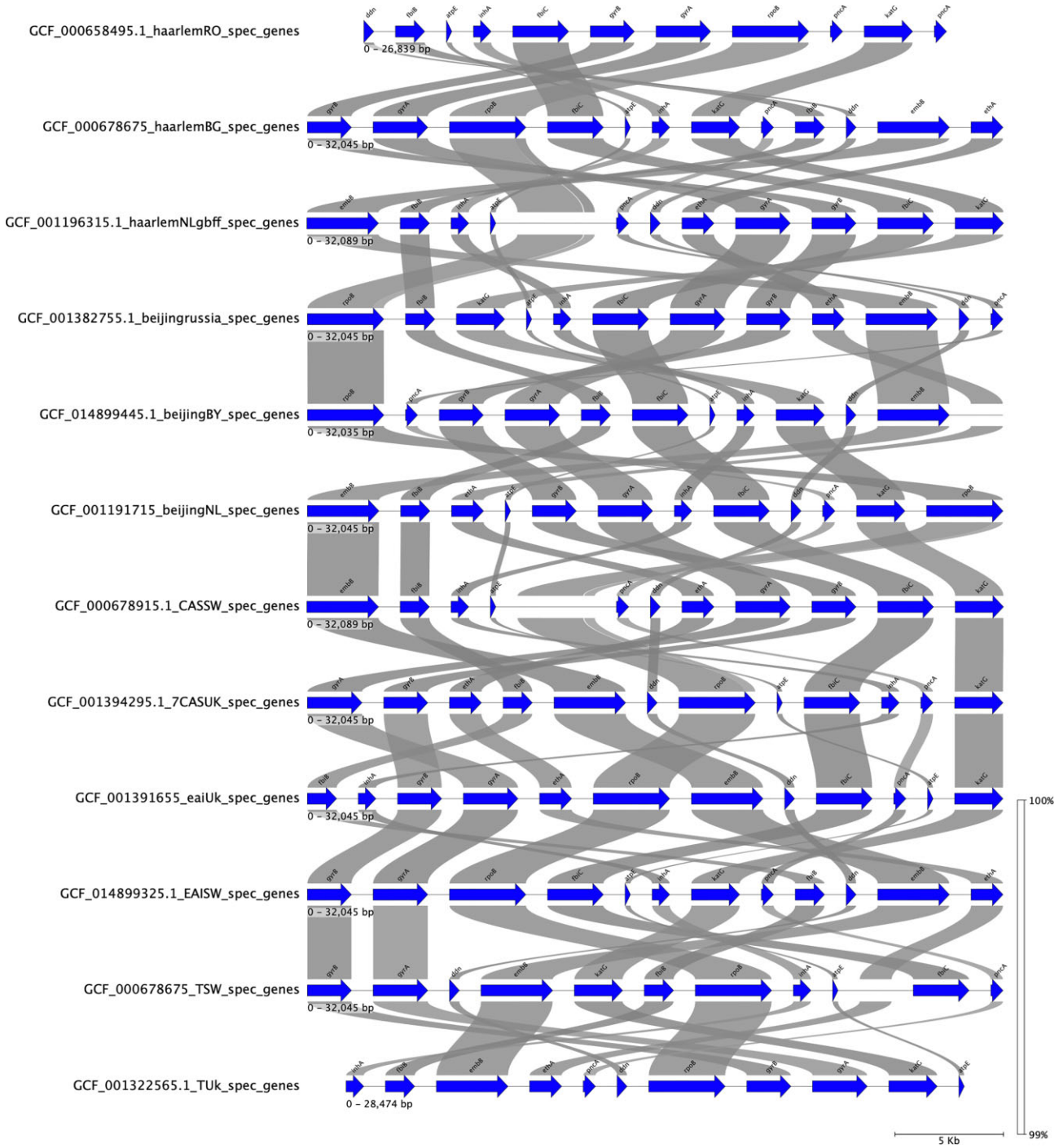

Figure 7. Synteny map showing the distribution and similarity between various drug resistance genes (e.g. rpoB, katG, ethA, inhA, pncA, embB, rrs, and gyrA) recovered from selected RefSeq genome assemblies.

23 countries (Albania AL, $n = 237$; Austria AT, $n = 1575$; Belgium BE, $n = 1369$; Bulgaria BG, $n = 639$; Czech Republic CZ, $n = 637$; Denmark DK, $n = 550$; Estonia EE, $n = 119$; Finland FI, $n = 1427$; France FR, $n = 3509$; Germany DE, $n = 455$; Greece GR, $n = 170$; Hungary HU, $n = 65$; Italy IT, $n = 3191$; Latvia LV, $n = 363$; Lithuania LT, $n = 200$; Netherlands NL, $n = 1355$; Norway NO, $n = 89$; Poland PL, $n = 523$; Portugal PT, $n = 722$; Romania RO, $n = 14$; Spain ES, $n = 2056$; Sweden SE, $n = 1409$; United Kingdom GB, $n = 900$). Genomic data from RefSeq and SITVITEXTEND were analyzed using simpiTB, Miru-Hero, SpolLineages, and TB-Profiler to infer spoligotyping families and drug resistance profiles. Comparative analyses showed that four major spoligotyping families (T, Beijing, Haarlem, and LAM) were widespread, though proportions varied. Beijing lineage strains were predominant in eastern European countries, such as Russia, Estonia, Lithuania, Latvia, Belarus, and Moldova (Fig. 6A, B), and accounted for nearly 70% of strains in Russia and Belarus. Conversely, some countries in Northern and Western Europe, such as Sweden, the UK, the Netherlands, and Belgium, showed a distinct and heterogeneous distribution of *M. tuberculosis* families. Nevertheless, the proportion of the T family remains relatively constant in these countries, accounting for around 25% of MTBC isolates. In French overseas territo-

ries, the MTBC family distribution in Guadeloupe, Martinique, and French Guiana resembled Western Europe, while Reunion Island displayed a higher prevalence of Beijing strains (Fig. 6B).

Drug resistance was categorized using SITVITEXTEND codes:

Code 1: Pan-susceptible strains.
Code 2: MDR-TB (resistance to INH and RIF, ± other drugs).
Code 3: Resistance to other drugs.
Code 4: XDR-TB (MDR-TB + fluoroquinolone + any 1 of 3 injectable 2nd-line drugs (capreomycin, kanamycin, amikacin). Note that newer drugs have been introduced to treat TB disease for managing resistant strains, which are not shown in this study.

Note that a significant proportion of MDR-TB and XDR-TB strains was found mainly in the Eastern European countries (both in RefSeq and STIVITEXTEND databases), such as Russia, Latvia, Poland, Bulgaria, Moldova, and Romania. On the other hand, higher proportions of pansusceptible strains were observed in France, Italy, Albania, and Finland (Fig. 6C, D). In order to harmonize the data recorded, it would be recommendable to carry out studies on European countries for which we have little or no data such as Slovakia, Slovenia, and Switzerland.

Antibiotic resistance, particularly MDR and XDR TB, is a global public health issue affecting the management of TB. New strategies, including the use of specific software tools and databases, are essential for studying drug-resistant tuberculosis and improving data visualization. These tools help answer broader microbiological questions on TB transmission, drug resistance, virulence, and epidemiology. At the Institut Pasteur de la Guadeloupe, tools like getSequenceInfo and "getGenesFromGenBank.py" (https://github.com/karubiotools/getSequenceInfo) enable the extraction of specific genes from genome assemblies to study resistance gene acquisition across TB families (Moco et al. 2022). To compare drug resistance-associated genes (e.g. *rpoB*, *katG*, *ethA*, *inhA*, *pncA*, *embB*, *rrs*, *gyrA*, etc.; Fig. 7) across genomes, pyGenomeViz (https://github.com/moshi4/pyGenomeViz) was used to construct a synteny map. Developments are underway to make these tools accessible to the public. Alternatives also exist for interrogating a specific region of the genome, notably with tools such as GenBank (https://www.ncbi.nlm.nih.gov/genbank/). These analyses show high similarity (>99%) among genes, though their positions in genomes are unstable. Despite this, *M. tuberculosis* genomes exhibit high conservation. Integrating such analyses into pipelines and databases can provide global genomic insights and aid in deciphering drug resistance profiles.

Conclusions and perspectives

In conclusion, molecular typing tools are crucial for studying TB epidemiology and evolution. Over the past three decades, various molecular methods have emerged, from IS*6110*-RFLP, once the gold standard, to more accessible techniques like spoligotyping and MIRU-VNTR typing, particularly in low-income countries. Currently, WGS is becoming the new gold standard for in-depth *M. tuberculosis* genome analysis. Despite this, no single typing system is ideal due to technical limitations and feasibility challenges, particularly in implementing WGS.

Recent advancements in TB-focused software and databases offer valuable insights into TB genomics, correlating data with geographical, demographic, and epidemiological information. However, further methods are needed to better analyze and extract meaningful knowledge from vast genomic data. Combining classical genotyping with WGS can provide a more comprehensive understanding of TB molecular epidemiology. Additionally, developing cost-effective, accessible WGS tools will help expand research in low-income settings. Future research will likely focus on leveraging AI to enhance the depth, accuracy, and efficiency of TB WGS data analysis.

Acknowledgments

We thank Damien Cazenave for his help in the development of getSequenceInfo suite of tools. We are also grateful to various contributors who kindly provided genotyping data to the SpolDB/SITVIT databases as well as all the users of these tools.

Conflict of interest: None declared.

Funding

This review was supported by a grant from the National Science Centre, Poland, under contract number 2017/27/L/NZ6/03279.

References

Acharya B, Acharya A, Gautam S *et al.* Advances in diagnosis of tuberculosis: an update into molecular diagnosis of *Mycobacterium tuberculosis*. *Mol Biol Rep* 2020;**47**:4065–75. https://doi.org/10.1007/s11033-020-05413-7.

