



OPEN ACCESS

EDITED BY

João Perdigão,
University of Lisbon, Portugal

REVIEWED BY

David Couvin,
Institut Pasteur de Guadeloupe, Guadeloupe
Addy Cecilia Helguera-Repetto,
Instituto Nacional de Perinatología
(INPER), Mexico

*CORRESPONDENCE

Daniel Mekonnen
✉ nigusdaniel@gmail.com

RECEIVED 24 April 2023

ACCEPTED 30 May 2023

PUBLISHED 30 June 2023

CITATION

Mekonnen D, Munshea A, Nibret E, Adnew B, Herrera-Leon S, Amor Aramendia A, Benito A, Abascal E, Jacqueline C, Aseffa A and Herrera-Leon L (2023) Comparative whole-genome sequence analysis of *Mycobacterium tuberculosis* isolated from pulmonary tuberculosis and tuberculous lymphadenitis patients in Northwest Ethiopia. *Front. Microbiol.* 14:1211267. doi: 10.3389/fmicb.2023.1211267

COPYRIGHT

© 2023 Mekonnen, Munshea, Nibret, Adnew, Herrera-Leon, Amor Aramendia, Benito, Abascal, Jacqueline, Aseffa and Herrera-Leon. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comparative whole-genome sequence analysis of *Mycobacterium tuberculosis* isolated from pulmonary tuberculosis and tuberculous lymphadenitis patients in Northwest Ethiopia

Daniel Mekonnen^{1,2,3*}, Abaineh Munshea^{2,4}, Endalkachew Nibret^{2,4}, Bethlehem Adnew⁵, Silvia Herrera-Leon⁶, Aranzazu Amor Aramendia⁷, Agustín Benito⁸, Estefanía Abascal⁶, Camille Jacqueline^{6,9}, Abraham Aseffa⁵ and Laura Herrera-Leon^{6,10}

¹Department of Medical Laboratory Sciences, School of Health Science, College of Medicine and Health Sciences, Bahir Dar University, Bahir Dar, Ethiopia, ²Health Biotechnology Division, Institute of Biotechnology, Bahir Dar University, Bahir Dar, Ethiopia, ³Amhara Public Health Institute, Bahir Dar, Ethiopia, ⁴Department of Biology, Bahir Dar University, Bahir Dar, Ethiopia, ⁵Armauer Hansen Research Institute, Addis Ababa, Ethiopia, ⁶National Centre for Microbiology, Instituto de Salud Carlos III, Madrid, Spain, ⁷Fundación Mundo Sano, Madrid, Spain, ⁸National Center of Tropical Medicine, Instituto de Salud Carlos III, Centro de Investigación Biomédica en Red de Enfermedades Infecciosas, Madrid, Spain, ⁹European Public Health Microbiology Training Programme, European Centre for Disease Prevention and Control, Stockholm, Sweden, ¹⁰CIBER Epidemiología y Salud Pública, Madrid, Spain

Background: Tuberculosis (TB), caused by the *Mycobacterium tuberculosis* complex (MTBC), is a chronic infectious disease with both pulmonary and extrapulmonary forms. This study set out to investigate and compare the genomic diversity and transmission dynamics of *Mycobacterium tuberculosis* (*Mtb*) isolates obtained from tuberculous lymphadenitis (TBLN) and pulmonary TB (PTB) cases in Northwest Ethiopia.

Methods: A facility-based cross-sectional study was conducted using two groups of samples collected between February 2021 and June 2022 (Group 1) and between June 2020 and June 2022 (Group 2) in Northwest Ethiopia. Deoxyribonucleic acid (DNA) was extracted from 200 heat-inactivated *Mtb* isolates. Whole-genome sequencing (WGS) was performed from 161 isolates having ≥ 1 ng DNA/ μ l using Illumina NovaSeq 6000 technology.

Results: From the total 161 isolates sequenced, 146 *Mtb* isolates were successfully genotyped into three lineages (L) and 18 sub-lineages. The Euro-American (EA, L4) lineage was the prevailing ($n = 100$; 68.5%) followed by Central Asian (CAS, L3, $n = 43$; 25.3%) and then L7 ($n = 3$; 2.05%). The L4.2.2.ETH sub-lineage accounted for 19.9%, while Haarlem estimated at 13.7%. The phylogenetic tree revealed distinct *Mtb* clusters between PTB and TBLN isolates even though there was no difference at lineages and sub-lineages levels. The clustering rate (CR) and recent transmission index (RTI) for PTB were 30 and 15%, respectively. Similarly, the CR and RTI for TBLN were 31.1 and 18 %, respectively.