Afshar B, Carless J, Roche A *et al.* Surveillance of tuberculosis (TB) cases attributable to relapse or reinfection in London, 2002–2015. *PLoS One* 2019;**14**:e0211972. https://doi.org/10.1371/journal.pone.0211972.

Ahmad F, Alam A, Kumari I *et al.* Next generation sequencing: opportunities and challenges in tuberculosis research. In: Hameed S, Fatima Z (eds), *Integrated Omics Approaches to Infectious Diseases*. Singapore: Springer, 2021. https://doi.org/10.1007/978-981-16-0691-5_2.

Alaridah N, Hallbäck ET, Tångrot J *et al.* Transmission dynamics study of tuberculosis isolates with whole genome sequencing in southern Sweden. *Sci Rep* 2019;**9**:4931. https://doi.org/10.1038/s41598-019-39971-z.

Allix-Béguec C, Fauville-Dufaux M, Supply P. Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 2008;**46**:1398–406. https://doi.org/10.1128/JCM.02089-07.

Allix-Be´guec C, Harmsen D, Weniger T *et al.* Evaluation and strategy for use of MIRU-VNTR *plus*, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. *J Clin Microbiol* 2008;**46**:2692–9. https://doi.org/10.1128/JCM.00540-08.

Allix-Béguec C, Wahl C, Hanekom M *et al.* Proposal of a consensus set of hypervariable mycobacterial interspersed repetitive-unit-variable-number tandem-repeat loci for subtyping of *Mycobacterium tuberculosis* Beijing isolates. *J Clin Microbiol* 2014;**52** 164–72. https://doi.org/10.1128/JCM.02519-13.

Ani A, Bruvik T, Okoh Y *et al.* Genetic diversity of *Mycobacterium tuberculosis* Complex in Jos, Nigeria. *BMC Infect Dis* 2010;**10**:189. https://doi.org/10.1186/1471-2334-10-189.

Augustynowicz-Kopeć E, Jagielski T, Zwolska Z. Genetic diversity of isoniazid-resistant *Mycobacterium tuberculosis* isolates collected in Poland and assessed by spoligotyping. *J Clin Microbiol* 2008;**46**:4041–4. https://doi.org/10.1128/JCM.01315-08.

Bakuła Z, Dziurzyński M, Decewicz P *et al.* Spoligotyping of *Mycobacterium tuberculosis*—comparing in vitro and in silico approaches.

Infect Genet Evol 2023a;**115**:105508. https://doi.org/10.1016/j.meegid.2023.105508.

Bakuła Z, Marczak M, Bluszcz A *et al.* Phylogenetic relationships of *Mycobacterium tuberculosis* isolates in Poland: the emergence of Beijing genotype among multidrug-resistant cases. *Front Cell Infect Microbiol* 2023b;**13**:1161905. https://doi.org/10.3389/fcimb.2023.1161905.

Bakuła Z, Javed H, Pleń M *et al.* Genetic diversity of multidrug-resistant *Mycobacterium tuberculosis* isolates in Punjab, Pakistan. *Infect Genet Evol* 2019;**72**:16–24. https://doi.org/10.1016/j.meegid.2019.02.029.

Bellad R, Nagamoti M, Sharma P *et al.* Spoligotyping of *Mycobacterium tuberculosis* isolates from pulmonary tuberculosis patients from North Karnataka, India. *Trop Doct* 2022;**52**:386–90. https://doi.org/10.1177/00494755221080584.

Bespyatykh JA, Zimenkov DV, Shitikov EA *et al.* Spoligotyping of *Mycobacterium tuberculosis* complex isolates using hydrogel oligonucleotide microarrays. *Infect Genet Evol* 2014;**26**:41–46. https://doi.org/10.1016/j.meegid.2014.04.024.

Bidovec-Stojkovic U, Zolnir-Dovc M, Supply P. One year nationwide evaluation of 24-locus MIRU-VNTR genotyping on Slovenian *Mycobacterium tuberculosis* isolates. *Respir Med* 2011;**105**:S67–73. https://doi.org/10.1016/S0954-6111(11)70014-2.

Bifani P, Kurepina N, Mathema B *et al.* Genotyping of *Mycobacterium tuberculosis* clinical isolates using IS6110-based restriction fragment length polymorphism analysis. *Methods Mol Biol* 2009;**551**:173–88. https://doi.org/10.1007/978-1-60327-999-4_14.

Blackwood KS, Wolfe JN, Kabani AM. Application of mycobacterial interspersed repetitive unit typing to Manitoba tuberculosis cases: can restriction fragment length polymorphism be forgotten? *J Clin Microbiol* 2004;**42**:5001–6. https://doi.org/10.1128/JCM.42.11.5001-5006.2004.

Bogaerts B, Delcourt T, Soetaert K *et al.* A bioinformatics whole-genome sequencing workflow for clinical *Mycobacterium tuberculosis* complex isolate analysis, validated using a reference collection extensively characterized with conventional methods and in silico approaches. *J Clin Microbiol* 2021;**59**:e00202–21. https://doi.org/10.1128/JCM.00202-21.

Bouklata N, Supply P, Jaouhari S *et al.* Molecular typing of *Mycobacterium tuberculosis* complex by 24-locus based MIRU-VNTR typing in conjunction with spoligotyping to assess genetic diversity of strains circulating in Morocco. *PLoS One* 2015;**10**:e0135695. https://doi.org/10.1371/journal.pone.0135695.

Braden CR, Crawford JT, Schable BA. Quality assessment of *Mycobacterium tuberculosis* genotyping in a large laboratory network. *Emerg Infect Dis* 2002;**8**:1210–5. https://doi.org/10.3201/eid0811.020401.

Brudey K, Driscoll JR, Rigouts L *et al.* *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol* 2006;**6**:23. https://doi.org/10.1186/1471-2180-6-23.

Bryant JM, Harris SR, Parkhill J *et al.* Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study. *Lancet Respir Med* 2013;**1**:786–92. https://doi.org/10.1016/S2213-2600(13)70231-5.

Cancino-Muñoz I, López MG, Torres-Puente M *et al.* Population-based sequencing of *Mycobacterium tuberculosis* reveals how current population dynamics are shaped by past epidemics. *eLife* 2022;**11**:e76605. https://doi.org/10.7554/ELIFE.76605.

Casali N, Broda A, Harris SR *et al.* Whole genome sequence analysis of a large isoniazid-resistant tuberculosis outbreak in London: a retrospective observational study. *PLoS Med* 2016;**13**:e1002137. https://doi.org/10.1371/journal.pmed.1002137.

Chaoui I, Zozio T, Lahlou O *et al.* Contribution of spoligotyping and MIRU-VNTRs to characterize prevalent *Mycobacterium tuberculosis* genotypes infecting tuberculosis patients in Morocco. *Infect Genet Evol* 2014;**21**:463–71. https://doi.org/10.1016/j.meegid.2013.05.023.

Chauhan DS, Sharma VD, Parashar D *et al.* Molecular typing of *Mycobacterium tuberculosis* isolates from different parts of India based on IS6110 element polymorphism using RFLP analysis. *Indian J Med Res* 2007;**125**:577–81.

Chen J, Chen L, Zhou M *et al.* Transmission of multidrug-resistant tuberculosis within family households by DTM-PCR and MIRU-VNTR genotyping. *BMC Infect Dis* 2022;**22**:192. https://doi.org/10.1186/s12879-022-07188-7.

Chernyaeva EN, Shulgina MV, Rotkevich MS *et al.* Genome-wide *Mycobacterium tuberculosis* variation (GMTV) database: a new tool for integrating sequence variations and epidemiology. *BMC Genomics* 2014;**15**:308. https://doi.org/10.1186/1471-2164-15-308.

Chesov E, Chesov D, Maurer FP *et al.* Emergence of bedaquiline resistance in a high tuberculosis burden country. *Eur Respir J* 2022;**59**:2100621. https://doi.org/10.1183/13993003.00621-2021.

Chisompola NK, Streicher EM, Dippenaar A *et al.* Drug resistant tuberculosis cases from the Copperbelt province and Northern regions of Zambia: genetic diversity, demographic and clinical characteristics. *Tuberculosis* 2021;**130**:102122. https://doi.org/10.1016/j.tube.2021.102122.

Clark CM, Driver CR, Munsiff SS *et al.* Universal genotyping in Tuberculosis Control Program, New York City, 2001–2003. *Emerg Infect Dis* 2006;**12**:719–24. https://doi.org/10.3201/eid1205.050446.

Cohen KA, Abeel T, Manson McGuire A *et al.* Evolution of extensively drug-resistant tuberculosis over four decades: whole genome sequencing and dating analysis of *Mycobacterium tuberculosis* isolates from KwaZulu-Natal. *PLoS Med* 2015;**12**:e1001880. https://doi.org/10.1371/journal.pmed.1001880.

Cole S, Brosch R, Parkhill J *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998;**396**:190.

Coll F, Mallard K, Preston MD *et al.* SpolPred: rapid and accurate prediction of *Mycobacterium tuberculosis* spoligotypes from short genomic sequences. *Bioinformatics* 2012;**28**:2991–3. https://doi.org/10.1093/bioinformatics/bts544.

Comas I, Homolka S, Niemann S *et al.* Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS One* 2009;**4**:e7815. https://doi.org/10.1371/journal.pone.0007815.

Coscolla M, Gagneux S, Menardo F *et al.* Phylogenomics of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. *Microb Genom* 2021;**7**:000477. https://doi.org/10.1099/mgen.0.000477.

Couvin D, Cervera-Marzal I, David A *et al.* SITVITBovis—a publicly available database and mapping tool to get an improved overview of animal and human cases caused by *Mycobacterium bovis*. *Database* 2022;**2022**:baab081. https://doi.org/10.1093/database/baab081.

Couvin D, David A, Zozio T *et al.* Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the *Mycobacterium tuberculosis* genotyping database. *Infect Genet Evol* 2019;**72**:31–43. https://doi.org/10.1016/j.meegid.2018.12.030.

Couvin D, Reynaud Y, Rastogi N. MTBCtools: a non-exhaustive list of software tools/resources for bioinformatics analyses of *Mycobacterium tuberculosis* complex, the causative agent of tuberculosis. *Int J Mycobacteriol* 2021;**10**:S18. https://doi.org/10.4103/2212-5531.307066.

Couvin D, Segretier W, Stattner E *et al.* Novel methods included in SpolLineages tool for fast and precise prediction of *Mycobacterium tuberculosis* complex spoligotype families. *Database* 2020;**2020**:baaa108. https://doi.org/10.1093/database/baaa108.

Couvin D, Zozio T, Rastogi N. SpolSimilaritySearch—a web tool to compare and search similarities between spoligotypes of *Mycobacterium tuberculosis* complex. *Tuberculosis* 2017;**105**:49–52. https://doi.org/10.1016/j.tube.2017.04.007.

Cowan LS, Diem L, Brake MC *et al.* Transfer of a *Mycobacterium tuberculosis* genotyping method, spoligotyping, from a reverse line-blot hybridization, membrane-based assay to the Luminex multianalyte profiling system. *J Clin Microbiol* 2004;**42**:474–7. https://doi.org/10.1128/JCM.42.1.474-477.2004.

Cowan LS, Diem L, Monson T *et al.* Evaluation of a two-step approach for large-scale, prospective genotyping of *Mycobacterium tuberculosis* isolates in the United States. *J Clin Microbiol* 2005;**43**:688–95. https://doi.org/10.1128/JCM.43.2.688-695.2005.

Cowan LS, Mosher L, Diem L *et al.* Variable-number tandem repeat typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS *6110* by using mycobacterial interspersed repetitive units. *J Clin Microbiol* 2002;**40**:1592–602. https://doi.org/10.1128/JCM.40.5.1592-1602.2002.

Crawford JT, Braden CR, Schable BA *et al.* National tuberculosis genotyping and surveillance network: design and methods. *Emerg Infect Dis* 2002;**8**:1192–6. https://doi.org/10.3201/eid0811.020296.

Cronin WA, Golub JE, Magder LS *et al.* Epidemiologic usefulness of spoligotyping for secondary typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS *6110*. *J Clin Microbiol* 2001;**39**:3709–11. https://doi.org/10.1128/JCM.39.10.3709-3711.2001.

Das SD, Narayanan S, Hari L *et al.* Differentiation of highly prevalent IS6110 single-copy strains of *Mycobacterium tuberculosis* from a rural community in South India with an ongoing DOTS programme. *Infect Genet Evol* 2005;**5**:67–77. https://doi.org/10.1016/j.meegid.2004.06.007.

de Boer AS, Borgdorff MW, de Haas PEW *et al.* Analysis of rate of change of IS *6110* RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates. *J Infect Dis* 1999;**180**:1238–44. https://doi.org/10.1086/314979.

Dekhil N, Skhairia MA, Mhenni B *et al.* Automated IS6110-based fingerprinting of *Mycobacterium tuberculosis*: reaching unprecedented discriminatory power and versatility. *PLoS One* 2018;**13**:e0197913. https://doi.org/10.1371/journal.pone.0197913.

De La Salmonière YOG, Li HM, Torrea G *et al.* Evaluation of spoligotyping in a study of the transmission of *Mycobacterium tuberculosis*. *J Clin Microbiol* 1997;**35**:2210–4. https://doi.org/10.1128/jcm.35.9.2210-2214.1997.

Demay C, Liens B, Burguière T *et al.* SITVITWEB—a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect Genet Evol* 2012;**12**:755–66. https://doi.org/10.1016/j.meegid.2012.02.004.

Devaux I, Kremer K, Heersma H *et al.* Clusters of multidrug-resistant *Mycobacterium tuberculosis* cases, Europe. *Emerg Infect Dis* 2009;**15**:1052–60. https://doi.org/10.3201/eid1507.080994.

Diaz R, Gomez RI, Restrepo E *et al.* Transmission of tuberculosis in Havana, Cuba: a molecular epidemiological study by IS6110 restriction fragment length polymorphism typing. *Mem Inst Oswaldo Cruz* 2001;**96**:437–43. https://doi.org/10.1590/S0074-02762001000400001.

Diaz R, Kremer K, De Haas PEW *et al.* Molecular epidemiology of tuberculosis in Cuba outside of Havana, July 1994–June 1995: utility of spoligotyping versus IS6110 restriction fragment length polymorphism. *Int J Tuberc Lung Dis* 1998;**2**:743–50.

Diel R, Meywald-Walter K, Gottschalk R *et al.* Ongoing outbreak of tuberculosis in a low-incidence community: a molecular-epidemiological evaluation. *Int J Tuberc Lung Dis* 2004;**8**:855–61.

Dong H, Liu Z, Lv B *et al.* Spoligotypes of *Mycobacterium tuberculosis* from different provinces of China. *J Clin Microbiol* 2010;**48**:4102–6. https://doi.org/10.1128/JCM.00549-10.

Dookie N, Rambaran S, Padayatchi N *et al.* Evolution of drug resistance in *Mycobacterium tuberculosis*: a review on the molecular determinants of resistance and implications for personalized care. *J Antimicrob Chemother* 2018;**73**:1138–51. https://doi.org/10.1093/jac/dkx506.

Ei PW, Lee JS, Aung WW *et al.* Genotypes and genetic characters of *Mycobacterium tuberculosis* from Myanmar using three typing methods. *Infect Genet Evol* 2019;**75**:104005. https://doi.org/10.1016/j.meegid.2019.104005.

Eldholm V, Matee M, Mfinanga SGM *et al.* A first insight into the genetic diversity of *Mycobacterium tuberculosis* in Dar es Salaam, Tanzania, assessed by spoligotyping. *BMC Microbiol* 2006;**6**:76. https://doi.org/10.1186/1471-2180-6-76.

Eldholm V, Norheim G, von der Lippe B *et al.* Evolution of extensively drug-resistant *Mycobacterium tuberculosis* from a susceptible ancestor in a single patient. *Genome Biol* 2014;**15**:490. https://doi.org/10.1186/s13059-014-0490-3.

Elegail A, Ibrahim Mohamed NY, Mohamed Nour EO *et al.* Molecular characterization of *Mycobacterium tuberculosis* isolates from pulmonary tuberculosis patients in Khartoum, Sudan. *Int J Mycobacteriol* 2018;**7**:236. https://doi.org/10.4103/ijmy.ijmy_82_18.

Essahale A, Nia F, Sfendla A *et al.* Finger printing-RFLP analysis of chromosomal IS6110 insertion sequence and PCR diagnosis of pulmonary tuberculosis, isolated from patients in El Hajeb region of Morocco. *Indian J Tuberc* 2024;**71**:117–22. https://doi.org/10.1016/j.ijtb.2023.03.020.

Fang Z, Morrison N, Watt B *et al.* IS *6110* transposition and evolutionary scenario of the direct repeat locus in a group of closely related *Mycobacterium tuberculosis* strains. *J Bacteriol* 1998;**180**:2102–9. https://doi.org/10.1128/JB.180.8.2102-2109.1998.

Farmer PE, Nizeye B, Stulac S *et al.* Structural violence and clinical medicine. *PLoS Med* 2006;**3**:e449. https://doi.org/10.1371/journal.pmed.0030449.

Feuerriegel S, Schleusener V, Beckert P *et al.* PhyResSE: a web tool delineating *Mycobacterium tuberculosis* antibiotic resistance and lineage from whole-genome sequencing data. *J Clin Microbiol* 2015;**53**:1908–14. https://doi.org/10.1128/JCM.00025-15.

Filliol I, Driscoll JR, van Soolingen D *et al.* Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. *J Clin Microbiol* 2003;**41**:1963–70. https://doi.org/10.1128/JCM.41.5.1963-1970.2003.

Fox GJ, Barry SE, Britton WJ *et al.* Contact investigation for tuberculosis: a systematic review and meta-analysis. *Eur Respir J* 2013;**41**:140–56. https://doi.org/10.1183/09031936.00070812.

Friedman CR, Stoeckle MY, Johnson WD *et al.* Double-repetitive-element PCR method for subtyping *Mycobacterium tuberculosis* clinical isolates. *J Clin Microbiol* 1995;**33**: 1383–4. https://doi.org/10.1128/jcm.33.5.1383-1384.1995.

Frothingham R, Meeker-O'Connell WA. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology* 1998;**144**:1189–96. https://doi.org/10.1099/00221287-144-5-1189.

Gagneux S. Gagneux S. (ed.), *Strain Variation in the Mycobacterium Tuberculosis Complex: Its Role in Biology, Epidemiology and Control*. vol.

1019. Cham: Springer. 2017. https://doi.org/10.1007/978-3-319-64371-7.

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

Gardy JL, Johnston JC, Sui SJH *et al.* Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 2011;**364**:730–9. https://doi.org/10.1056/nejmoa1003176.

Gautam SS, Aogáin M Mac, Cooley LA *et al.* Molecular epidemiology of tuberculosis in Tasmania and genomic characterisation of its first known multi-drug resistant case. *PLoS One* 2018;**13**:e0192351. https://doi.org/10.1371/journal.pone.0192351.

Genestet C, Hodille E, Bernard A *et al.* Consistency of *Mycobacterium tuberculosis* complex spoligotyping between the membrane-based method and in silico approach. *Microbiol Spectr* 2022;**10**:e0022322. https://doi.org/10.1128/spectrum.00223-22.

Gomgnimbou MK, Abadia E, Zhang J *et al.* "Spoligoriftyping," a dual-priming-oligonucleotide-based direct-hybridization assay for tuberculosis control with a multianalyte microbead-based hybridization system. *J Clin Microbiol* 2012;**50**:3172–9. https://doi.org/10.1128/JCM.00976-12.