Conclusion and recommendations: PTB and TBLN isolates showed no *Mtb* lineages and sub-lineages difference. However, at the threshold of five allelic distances, *Mtb* isolates obtained from PTB and TBLN form distinct complexes in the phylogenetic tree, which indicates the presence of *Mtb* genomic variation among the two clinical forms. The high rate of clustering and RTI among TBLN implied that TBLN was likely the result of recent transmission and/or reactivation from short latency. Hence, the high incidence rate of TBLN in the Amhara region could be the result of *Mtb* genomic diversity and rapid clinical progression from primary infection and/or short latency. To validate this conclusion, a similar community-based study with a large sample size and better sampling technique is highly desirable. Additionally, analysis of genomic variants other than phylogenetic informative regions could give insightful information. Combined analysis of the host and the pathogen genome (GXG) together with environmental (GxGxE) factors could give comprehensive co-evolutionary information.

KEYWORDS

Mycobacterium tuberculosis, tuberculous lymphadenitis, pulmonary tuberculosis, whole-genome sequencing, Ethiopia

1. Background

Tuberculosis (TB) is a chronic infectious disease with complex epidemiological characteristics (Abel et al., 2014; Barberis et al., 2017; Adhikari, 2022). It is an ancient disease causing illness and death in humans since the primordial times (Palomino et al., 2007). Over 10.9 and 1.4 million TB cases and deaths were reported in 2021 alone (WHO, 2022). *Mycobacterium tuberculosis* complex (MTBC), the etiological agents of TB, includes *Mycobacterium tuberculosis* (*M. tuberculosis*, *Mtb*) (Sakula, 1982), *M. africanum* (de Jong et al., 2010), *M. bovis*, *M. microti*, *M. pinnipedii*, *M. orygis*, *M. mungi*, and *M. suricattae* (Gagneux, 2018). The human-adapted members of MTBC (*Mtb* and *M. africanum*) are further classified into nine lineages with distinct geographic structures (Coscolla et al., 2021).

Ethiopia is a country in East Africa with high TB incidence and a complex *Mtb* population genetic structure (Mekonnen et al., 2019c). This complex population genetic structure is likely shaped by the consecutive entry of strains through long trade across the red sea, human migration over the millennia (Comas et al., 2013, 2015), and the long-established co-evolutionary trajectory (May and Anderson, 1983). The incidence of tuberculous lymphadenitis (TBLN), which is clinically characterized by the inflammation of the lymph node (cervical, axillary, or inguinal) with or without TB constitutional symptoms (cough, weight loss, fever, and night sweats), is exceptionally high in Ethiopia (Berg et al., 2015). However, causal factors leading to a high incidence rate of TBLN in Ethiopia remain speculative. Factors including ethnicity, *Mtb* strain type, HIV co-infection (Firdessa et al., 2013; Berg et al., 2015; Mekonnen et al., 2019b), over diagnosis (Iwnetu et al., 2009), and spill-over transmission of bovine TB (Ameni et al., 2011; Tadesse et al., 2017) were excluded. On the other hand, female sex, younger age (Mekonnen et al., 2019b), delayed diagnosis (Asres et al., 2019),

and rural residency (Mekonnen et al., 2022) demonstrated a significant association.

The controversy about scientific evidence for “the incidence of TBLN with no lung involvement” is the subject of ongoing debate (Ganchua et al., 2020). For instance, Yew and Lee (1995) and Deveci et al. (2016) concluded that TBLN without pulmonary TB (PTB) is rare. To accept the theory that TBLN can exist as a distinct localized form of TB without pulmonary involvement, pathogenesis pathways other than lymphohematogenous spread, such as via tonsils and adenoids, must be proposed (Yew and Lee, 1995; Deveci et al., 2016). Herath and Lewis (2014) support the above claims and found a higher rate of concomitant PTB among extrapulmonary TB (EPTB) patients but with normal chest X-ray findings. In contrast to the above findings, a higher number of TBLN cases without concurrent PTB has been reported by multiple studies. Briefly, in Polesky et al.’s (2005) study, out of 106 TBLN cases, only 16 (15%) had evidence of pulmonary involvement. Of the total TBLN cases, concomitant PTB was reported among 20.8% of TBLN cases (Qian et al., 2018). In another study by Bomanji et al. (2020), only 100 (28%) out of 358 EPTB patients showed pulmonary involvement. Nevertheless, to disprove either of the two scientific narratives, a clinical, radiological, and pathological investigation should be coupled with bacteriological methods.