Gopaul KK, Brown TJ, Gibson AL *et al.* Progression toward an improved DNA amplification-based typing technique in the study of *Mycobacterium tuberculosis* epidemiology. *J Clin Microbiol* 2006;**44**:2492–8. https://doi.org/10.1128/JCM.01428-05.

Goyal M, Saunders NA, Van Embden JDA *et al.* Differentiation of *Mycobacterium tuberculosis* isolates by spoligotyping and IS6110 restriction fragment length polymorphism. *J Clin Microbiol* 1997;**35**:647–51. https://doi.org/10.1128/jcm.35.3.647-651.1997.

Groenen PMA, Bunschoten AE, Soolingen D van. *et al.* Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol Microbiol* 1993;**10**:1057–65. https://doi.org/10.1111/j.1365-2958.1993.tb00976.x.

Groenheit R, Ghebremichael S, Svensson J *et al.* The Guinea-Bissau family of *Mycobacterium tuberculosis* complex revisited. *PLoS One* 2011;**6**:e18601. https://doi.org/10.1371/journal.pone.0018601.

Gröschel MI, Owens M, Freschi L *et al.* GenTB: a user-friendly genome-based predictor for tuberculosis resistance powered by machine learning. *Genome Med* 2021;**13**:138. https://doi.org/10.1186/s13073-021-00953-4.

Gurjav U, Outhred AC, Jelfs P *et al.* Whole genome sequencing demonstrates limited transmission within identified *Mycobacterium tuberculosis* clusters in New South Wales, Australia. *PLoS One* 2016;**11**:e0163612. https://doi.org/10.1371/journal.pone.0163612.

Guyeux C, Senelle G, Le Meur A *et al.* Newly identified *Mycobacterium africanum* lineage 10, Central Africa. *Emerg Infect Dis* 2024;**30**:560–63. https://doi.org/10.3201/eid3003.231466.

Guyeux C, Sola C, Noûs C *et al.* CRISPRbuilder-TB: "CRISPR-builder for tuberculosis". Exhaustive reconstruction of the CRISPR locus in *Mycobacterium tuberculosis* complex using SRA. *PLoS Comput Biol* 2021;**17**:e1008500. https://doi.org/10.1371/journal.pcbi.1008500.

Haft DH, Badretdin A, Coulouris G *et al.* RefSeq and the prokaryotic genome annotation pipeline in the age of metagenomes. *Nucleic Acids Res* 2024;**52**:D762–9. https://doi.org/10.1093/nar/gkad988.

Hatherell HA, Colijn C, Stagg HR *et al.* Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. *BMC Med* 2016;**14**:21. https://doi.org/10.1186/s12916-016-0566-x.

He G, Li Y, Chen X *et al.* Prediction of treatment outcomes for multidrug-resistant tuberculosis by whole-genome sequencing. *Int J Infect Dis* 2020;**96**:68–72. https://doi.org/10.1016/j.ijid.2020.04.043.

Heersma HF, Kremer K, van Embden JDA. Computer analysis of IS6110 RFLP patterns of *Mycobacterium tuberculosis*. In: Parish T, Stoker NG (eds), *Mycobacteria Protocols*. Humana Press, 1998, 395–422. https://doi.org/10.1385/0-89603-471-2:395.

Hermans PWM, Van Soolingen D, Bik EM *et al.* Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect Immun* 1991;**59**:2695–705. https://doi.org/10.1128/iai.59.8.2695-2705.1991.

Heyderman RS, Goyal M, Roberts P *et al.* Pulmonary tuberculosis in harare, zimbabwe: analysis by spoligotyping. *Thorax* 1998;**53**:346–50. https://doi.org/10.1136/thx.53.5.346.

Hijikata M, Keicho N, Duc L *et al.* Spoligotyping and whole-genome sequencing analysis of lineage 1 strains of *Mycobacterium tuberculosis* in Da Nang, Vietnam. *PLoS One* 2017;**12**:e0186800. https://doi.org/10.1371/journal.pone.0186800.

Honisch C, Mosko M, Arnold C *et al.* Replacing reverse line blot hybridization spoligotyping of the *Mycobacterium tuberculosis* complex. *J Clin Microbiol* 2010;**48**:1520–6. https://doi.org/10.1128/JCM.02299-09.

Hunt M, Bradley P, Lapierre SG *et al.* Antibiotic resistance prediction for *Mycobacterium tuberculosis* from genome sequence data with mykrobe. *Wellcome Open Res* 2019;**4**:191. https://doi.org/10.12688/wellcomeopenres.15603.1.

Hussien B, Zewude A, Wondale B *et al.* Spoligotyping of clinical isolates of *Mycobacterium tuberculosis* complex species in the Oromia Region of Ethiopia. *Front Public Health* 2022;**10**:808626. https://doi.org/10.3389/fpubh.2022.808626.

Iwai H, Kato-Miyazawa M, Kirikae T *et al.* CASTB (the comprehensive analysis server for the *Mycobacterium tuberculosis* complex): a publicly accessible web server for epidemiological analyses, drug-resistance prediction and phylogenetic comparison of clinical isolates. *Tuberculosis* 2015;**95**:843–44. https://doi.org/10.1016/j.tube.2015.09.002.

Iwamoto T, Arikawa K, Murase Y *et al.* Transmission dynamics variability of lineage 2 *Mycobacterium tuberculosis* strains in Kobe, Japan, determined using population-based whole-genome sequencing analysis. *Infect Genet Evol* 2023;**114**:105495. https://doi.org/10.1016/j.meegid.2023.105495.

Iwamoto T, Yoshida S, Suzuki K *et al.* Hypervariable loci that enhance the discriminatory ability of newly proposed 15-loci and 24-loci variable-number tandem repeat typing method on *Mycobacterium tuberculosis* strains predominated by the Beijing family. *FEMS Microbiol Lett* 2007;**270**:67–74. https://doi.org/10.1111/j.1574-6968.2007.00658.x.

Jagielski T, Brzostek A, van Belkum A *et al.* A close-up on the epidemiology and transmission of multidrug-resistant tuberculosis in Poland. *Eur J Clin Microbiol Infect Dis* 2015;**34**:41–53. https://doi.org/10.1007/s10096-014-2202-z.

Jagielski T, Minias A, van Ingen J *et al.* Methodological and clinical aspects of the molecular epidemiology of *Mycobacterium tuberculosis* and other mycobacteria. *Clin Microbiol Rev* 2016;**29**:239–90. https://doi.org/10.1128/CMR.00055-15.

Jagielski T, Van Ingen J, Rastogi N *et al.* Current methods in the molecular typing of *Mycobacterium tuberculosis* and other mycobacteria. *BioMed Res Int* 2014;**2014**:645802. https://doi.org/10.1155/2014/645802.

Jajou R, De Neeling A, Van Hunen R *et al.* Epidemiological links between tuberculosis cases identified twice as efficiently by whole genome sequencing than conventional molecular typing:

a population-based study. *PLoS One* 2018;**13**:e0195413. https://doi.org/10.1371/journal.pone.0195413.

Jajou R, Kohl TA, Walker T *et al.* Towards standardisation: comparison of five whole genome sequencing (WGS) analysis pipelines for detection of epidemiologically linked tuberculosis cases. *Eurosurveillance* 2019a;**24**:1900130. https://doi.org/10.2807/1560-7917.ES.2019.24.50.1900130.

Jajou R, Van Der Laan T, De Zwaan R *et al.* WGS more accurately predicts susceptibility of *Mycobacterium tuberculosis* to first-line drugs than phenotypic testing. *J Antimicrob Chemother* 2019b;**74**:2605–16. https://doi.org/10.1093/jac/dkz215.

Jones RC, Harris LG, Morgan S *et al.* Phylogenetic analysis of *Mycobacterium tuberculosis* strains in Wales by use of core genome multilocus sequence typing to analyze whole-genome sequencing data. *J Clin Microbiol* 2019;**57**:e02025–18. https://doi.org/10.1128/JCM.02025-18.

Joseph BV, Soman S, Radhakrishnan I *et al.* Molecular epidemiology of *Mycobacterium tuberculosis* isolates from Kerala, India using IS6110-RFLP, spoligotyping and MIRU-VNTRs. *Infect Genet Evol* 2013;**16**:157–64. https://doi.org/10.1016/j.meegid.2013.01.012.

Joshi KR, Dhiman H, Scaria V. tbvar: a comprehensive genome variation resource for *Mycobacterium tuberculosis*. *Database* 2014;**2014**:bat083. https://doi.org/10.1093/database/bat083.

Jünemann S, Sedlazeck FJ, Prior K *et al.* Updating benchtop sequencing performance comparison. *Nat Biotechnol* 2013;**31**:294–6. https://doi.org/10.1038/nbt.2522.

Kamerbeek J, Schouls L, Kolk A *et al.* Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 1997;**35**:907–14. https://doi.org/10.1128/jcm.35.4.907-914.1997.

Kapopoulou A, Lew JM, Cole ST. The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis* 2011;**91**:8–13. https://doi.org/10.1016/j.tube.2010.09.006.

Kargarpour Kamakoli M, Farmanfarmaei G, Masoumi M *et al.* Prediction of the hidden genotype of mixed infection strains in Iranian tuberculosis patients. *Int J Infect Dis* 2020;**95**:22–27. https://doi.org/10.1016/j.ijid.2020.03.056.

Katale BZ, Mbelele PM, Lema NA *et al.* Whole genome sequencing of *Mycobacterium tuberculosis* isolates and clinical outcomes of patients treated for multidrug-resistant tuberculosis in Tanzania. *BMC Genomics* 2020;**21**:174. https://doi.org/10.1186/s12864-020-6577-1.

Keshavjee S, Gelmanova IY, Pasechnikov AD *et al.* Treating multidrug-resistant tuberculosis in Tomsk, Russia. *Ann NY Acad Sci* 2008;**1136**:1–11. https://doi.org/10.1196/annals.1425.009.

Kohl TA, Diel R, Harmsen D *et al.* Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. *J Clin Microbiol* 2014;**52**:2479–86. https://doi.org/10.1128/JCM.00567-14.

Kohl TA, Harmsen D, Rothgänger J *et al.* Harmonized genome wide typing of tubercle Bacilli using a web-based gene-by-gene nomenclature system. *EBioMedicine* 2018a;**34**:131–8. https://doi.org/10.1016/j.ebiom.2018.07.030.

Kohl TA, Utpatel C, Schleusener V *et al.* MTBseq: a comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex isolates. *PeerJ* 2018b;**6**:e5895. https://doi.org/10.7717/peerj.5895.

Kremer K, Arnold C, Cataldi A *et al.* Discriminatory power and reproducibility of novel DNA typing methods for *Mycobacterium tuberculosis* complex strains. *J Clin Microbiol* 2005;**43**:5628–38. https://doi.org/10.1128/JCM.43.11.5628-5638.2005.

Kremer K, Glynn JR, Lillebaek T *et al.* Definition of the Beijing/W lineage of *Mycobacterium tuberculosis* on the basis of genetic markers. *J Clin Microbiol* 2004;**42**:4040–9. https://doi.org/10.1128/JCM.42.9.4040-4049.2004.

Kremer K, Van Soolingen D, Frothingham R *et al.* Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. *J Clin Microbiol* 1999;**37**:2607–18. https://doi.org/10.1128/jcm.37.8.2607-2618.1999.

Kubín M, Havelková M, Hynčicová I *et al.* A multidrug-resistant tuberculosis microepidemic caused by genetically closely related *Mycobacterium tuberculosis* strains. *J Clin Microbiol* 1999;**37**:2715–6. https://doi.org/10.1128/jcm.37.8.2715-2716.1999.

Lam C, Martinez E, Crighton T *et al.* Value of routine whole genome sequencing for *Mycobacterium tuberculosis* drug resistance detection. *Int J Infect Dis* 2021;**113**:S48–54. https://doi.org/10.1016/j.ijid.2021.03.033.

Lee ASG, Tang LLH, Lim IHK *et al.* Discrimination of single-copy IS *6110* DNA fingerprints of *Mycobacterium tuberculosis* isolates by high-resolution minisatellite-based typing. *J Clin Microbiol* 2002;**40**:657–9. https://doi.org/10.1128/JCM.40.2.657-659.2002.

Leong KWC, Gautam SS, Pradhan M *et al.* Comparative genomic analyses of multi-drug resistant *Mycobacterium tuberculosis* from Nepal and other geographical locations. *Genomics* 2022;**114**:110278. https://doi.org/10.1016/j.ygeno.2022.110278.

Li J, Yang T, Hong C *et al.* Whole-genome sequencing for resistance level prediction in multidrug-resistant tuberculosis. *Microbiol Spectr* 2022;**10**:e0271421. https://doi.org/10.1128/spectrum.02714-21.

Liang PK, Zheng C, Xu XF *et al.* Local adaptive evolution of two distinct clades of Beijing and T families of *Mycobacterium tuberculosis* in Chongqing: a bayesian population structure and phylogenetic study. *Infect Dis Poverty* 2020;**9**:59. https://doi.org/10.1186/s40249-020-00674-7.

Liu L, Li Y, Li S *et al.* Comparison of next-generation sequencing systems. *Biomed Res Int* 2012;**2012**:251364.

Lu H, Giordano F, Ning Z. Oxford Nanopore MinION sequencing and genome assembly. *Genomics, Proteomics and Bioinformatics* 2016;**14**:265–79.

Luo T, Yang C, Peng Y *et al.* Whole-genome sequencing to detect recent transmission of *Mycobacterium tuberculosis* in settings with a high burden of tuberculosis. *Tuberculosis* 2014;**94**:434–40. https://doi.org/10.1016/j.tube.2014.04.005.

Maeda S, Hijikata M, Hang NT *et al.* Genotyping of *Mycobacterium tuberculosis* spreading in Hanoi, Vietnam using conventional and whole genome sequencing methods. *Infect Genet Evol* 2020;**78**:104107. https://doi.org/10.1016/j.meegid.2019.104107.

Maes M, Kremer K, van Soolingen D *et al.* 24-Locus MIRU-VNTR genotyping is a useful tool to study the molecular epidemiology of tuberculosis among Warao Amerindians in Venezuela. *Tuberculosis* 2008;**88**:490–4. https://doi.org/10.1016/j.tube.2008.04.003.

Maghradze N, Jugheli L, Borrell S *et al.* Classifying recurrent *Mycobacterium tuberculosis* cases in Georgia using MIRU-VNTR typing. *PLoS One* 2019;**14**:e0223610. https://doi.org/10.1371/journal.pone.0223610.

Maghradze N, Jugheli L, Borrell S *et al.* Developing customized stepwise MIRU-VNTR typing for tuberculosis surveillance in Georgia. *PLoS One* 2022;**17**:e0264472. https://doi.org/10.1371/journal.pone.0264472.

Maiden MCJ, Van Rensburg MJJ, Bray JE *et al.* MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol* 2013;**11**:728–36. https://doi.org/10.1038/nrmicro3093.

Manson AL, Cohen KA, Abeel T *et al.* Genomic analysis of globally diverse *Mycobacterium tuberculosis* strains provides insights into the emergence and spread of multidrug resistance. *Nat Genet* 2017;**49**:395–402. https://doi.org/10.1038/ng.3767.

Mansoori N, Yaseri M, Vaziri F *et al.* Genetic diversity of *Mycobacterium tuberculosis* complex isolates circulating in an area with high tuberculosis incidence: using 24-locus MIRU-VNTR method. *Tuberculosis* 2018;**112**:89–97. https://doi.org/10.1016/j.tube.2018.08.003.

Margulies M, Egholm M, Altman WE *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;**437**:376–80.

Martín A, Herranz M, Lirola MM *et al.* Optimized molecular resolution of cross-contamination alerts in clinical mycobacteriology laboratories. *BMC Microbiol* 2008;**8**:30. https://doi.org/10.1186/1471-2180-8-30.

Mathema B, Kurepina NE, Bifani PJ *et al.* Molecular epidemiology of tuberculosis: current insights. *Clin Microbiol Rev* 2006;**19**:658–85. https://doi.org/10.1128/CMR.00061-05.

Mazars E, Lesjean S, Banuls AL *et al.* High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc Natl Acad Sci USA* 2001;**98**:1901–6. https://doi.org/10.1073/pnas.98.4.1901.

Meacham F, Boffelli D, Dhahbi J *et al.* Identification and correction of systematic error in high-throughput sequence data. *BMC Bioinformatics* 2011;**12**:451. https://doi.org/10.1186/1471-2105-12-451.

Meehan CJ, Goig GA, Kohl TA *et al.* Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. *Nat Rev Microbiol* 2019;**17**:533–45. https://doi.org/10.1038/s41579-019-0214-5.

Meehan CJ, Moris P, Kohl TA *et al.* The relationship between transmission time and clustering methods in *Mycobacterium tuberculosis* epidemiology. *EBioMedicine* 2018;**37**:410–6. https://doi.org/10.1016/j.ebiom.2018.10.013.

Mekonnen D, Munshea A, Nibret E *et al.* Comparative whole-genome sequence analysis of *Mycobacterium tuberculosis* isolated from pulmonary tuberculosis and tuberculous lymphadenitis patients in Northwest Ethiopia. *Front Microbiol* 2023;**14**:1211267. https://doi.org/10.3389/fmicb.2023.1211267.

Merker M, Blin C, Mona S *et al.* Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat Genet* 2015;**47**:242–9. https://doi.org/10.1038/ng.3195.

Merker M, Egbe NF, Ngangue YR *et al.* Transmission patterns of rifampicin resistant *Mycobacterium tuberculosis* complex strains in Cameroon: a genomic epidemiological study. *BMC Infect Dis* 2021;**21**:891. https://doi.org/10.1186/s12879-021-06593-8.

Merker M, Kohl TA, Niemann S *et al.* The evolution of strain typing in the *Mycobacterium tuberculosis* complex. *Adv Exp Med Biol* 2017;**1019**:43–78. https://doi.org/10.1007/978-3-319-64371-7_3.

Micheni LN, Kassaza K, Kinyi H *et al.* Detection of *Mycobacterium tuberculosis* multiple strains in sputum samples from patients with pulmonary tuberculosis in south western Uganda using MIRU-VNTR. *Sci Rep* 2022;**12**:1656. https://doi.org/10.1038/s41598-022-05591-3.

Mistry NF, Iyer AM, D'Souza DTB *et al.* Spoligotyping of *Mycobacterium tuberculosis* isolates from multiple-drug-resistant tuberculosis patients from Bombay, India. *J Clin Microbiol* 2002;**40**:2677–80. https://doi.org/10.1128/JCM.40.7.2677-2680.2002.

,Moco V., ,Cazenave D., ,Garnier M. *et al.*. getSequenceInfo: a suite of tools allowing to get genome sequence information from public repositories. *BMC Bioinformatics* 2022;**23**:268. https://doi.org/10.1186/s12859-022-04809-5.

Mokrousov I, Narvskaya O, Vyazovaya A *et al.* *Mycobacterium tuberculosis* Beijing genotype in Russia: in search of informative variable-number tandem-repeat loci. *J Clin Microbiol* 2008;**46**:3576–84. https://doi.org/10.1128/JCM.00414-08.

Molina-Moya B, Gomgnimbou MK, Spinasse L *et al.* *Mycobacterium tuberculosis* complex genotypes circulating in Nigeria based on spoligotyping obtained from Ziehl-Neelsen stained slides extracted DNA. *PLoS Negl Trop Dis* 2018;**12**:e0006242. https://doi.org/10.1371/journal.pntd.0006242.