Whole-genome sequencing (WGS) technology demonstrated an incredible potential for investigating the population’s genomic structure and transmission clusters (Gardy et al., 2011; Walker et al., 2013; and Meehan et al., 2018). Core genome multilocus sequence typing (cgMLST) and single-nucleotide polymorphism (SNP) with 5–12 SNP or allelic differences were suggested for identifying epidemiological links with temporal scales (Gardy et al., 2011; Walker et al., 2013; and Meehan et al., 2018). Recent transmission index (RTI) is the measure of the severity of the *Mtb* epidemic and indicates the rapid progression of the infection into clinical illness (Tanaka and Francis, 2005; Tessema et al., 2013).

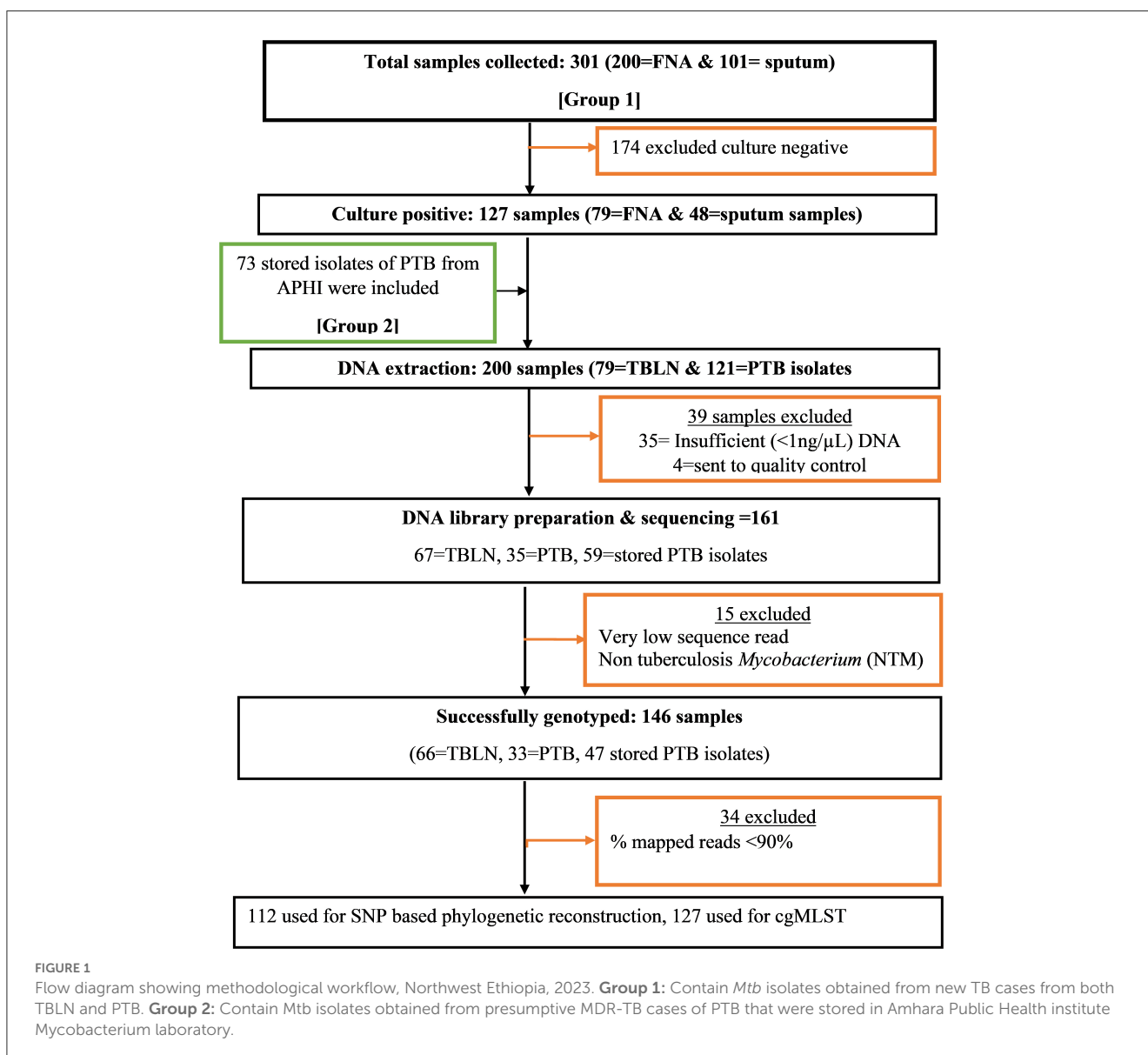
A few number of studies examined whether the high incidence rate of TBLN in Ethiopia was associated with *Mtb* lineages and sub-lineages (Firdessa et al., 2013; Ejo et al., 2021). These findings indicated that lineages and sub-lineages were identical between PTB and TBLN forms. Furthermore, the rate of clustering was also similar among the two clinical forms (10% for PTB and 11% for TBLN) (Firdessa et al., 2013). Additionally, a comparative literature review in Africa showed no difference with regard to lineages and sub-lineages between the two forms of TB (Mekonnen et al., 2019a). However, as can be seen from the figures in our previous research studies (Mekonnen et al., 2019b, 2022), TBLN showed a geographic pattern similar to the MTBC geographic structuring. To explain further, TBLN is mainly reported in countries having a high prevalence of geographically restricted specialist MTBC (sub)-lineages (Mekonnen et al., 2019a). Hence, we speculated the presence of MTBC genomics influences on TB clinical phenotypes.

Hence, this study was designed to investigate the genomic diversity and transmission dynamics of *Mycobacterium tuberculosis* (*Mtb*) isolates obtained from TBLN and PTB cases in Northwest Ethiopia using WGS data.

2. Materials and methods

2.1. Study design and setting

A facility-based cross-sectional survey was carried out with two sets of samples (new TB patients and presumptive MDR-TB cases). The first group of participants included new PTB and TBLN cases identified at outpatient departments in Bahir Dar health facilities (Felege-Hiwot Comprehensive Specialized Hospital/FHCSH, Bahir Dar Health Center, Han Health Center, and Shum-Abo Health Center) between February



2021 and June 2022. The second group included presumptive multidrug-resistant tuberculosis (MDR-TB) isolates archived at Amhara Public Health Institute's (APHI) *Mycobacterium* reference laboratory between June 2020 and June 2022 (Figure 1). Patients were selected from both urban and rural parts of Northwest Ethiopia.

2.2. *Mycobacterium tuberculosis* culture

At the APHI *Mycobacterium* reference laboratory, the stored isolates were re-grown using Löwenstein–Jensen (LJ) medium and *Mycobacterium* growth indicator tube (MGIT) media. When both LJ and MGIT gave confluent growth, the colony was taken from LJ, or otherwise, MGIT was used. When MGIT became positive, the culture was further incubated at 37°C for an additional 7 days to enrich the MTBC population and make the MTBC cells aggregate and clearly visible in the transparent media. One milliliter of MGIT media containing MTBC aggregates was transferred into 1.5-ml screw-cap vials. The MTBC grown on LJ media was harvested using a sterile plastic loop and transferred into 1.5-ml screw-cap vials containing 1 ml of distilled water. Then the colonies were inactivated by heating at 95°C for 30 min. The heat-inactivated MTBC was transported to the National Center of Microbiology, Institute of Carlos III (ISCIII), Madrid, Spain, for DNA extraction and WGS.

TABLE 1 Mtb lineages (L) and sub-lineages and their distribution in PTB and TBLN cases, Northwest Ethiopia, 2023.

Mtb genotypes	TBLN N (%)	PTB N (%)	Total (%)
Delhi-CAS (L3)	21 (56.8)	16 (43.2)	37 (25.3)
Delhi-CAS (L3.1.1)	2 (33.3)	4 (66.7)	6 (4.1)
EA (L4.2)	0 (0.0)	1 (100.0)	1 (0.68)
EA (L4.6)	2 (28.6)	5 (71.4)	7 (4.8)
T (L4.8)	4 (44.4)	5 (55.6)	9 (6.2)
L7	2 (66.7)	1 (33.3)	3 (2.1)
X-type (L4.1.1.1)	0 (0.0)	4 (100.0)	4 (2.7)
X-type (L4.1.1.3)	0 (0.0)	2 (100.0)	2 (1.4)
Haarlem (L4.1.2.1)	9 (45.0)	11 (55.0)	20 (13.7)
Ural (L4.2.1)	3 (75.0)	1 (25.0)	4 (2.7)
EA.ETH (L4.2.2)	12 (41.4)	17 (58.6)	29 (19.9)
TUR (L4.2.2.1)	0 (0.0)	1 (100.0)	1 (0.68)
LAM (L4.3.3)	2 (66.7)	1 (33.3)	3 (2.1)
LAM (L4.3.4.1)	0 (0.0)	1 (100.0)	1 (0.7)
LAM (L4.3.4.2)	5 (100.0)	0 (0.0)	5 (3.4)
EA/T1 (L4.4.1)	3 (75.0)	1 (25.0)	4 (2.7)
EA/T1 (L4.4.1.2)	0 (0.0)	2 (100.0)	2 (1.4)
EA/T3 (4.6.2.1)	1 (12.5)	7 (87.5)	8 (5.5)
Total	66 (45.2)	80 (54.8)	146 (100.0)