Morey-León G, Mejía-Ponce PM, Granda Pardo JC *et al.* A precision overview of genomic resistance screening in isolates of *Mycobacterium tuberculosis* using web-based bioinformatics tools. *bioRxiv*, 2023. https://doi.org/10.1101/2023.01.10.523521, 13 January 2023, preprint: not peer reviewed.

Mudliar SKR, Kulsum U, Rufai SB *et al.* Snapshot of *Mycobacterium tuberculosis* phylogenetics from an Indian State of Arunachal Pradesh bordering China. *Genes* 2022;**13**:263. https://doi.org/10.3390/genes13020263.

Muzondiwa D, Mutshembele A, Pierneef RE *et al.* Resistance Sniffer: an online tool for prediction of drug resistance patterns of *Mycobacterium tuberculosis* isolates using next generation sequencing data. *Int J Med Microbiol* 2020;**310**:151399. https://doi.org/10.1016/j.ijmm.2020.151399.

Napier G, Couvin D, Refrégier G *et al.* Comparison of in silico predicted *Mycobacterium tuberculosis* spoligotypes and lineages from whole genome sequencing data. *Sci Rep* 2023;**13**:11368. https://doi.org/10.1038/s41598-023-38384-3.

Ngabonziza JCS, Loiseau C, Marceau M *et al.* A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. *Nat Commun* 2020;**11**:2917. https://doi.org/10.1038/s41467-020-16626-6.

Niang MN, Goguet De La Salmoniere Y, Samb A *et al.* Characterization of *M. tuberculosis* strains from West African patients by spoligotyping. *Microbes Infect* 1999;**1**:1189–92. https://doi.org/10.1016/S1286-4579(99)00243-9.

Niedringhaus TP, Milanova D, Kerby MB *et al.* Landscape of next-generation sequencing technologies. *Anal Chem* 2011;**83**:4327–41.

Niemann S, Ru¨sch-Gerdes S, Richter E *et al.* Stability of IS *6110* restriction fragment length polymorphism patterns of *Mycobacterium tuberculosis* strains in actual chains of transmission. *J Clin Microbiol* 2000;**38**:2563–7. https://doi.org/10.1128/JCM.38.7.2563-2567.2000.

Nikolayevskyy V, Kranzer K, Niemann S *et al.* Whole genome sequencing of *Mycobacterium tuberculosis* for detection of recent transmission and tracing outbreaks: a systematic review. *Tuberculosis* 2016a;**98**:77–85. https://doi.org/10.1016/j.tube.2016.02.009.

Nikolayevskyy V, Niemann S, Anthony R *et al.* Role and value of whole genome sequencing in studying tuberculosis transmission. *Clin Microbiol Infect* 2019;**25**:1377–82. https://doi.org/10.1016/j.cmi.2019.03.022.

Nikolayevskyy V, Trovato A, Broda A *et al.* MIRU-VNTR genotyping of *Mycobacterium tuberculosis* strains using qiaxcel technology: a multicentre evaluation study. *PLoS One* 2016b;**11**:e0149435. https://doi.org/10.1371/journal.pone.0149435.

Norheim G, Seterelv S, Arnesen TM *et al.* Tuberculosis outbreak in an educational institution in Norway. *J Clin Microbiol* 2017;**55**:1327–33. https://doi.org/10.1128/JCM.01152-16.

Ocheretina O, Merveille YM, Mabou MM *et al.* Use of Luminex MagPlex magnetic microspheres for high-throughput spoligotyping of *Mycobacterium tuberculosis* isolates in Port-au-Prince, Haiti. *J Clin Microbiol* 2013;**51**:2232–7. https://doi.org/10.1128/JCM.00268-13.

Oelemann MC, Diel R, Vatin V *et al.* Assessment of an optimized mycobacterial interspersed repetitive-unit- variable-

number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. *J Clin Microbiol* 2007;**45**:691–7. https://doi.org/10.1128/JCM.01393-06.

Otal I, Samper S, Asensio MP *et al.* Use of a PCR method based on IS6110 polymorphism for typing *Mycobacterium tuberculosis* strains from BACTEC cultures. *J Clin Microbiol* 1997;**35**:273–7. https://doi.org/10.1128/jcm.35.1.273-277.1997.

O'Toole RF. Development of a new genome-wide MLST scheme for high-resolution typing of diverse *Mycobacterium tuberculosis* complex strains. *EBioMedicine* 2018;**34**:6–7. https://doi.org/10.1016/j.ebiom.2018.07.038.

Page AJ, Alikhan N-F, Strinden M *et al.* Rapid *Mycobacterium tuberculosis* spoligotyping from uncorrected long reads using Galru. *bioRxiv*, 2020. https://doi.org/10.1101/2020.05.31.126490, 1 June 2020, preprint: not peer reviewed.

Pan J, Li X, Zhang M *et al.* TransFlow: a Snakemake workflow for transmission analysis of *Mycobacterium tuberculosis* whole-genome sequencing data. *Bioinformatics* 2023;**39**:btac785. https://doi.org/10.1093/bioinformatics/btac785.

Papaventsis D, Casali N, Kontsevaya I *et al.* Whole genome sequencing of *Mycobacterium tuberculosis* for detection of drug resistance: a systematic review. *Clin Microbiol Infect* 2017;**23**:61–8. https://doi.org/10.1016/j.cmi.2016.09.008.

Parson W, Strobl C, Huber G *et al.* Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome machine (PGM). *Forensic Sci Int Genet* 2013;**7**:543–9.

Peker N, Schuele L, Kok N *et al.* Evaluation of whole-genome sequence data analysis approaches for short- and long-read sequencing of *Mycobacterium tuberculosis*. *Microbial Genom* 2021;**7**:000695. https://doi.org/10.1099/mgen.0.000695.

Perdigão J, Silva C, Diniz J *et al.* Clonal expansion across the seas as seen through CPLP-TB database: a joint effort in cataloguing *Mycobacterium tuberculosis* genetic diversity in Portuguese-speaking countries. *Infect Genet Evol* 2019;**72**:44–58. https://doi.org/10.1016/j.meegid.2018.03.011.

Peres RL, Vinhas SA, Ribeiro FKC *et al.* Risk factors associated with cluster size of *Mycobacterium tuberculosis* (Mtb) of different RFLP lineages in Brazil. *BMC Infect Dis* 2018;**18**:71. https://doi.org/10.1186/s12879-018-2969-0.

Pescarini JM, Simonsen V, Ferrazoli L *et al.* Migration and tuberculosis transmission in a middle-income country: a cross-sectional study in a central area of São Paulo, Brazil. *BMC Med* 2018;**16**:62. https://doi.org/10.1186/s12916-018-1055-1.

Phelan JE, O'Sullivan DM, Machado D *et al.* Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Med* 2019;**11**:41. https://doi.org/10.1186/s13073-019-0650-x.

Plikaytis BB, Crawford JT, Woodley CL *et al.* Rapid, amplification-based fingerprinting of *Mycobacterium tuberculosis*. *J Gen Microbiol* 1993;**139**:1537–42. https://doi.org/10.1099/00221287-139-7-1537.

Pokam BDT, Yeboah-Manu D, Lawson L *et al.* Molecular analysis of *Mycobacterium tuberculosis* isolated in the north central zone of Nigeria. *JEGH* 2019;**9**:259. https://doi.org/10.2991/jegh.k.191015.001.

Puustinen K, Marjamäki M, Rastogi N *et al.* Characterization of Finnish *Mycobacterium tuberculosis* isolates by spoligotyping. *J Clin Microbiol* 2003;**41**:1525–8. https://doi.org/10.1128/JCM.41.4.1525-1528.2003.

Quan Z, Li M, Chen Y *et al.* Performance evaluation of core genome multilocus sequence typing for genotyping of *Mycobacterium tuberculosis* strains in China: based on multicenter, population-based collection. *Eur J Clin Microbiol Infect Dis* 2024;**43**:297–304. https://doi.org/10.1007/s10096-023-04720-8.

Rajwani R, Shehzad S, Siu GKH. MIRU-profiler: a rapid tool for determination of 24-loci MIRU-VNTR profiles from assembled genomes of *Mycobacterium tuberculosis*. *PeerJ* 2018;**6**:e5090. https://doi.org/10.7717/peerj.5090.

Ramazanzadeh R, Shakib P, Rouhi S *et al.* Molecular epidemiology of *Mycobacterium tuberculosis* isolates in Iran using spoligotyping. *New Microbes New Infect* 2020;**38**:100767. https://doi.org/10.1016/j.nmni.2020.100767.

Ramirez LMN, Vargas KQ, Diaz G. Whole genome sequencing for the analysis of drug resistant strains of *Mycobacterium tuberculosis*: a systematic review for bedaquiline and delamanid. *Antibiotics* 2020;**9**:133. https://doi.org/10.3390/antibiotics9030133.

Razanamparany VR, Ménard D, Aurégan G *et al.* Extrapulmonary and pulmonary tuberculosis in Antananarivo (Madagascar): high clustering rate in female patients. *J Clin Microbiol* 2002;**40**:3964–9. https://doi.org/10.1128/JCM.40.11.3964-3969.2002.

Razo CAP, Hernández ER, Ponce SIR *et al.* Molecular epidemiology of cattle tuberculosis in mexico through whole-genome sequencing and spoligotyping. *PLoS One* 2018;**13**:e0201981. https://doi.org/10.1371/journal.pone.0201981.

Reisig F, Kremer K, Amthor B *et al.* Fast ligation-mediated PCR, a Fast and reliable method for IS *6110* -based typing of *Mycobacterium tuberculosis* complex. *J Clin Microbiol* 2005;**43**:5622–7. https://doi.org/10.1128/JCM.43.11.5622-5627.2005.