2.3. *Mycobacterium tuberculosis* DNA extraction

The DNA extraction was carried out using two methods: manual NZY Tissue gDNA isolation kit (*nzytech genes and enzymes, Lisboa, Portugal*) and robotic Maxwell RSC cultured cells DNA extraction kit (*Promega Biotech Ibérica S.L.*). DNA extraction using NZY tissue kit was performed according to the manufacturer's instructions with the following modifications: heat-inactivated MTBC isolate was centrifuged at 11,000 rpm for 2 min. After removing the supernatant, 180 µl of pre-aliquoted lysis buffer and 60 µl of reconstituted lysozyme (20 mg/ml) were added. The mixture was vortexed and incubated for 3 h at 37°C. Then 75 µl of sodium dodecylbenzenesulfonate (SDS) 10× and 20 µl of proteinase K (20 mg/ml) were added, vortexed, and then incubated overnight at 56°C. The same pre-treatment was applied to the samples before DNA extraction using the robotic Maxwell RSC DNA kit-based extraction. The concentration of double-stranded DNA was measured using a fluorometer (Promega corporation), and the purity was measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific) at a 260/280 wavelength absorbance ratio. A DNA quantity of ≥1 ng/µl was considered a cutoff value for library preparation. Figure 1 summarizes the key methods of the study.

2.4. Library preparation and sequencing

The input DNA library was prepared using Nextera DNA Flex Library Prep Workflow according to the manufacturer's protocol (Illumina, San Diego, CA, USA). Sequencing was carried out using Illumina NovaSeq 6000 technology (Modi et al., 2021).

2.5. Bioinformatics analysis

2.5.1. Sequence read quality control and assembly

Quality control analysis was carried out using fastQC v.0.11.3 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and Trimmomatic v.0.36 software (Bolger et al., 2014) by the Bioinformatics Unit at the ISCIII. Assembly of the reads was performed using SPAdes v.3.8.0 (Bankevich et al., 2012), and the quality of the assembly was evaluated using QUAST software (Gurevich et al., 2013).

2.5.2. Mapping and genotyping

Mapping was performed with MTBseq 1.0.3 (Kohl et al., 2018b) bioinformatic tool using Mtb RefSeq (GCF_003287165.1_ASM328716v1). Briefly, mapping was done using Burrows–Wheeler Aligner–Maximal Exact Match (BWA-MEM) algorithm. Sequence Alignment/Map tools (SAM tools) were used for sorting, indexing, removing putative PCR duplicates, and removing temporary files. Then, Genome Analysis Toolkit (GATK) was used for base call recalibration and realignment of reads around insertions or deletions. Later,

variant calling/discovery and parsing were done for strains and group calling.

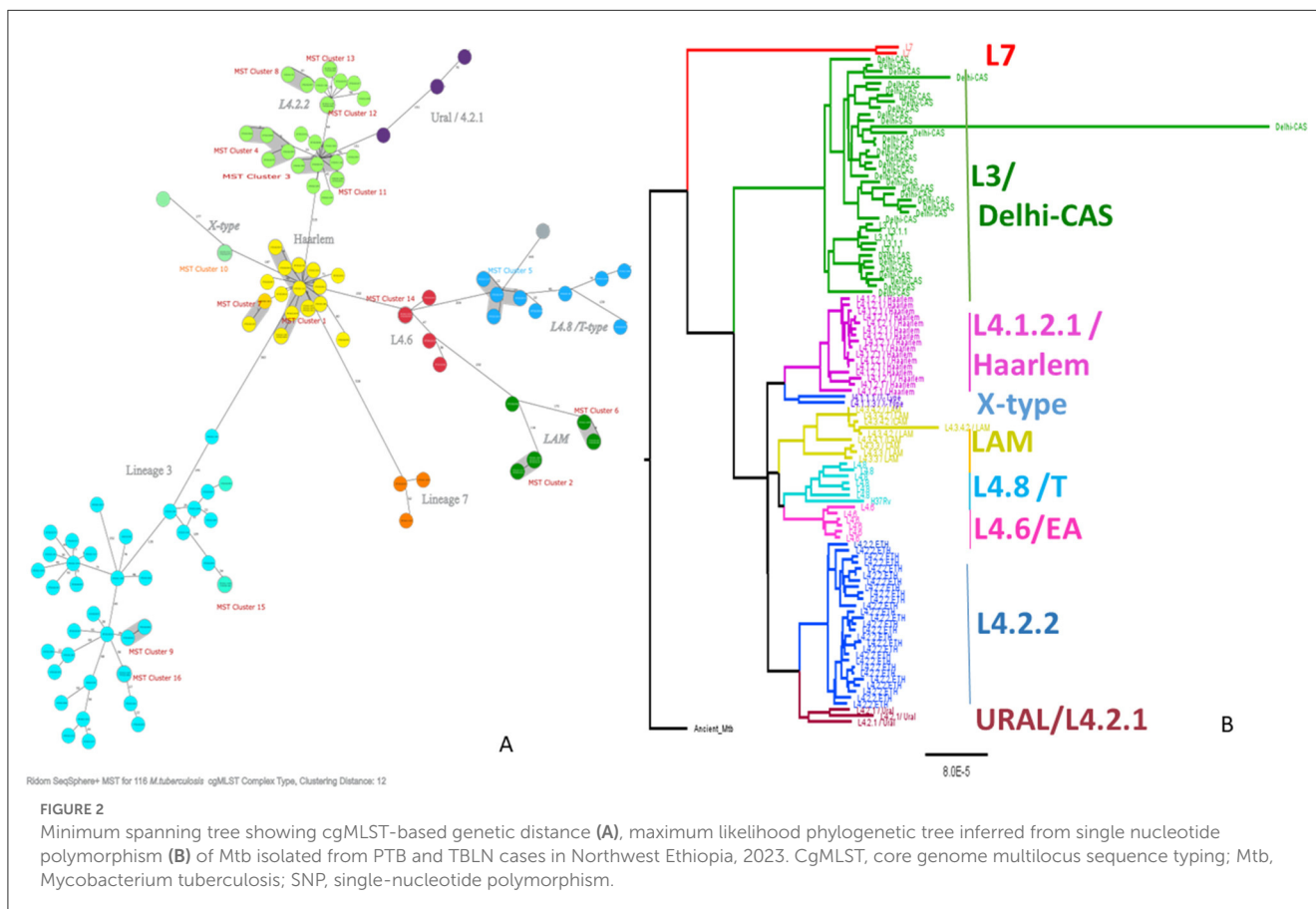
With default settings, variants need to be indicated by four reads mapped in each forward and reverse orientation, respectively, at 75% allele frequency, and by at least four calls with a phred quality score of at least 20. The last sample-specific module enables the phylogenetic classification of the input sample(s) according to Homolka et al. (2012) and Coll et al. (2014) SNP-based typing. In SNP-based barcoding, variant subsets are automatically generated and filtered for repetitive regions and resistance-associated genes (Kohl et al., 2018b).

2.5.3. SNP and cgMLST-based phylogeny

Multiple sequence alignment (MSA) was carried out using multiple alignments using fast fourier transform (MAFFT) (Katoh et al., 2002) against H37Rv (Gen Bank accession number NC_000962.3) which can be assessed at https://www.ncbi.nlm.nih.gov/nuccore/NC_000962.3. Then, the SNP-based phylogenetic reconstruction was carried out using randomized accelerated maximum likelihood (RaxML) using the general time reversal (GTR) model and 1,000 bootstrap value (Stamatakis, 2006). Ancient *Mtb* strain was used as a root for the tree (Supplementary fasta file). Finally, the generated tree was edited using Figtree (<https://github.com/rambaut/figtree/releases>).