Reyes JF, Chan CHS, Tanaka MM. Impact of homoplasy on variable numbers of tandem repeats and spoligotypes in *Mycobacterium tuberculosis*. *Infect Genet Evol* 2012;**12**:811–8. https://doi.org/10.1016/j.meegid.2011.05.018.

Reyes JF, Tanaka MM. Mutation rates of spoligotypes and variable numbers of tandem repeat loci in *Mycobacterium tuberculosis*. *Infect Genet Evol* 2010;**10**:1046–51. https://doi.org/10.1016/j.meegid.2010.06.016.

Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics, Proteomics and Bioinformatics* 2015;**13**:278–89.

Ribeiro FKC, Pan W, Bertolde A *et al.* Genotypic and spatial analysis of *Mycobacterium tuberculosis* transmission in a high-incidence urban setting. *Clin Infect Dis* 2015;**61**:758–66. https://doi.org/10.1093/cid/civ365.

Rienthong D, Ajawatanawong P, Rienthong S *et al.* Restriction fragment length polymorphism study of nationwide samples of *Mycobacterium tuberculosis* in Thailand, 1997–1998. *Int J Tuberc Lung Dis* 2005;**9**:576–81.

Rodriguez-Campos S, González S, de Juan L *et al.* A database for animal tuberculosis (mycoDB.es) within the context of the Spanish national programme for eradication of bovine tuberculosis. *Infect Genet Evol* 2012;**12**:877–82. https://doi.org/10.1016/j.meegid.2011.10.008.

Roetzer A, Diel R, Kohl TA *et al.* Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS Med* 2013;**10**:e1001387. https://doi.org/10.1371/journal.pmed.1001387.

Rosenthal A, Gabrielian A, Engle E *et al.* The TB portals: an open-access, web-based platform for global drug-resistant- tuberculosis data sharing and analysis. *J Clin Microbiol* 2017;**55**:3267–82. https://doi.org/10.1128/JCM.01013-17.

Ruddy MC, Davies AP, Yates MD *et al.* Outbreak of isoniazid resistant tuberculosis in north London. *Thorax* 2004;**59**:279–85. https://doi.org/10.1136/thx.2003.010405.

Rudeeaneksin J, Bunchoo S, Phetsuksiri B *et al.* The first insight into *Mycobacterium tuberculosis* complex isolates in the lower northern

region in Thailand. *Trans R Soc Trop Med Hyg* 2024;**118**:527–36. https://doi.org/10.1093/trstmh/trae014.

Said HM, Krishnamani K, Omar SV *et al.* Evaluation of semiautomated IS *6110* -based restriction fragment length polymorphism typing for *Mycobacterium tuberculosis* in a high-burden setting. *J Clin Microbiol* 2016;**54**:2547–52. https://doi.org/10.1128/JCM.00408-16.

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977;**74**:5463–7.

Satta G, Atzeni A, McHugh TD. *Mycobacterium tuberculosis* and whole genome sequencing: a practical guide and online tools available for the clinical microbiologist. *Clin Microbiol Infect* 2017;**23**:69–72. https://doi.org/10.1016/j.cmi.2016.09.005.

Schewe C, Goldmann T, Grosser M *et al.* Inter-laboratory validation of PCR-based detection of *Mycobacterium tuberculosis* in formalin-fixed, paraffin-embedded tissues. *Virchows Arch* 2005;**447**:573–85. https://doi.org/10.1007/s00428-005-1233-3.

Schürch AC, Kremer K, Daviena O *et al.* High-resolution typing by integration of genome sequencing data in a large tuberculosis cluster. *J Clin Microbiol* 2010;**48**:3403–6. https://doi.org/10.1128/JCM.00370-10.

Sekizuka T, Yamashita A, Murase Y *et al.* TGS-TB: total genotyping solution for *Mycobacterium tuberculosis* using short-read whole-genome sequencing. *PLoS One* 2015;**10**:e0142951. https://doi.org/10.1371/journal.pone.0142951.

Senelle G, Sahal MR, La K *et al.* Towards the reconstruction of a global TB history using a new pipeline "TB-Annotator". *Tuberculosis* 2023;**143**:102376. https://doi.org/10.1016/j.tube.2023.102376.

Shao Y, Song H, Li G *et al.* Relapse or re-infection, the situation of recurrent tuberculosis in Eastern China. *Front Cell Infect Microbiol* 2021;**11**:638990. https://doi.org/10.3389/fcimb.2021.638990.

Shea J, Halse TA, Lapierre P *et al.* Comprehensive whole-genome sequencing and reporting of drug resistance profiles on clinical cases of *Mycobacterium tuberculosis* in New York State. *J Clin Microbiol* 2017;**55**:1871–82. https://doi.org/10.1128/JCM.00298-17.

Shi J, Zheng D, Zhu Y *et al.* Role of MIRU-VNTR and spoligotyping in assessing the genetic diversity of *Mycobacterium tuberculosis* in Henan Province, China. *BMC Infect Dis* 2018;**18**: 447. https://doi.org/10.1186/s12879-018-3351-y.

Smith NH, Upton P. Naming spoligotype patterns for the RD9-deleted lineage of the *Mycobacterium tuberculosis* complex; www.Mbovis.Org. *Infect Genet Evol* 2012;**12**:873–6. https://doi.org/10.1016/j.meegid.2011.08.002.

Soini H, Pan X, Amin A *et al.* Characterization of *Mycobacterium tuberculosis* isolates from patients in Houston, Texas, by spoligotyping. *J Clin Microbiol* 2000;**38**:669–76. https://doi.org/10.1128/jcm.38.2.669-676.2000.

Sola C, Devallois A, Horgen L *et al.* Tuberculosis in the Caribbean: using spacer oligonucleotide typing to understand strain origin and transmission. *Emerg Infect Dis* 1999;**5**:404–11. https://doi.org/10.3201/eid0503.990311.

Sola C, Filliol I, Gutierrez MC *et al.* Spoligotype Database of *Mycobacterium tuberculosis* : biogeographic distribution of shared types and epidemiologic and phylogenetic perspectives. *Emerg Infect Dis* 2001;**7**:390–6. https://doi.org/10.3201/eid0703.010304.

Sola C, Filliol I, Legrand E *et al.* Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. *Infect Genet Evol* 2003;**3**:125–33. https://doi.org/10.1016/S1567-1348(03)00011-X.

Song EJ, Jeong HJ, Lee SM *et al.* A DNA chip-based spoligotyping method for the strain identification of *Mycobacterium tuberculosis* isolates. *J Microbiol Methods* 2007;**68**:430–3. https://doi.org/10.1016/j.mimet.2006.09.005.

Song Z, He W, Cao X *et al.* The recent transmission and associated risk factor of *Mycobacterium tuberculosis* in Golmud City, China. *Infect Drug Resist* 2024;**17**:417–25. https://doi.org/10.2147/IDR.S437026.

Starks AM, Avilés E, Cirillo DM *et al.* Collaborative effort for a centralized Worldwide tuberculosis relational sequencing Data platform: figure 1. *Clin Infect Dis* 2015;**61**:S141–6. https://doi.org/10.1093/cid/civ610.

Stucki D, Ballif M, Egger M *et al.* Standard genotyping overestimates transmission of *Mycobacterium tuberculosis* among immigrants in a low-incidence country. *J Clin Microbiol* 2016;**54**:1862–70. https://doi.org/10.1128/JCM.00126-16.

Stucki D, Gagneux S. Single nucleotide polymorphisms in *Mycobacterium tuberculosis* and the need for a curated database. *Tuberculosis* 2013;**93**:30–9. https://doi.org/10.1016/j.tube.2012.11.002.

Supply P, Allix C, Lesjean S *et al.* Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 2006;**44**:4498–510. https://doi.org/10.1128/JCM.01392-06.

Supply P, Lesjean S, Savine E *et al.* Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J Clin Microbiol* 2001;**39**:3563–71. https://doi.org/10.1128/JCM.39.10.3563-3571.2001.

Supply P, Magdalena J, Himpens S *et al.* Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. *Mol Microbiol* 1997;**26**:991–1003. https://doi.org/10.1046/j.1365-2958.1997.6361999.x.

Supply P, Mazars E, Lesjean S *et al.* Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol Microbiol* 2000;**36**:762–71. https://doi.org/10.1046/j.1365-2958.2000.01905.x.

Tafaj S, Ghariani A, Trovato A *et al.* Accuracy of the qiaxcel automated system for MIRU-VNTR genotyping of *Mycobacterium tuberculosis* in two limited resource settings. *J Clin Med* 2020;**9**:389. https://doi.org/10.3390/jcm9020389.

Tang CY, Ong RT-H. MIREader: MIRU-VNTR typing directly from long sequencing reads. *Bioinformatics* 2020;**36**:1625–6. https://doi.org/10.1093/bioinformatics/btz771.

Thabet S, Karboul A, Dekhil N *et al.* IS6110-5'3'FP: an automated typing approach for *Mycobacterium tuberculosis* complex strains simultaneously targeting and resolving IS6110 5' and 3' polymorphisms. *Int J Infect Dis* 2014;**29**:211–8. https://doi.org/10.1016/j.ijid.2014.10.004.

The CRyPTIC Consortium and the 100 000 Genomes Project. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. *N Engl J Med* 2018;**379**:1403–15. https://doi.org/10.1056/NEJMoa1800474.

Thierry D, Brisson-Noel A, Vincent-Levy-Frebault V *et al.* Characterization of a *Mycobacterium tuberculosis* insertion sequence, IS6110, and its application in diagnosis. *J Clin Microbiol* 1990;**28**:2668–73. https://doi.org/10.1128/jcm.28.12.2668-2673.1990.

Thorne N, Evans JT, Smith EG *et al.* An IS 6110-targeting fluorescent amplified fragment length polymorphism alternative to IS 6110 restriction fragment length polymorphism analysis for *Mycobacterium tuberculosis* DNA fingerprinting. *Clin Microbiol Infect* 2007;**13**:964–70. https://doi.org/10.1111/j.1469-0691.2007.01783.x.