The cgMLST uses alleles as the unit of comparison, rather than nucleotide sequences. In allele-based comparisons among isolates, each allelic change is counted as a single genetic event, regardless of the number of nucleotide polymorphisms involved (Maiden et al., 2013). Assembled genomes were uploaded onto the Ridom SeqSphere software (version 8.5; Ridom GmbH, Münster, Germany). Each isolate sequence was aligned to the Ridom SeqSphere *Mtb* core genome scheme of 2,891 core genes which was obtained using the H37Rv reference genome (NC 000962.3) (Kohl et al., 2014). Two genes overlapping by more than four bases and genes with an internal stop codon in more than 80% of all query genomes were removed. Furthermore, repetitive genes such as PPE/PE-PGRS gene families were manually excluded from the scheme (Kohl et al., 2014). The result was generated as a minimum spanning tree and then edited using the Inkscape online tool v1.2 (<https://inkscape.org/release/inkscape-1.2.2/>).

Clustering rate (CR) and RTI were calculated from the minimum spanning tree complexes. A cluster was defined as two or more *Mtb* isolates differing with ≤ 5 alleles at core genomes (Gardy et al., 2011; Walker et al., 2013; and Kohl et al., 2018a). Isolates differing with ≤ 5 alleles at core genomic regions were considered transmission linked cases. The clustering rate and RTI were calculated using the following formula: $CR\% = (n/N) \times 100$, $RTI\% = [(n - N_c)/N] \times 100$, where N is the total isolates in the sample, n is the total number of isolates within the clusters, and N_c is the number of clusters (Tanaka and Francis, 2005; Tessema et al., 2013).



2.6. Statistical analysis

Descriptive statistics, such as mean, median, standard error (Std.error), and interquartile range (IQR), were used to summarize the evolutionary events (SNP, deletion, insertion, and substitution) of genotypes. Furthermore, an independent sample *t*-test was used to assess the correlation between TB forms (TBLN and PTB) and evolutionary events. Logistic regression analysis was carried out to identify factors associated with forms of TB (TBLN vs. PTB) and clustering (cluster vs. singleton). The analyses were carried out using R software version 4.2.2 (R Core Team, 2022) and the SPSS statistical software version 25 (IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp).

3. Results

Whole-genome sequencing was performed for 161 Mtb isolates using Illumina NovaSeq 6000 technology. The overall sequence read quality was good. The sequence read length varied from 35 to 151 nucleotides (mean 148). The mean and median sequencing coverage were 267 and 201 times, respectively. The

mean (Std.error) and median (interquartile range/IQR) of SNP events were 1,163.62 (32.795) and 1,027.0 (498), respectively. The maximum deletion event among isolates was 610, and the minimum number of deletion was 114. The highest SNP, deletion, insertion, and substitution events were observed in L3 and L7 isolates than in L4 sub-lineages (Supplementary Table 1). An independent sample *t*-test was carried out to compare the mean evolutionary events between PTB and TBLN isolates. The analysis did not reveal any significant correlation ($p > 0.05$, Supplementary Table 2).

3.1. Genetic diversity of *Mycobacterium tuberculosis*

According to Homolka et al. (2012) and Coll et al. (2014) SNP-based typing, 146 isolates were genotyped and classified within three lineages and 18 sub-lineages. Lineage 4 accounted for 100 (68.5%) of the 146 total isolates, while L3 accounted for 43 (25.3%). The remaining 3 (2.05%) isolates were shared by Lineage 7. Furthermore, L4.2.2/L4.2.2.ETH/sub-lineage was estimated to