Tilahun M, Ameni G, Desta K *et al.* Molecular epidemiology and drug sensitivity pattern of *Mycobacterium tuberculosis* strains isolated from pulmonary tuberculosis patients in and around Ambo

Town, Central Ethiopia. *PLoS One* 2018;**13**:e0193083. https://doi.org/10.1371/journal.pone.0193083.

Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. *Am Hum Genet* 2009;**85**:142–54.

Tyler AD, Christianson S, Knox NC *et al.* Comparison of sample preparation methods used for the next-generation sequencing of *Mycobacterium tuberculosis*. *PLoS One* 2016;**11**:e0148676.

Valencia-Trujillo D, Avila-Trejo AM, García-Reyes RL *et al.* Genetic diversity of *Mycobacterium tuberculosis* strains isolated from HIV-infected patients in Mexico. *Pathogens* 2024;**13**:428. https://doi.org/10.3390/pathogens13050428.

Valway SE, Sanchez MP, Shinnick TF *et al.* An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. *N Engl J Med* 1998;**338**:633–9. [see comments] [published erratum appears in N Engl J Med 1998 Jun 11;338(24):1783].

Van Der Zanden AGM, Hoentjen AH, Heilmann FGC *et al.* Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* complex in paraffin wax embedded tissues and in stained microscopic preparations. *Mol Pathol* 1998;**51**:209–14. https://doi.org/10.1136/mp.51.4.209.

Van der Zanden AGM, Kremer K, Schouls LM *et al.* Improvement of differentiation and interpretability of spoligotyping for *Mycobacterium tuberculosis* complex isolates by introduction of new spacer oligonucleotides. *J Clin Microbiol* 2002;**40**:4628–39. https://doi.org/10.1128/JCM.40.12.4628-4639.2002.

Van Deutekom H, Supply P, De Haas PEW *et al.* Molecular typing of *Mycobacterium tuberculosis* by mycobacterial interspersed repetitive unit-variable-number tandem repeat analysis, a more accurate method for identifying epidemiological links between patients with tuberculosis. *J Clin Microbiol* 2005;**43**:4473–9. https://doi.org/10.1128/JCM.43.9.4473-4479.2005.

Van Duin JM, Pijnenburg JEM, Van Rijswoud CM *et al.* Investigation of cross contamination in a *Mycobacterium tuberculosis* laboratory using IS6110 DNA fingerprinting. *Int J Tuberc Lung Dis* 1998;**2**:425–9.

Van Embden JDA, Cave MD, Crawford JT *et al.* Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993;**31**:406–9. https://doi.org/10.1128/jcm.31.2.406-409.1993.

Van Embden JDA, Van Gorkom T, Kremer K *et al.* Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. *J Bacteriol* 2000;**182**:2393–401. https://doi.org/10.1128/JB.182.9.2393-2401.2000.

van Rie A, Warren R, Richardson M *et al.* Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. *N Engl J Med* 1999;**341**:1174–9. https://doi.org/10.1056/nejm199910143411602.

Van Soolingen D, Arbeit RD. Dealing with variation in molecular typing of *Mycobacterium tuberculosis*: low-intensity bands and other challenges. *J Med Microbiol* 2001;**50**:749–51. https://doi.org/10.1099/0022-1317-50-9-749.

Velji P, Nikolayevskyy V, Brown T *et al.* Discriminatory ability of hypervariable variable number tandem repeat loci in population-based analysis of *Mycobacterium tuberculosis* strains, London, UK. *Emerg Infect Dis* 2009;**15**:1609–16. https://doi.org/10.3201/eid1510.090463.

Veziris N, Bernard C, Guglielmetti L *et al.* Rapid emergence of *Mycobacterium tuberculosis* bedaquiline resistance: lessons to avoid repeating past errors. *Eur Respir J* 2017;**49**:1601719. https://doi.org/10.1183/13993003.01719-2016.

Vitol I, Driscoll J, Kreiswirth B *et al.* Identifying *Mycobacterium tuberculosis* complex strain families using spoligotypes. *Infect Genet Evol* 2006;**6**:491–504. https://doi.org/10.1016/j.meegid.2006.03.003.

Walker TM, Ip CLC, Harrell RH *et al.* Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis* 2013;**13**:137–46. https://doi.org/10.1016/S1473-3099(12)70277-3.

Walker TM, Lalor MK, Broda A *et al.* Assessment of *Mycobacterium tuberculosis* transmission in Oxfordshire, UK, 2007-12, with whole pathogen genome sequences: an observational study. *Lancet Respir Med* 2014;**2**:285–92. https://doi.org/10.1016/S2213-2600(14)70027-X.

Walker TM, Merker M, Knoblauch AM *et al.* A cluster of multidrug-resistant *Mycobacterium tuberculosis* among patients arriving in Europe from the Horn of Africa: a molecular epidemiological study. *Lancet Infect Dis* 2018;**18**:431–40. https://doi.org/10.1016/S1473-3099(18)30004-5.

Walker TM, Merker M, Kohl TA *et al.* Whole genome sequencing for M/XDR tuberculosis surveillance and for resistance testing. *Clin Microbiol Infect* 2017;**23**:161–6. https://doi.org/10.1016/j.cmi.2016.10.014.

Wang X, Liu H, Wei J *et al.* An investigation on the population structure of mixed infections of *Mycobacterium tuberculosis* in Inner Mongolia, China. *Tuberculosis* 2015;**95**:695–700. https://doi.org/10.1016/j.tube.2015.08.006.

Weniger T, Krawczyk J, Supply P *et al.* MIRU-VNTRplus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. *Nucleic Acids Res* 2010;**38**:W326–31. https://doi.org/10.1093/nar/gkq351.

Witney AA, Bateson ALE, Jindani A *et al.* Use of whole-genome sequencing to distinguish relapse from reinfection in a completed tuberculosis clinical trial. *BMC Med* 2017;**15**:71. https://doi.org/10.1186/s12916-017-0834-4.

Wu X, Gao R, Shen X *et al.* Use of whole-genome sequencing to predict *Mycobacterium tuberculosis* drug resistance in Shanghai, China. *Int J Infect Dis* 2020;**96**:48–53. https://doi.org/10.1016/j.ijid.2020.04.039.

Wyllie DH, Davidson JA, Grace Smith E *et al.* A quantitative evaluation of MIRU-VNTR typing against whole-genome sequencing for identifying *Mycobacterium tuberculosis* transmission: a prospective observational cohort study. *EBioMedicine* 2018;**34**:122–30. https://doi.org/10.1016/j.ebiom.2018.07.019.

Xia E, Teo YY, Ong RTH. SpoTyping: fast and accurate in silico mycobacterium spoligotyping from sequence reads. *Genome Med* 2016;**8**:19. https://doi.org/10.1186/s13073-016-0270-7.

Xia Z, Su B, Tu C *et al.* Single-tube protocol for culture-independent spoligotyping of *Mycobacterium tuberculosis* based on MeltArray. *J Clin Microbiol* 2024;**62**:e0118323. https://doi.org/10.1128/jcm.01183-23.

Xiao Y-X, Chan T-H, Liu K-H *et al.* Define SNP thresholds for delineation of tuberculosis transmissions using whole-genome sequencing. *Microbiol Spectr* 2024;**12**:e0041824. https://doi.org/10.1128/spectrum.00418-24.

Yang T, Gan M, Liu Q *et al.* SAM-TB: a whole genome sequencing data analysis website for detection of *Mycobacterium tuberculosis* drug resistance and transmission. *Briefings Bioinform* 2022;**23**:bbac030. https://doi.org/10.1093/bib/bbac030.

Yeh RW, De Leon AP, Agasino CB *et al.* Stability of *Mycobacterium tuberculosis* DNA genotypes. *J Infect Dis* 1998;**177**:1107–11. https://doi.org/10.1086/517406.

Yin C, Mijiti X, Liu H *et al.* Molecular epidemiology of clinical *Mycobacterium tuberculosis* isolates from Southern Xinjiang, China using spoligotyping and 15-locus MIRU-VNTR typing. *Infect Drug Resist* 2023;**16**:1313–26. https://doi.org/10.2147/IDR.S393192.

Zeng X, Xu Y, Zhou Y *et al.* McSpoligotyping, a one-step melting curve analysis-based protocol for spoligotyping of *Mycobacterium tuber-*

culosis. *J Clin Microbiol* 2018;**56**:e00539–18. https://doi.org/10.1128/JCM.00539-18.

Zewdie O, Mihret A, Ameni G *et al*. Molecular typing of mycobacteria isolated from tuberculous lymphadenitis cases in Addis Ababa, Ethiopia. *Int J Tuberc Lung Dis* 2016;**20**:1529–34. https://doi.org/10.5588/ijtld.15.1023.

Zhang J, Abadia E, Refregier G *et al*. *Mycobacterium tuberculosis* complex CRISPR genotyping: improving efficiency, throughput and discriminative power of 'spoligotyping' with new spacers and a microbead-based hybridization assay. *J Med Microbiol* 2010;**59**:285–94. https://doi.org/10.1099/jmm.0.016949-0.

Zhang X, Martinez E, Lam C *et al*. Exploring programmatic indicators of tuberculosis control that incorporate routine *Mycobacterium tuberculosis* sequencing in low incidence settings: a comprehensive (2017–2021) patient cohort analysis. *The Lancet Regional Health—Western Pacific* 2023;**41**:100910. https://doi.org/10.1016/j.lanwpc.2023.100910.

Zink AR, Sola C, Reischl U *et al*. Characterization of *Mycobacterium tuberculosis* complex DNAs from Egyptian mummies by spoligotyping. *J Clin Microbiol* 2003;**41**:359–67. https://doi.org/10.1128/JCM.41.1.359-367.2003.

Received 2 January 2025; **revised** 20 April 2025; **accepted** 25 April 2025