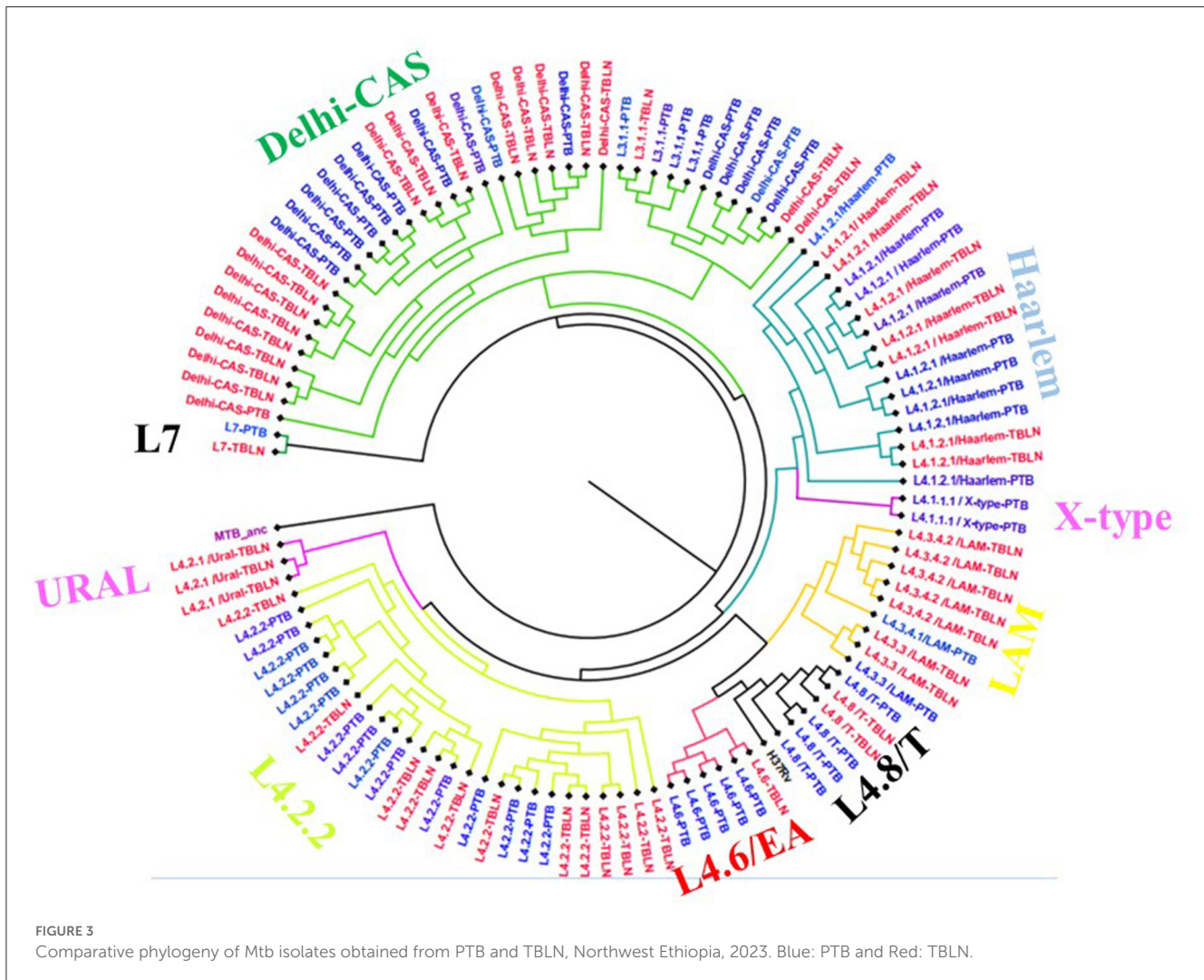


FIGURE 3 Comparative phylogeny of Mtb isolates obtained from PTB and TBLN, Northwest Ethiopia, 2023. Blue: PTB and Red: TBLN.

account for 19.9%, while the Haarlem/L4.1.2.1 accounted for 13.7%. The proportion and types of Mtb lineages and sub-lineages were comparable and/or the same between PTB and TBLN (Table 1). Briefly, from the total 37 Delhi-CAS sub-lineages, 21 (56.8%) and 16 (43.2%) Delhi-CAS sub-lineages were obtained from TBLN and PTB, respectively. Similarly, from 29 L4.2.2.ETH sub-lineages, 17 (58.6%) were from PTB and the rest 12 (41.4%) were from TBLN (Table 1; Figure 2).

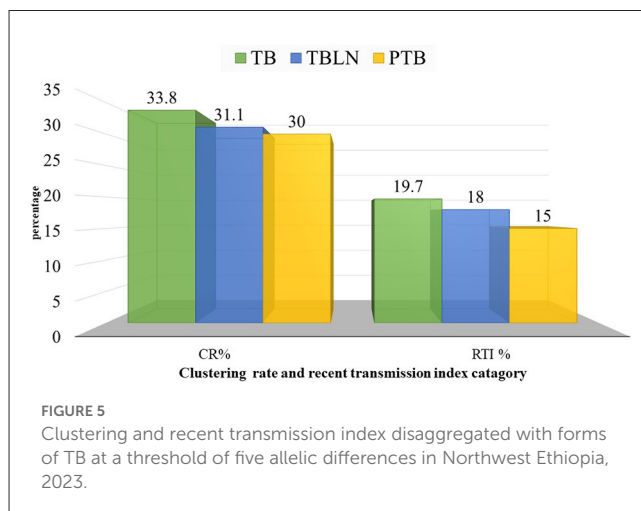
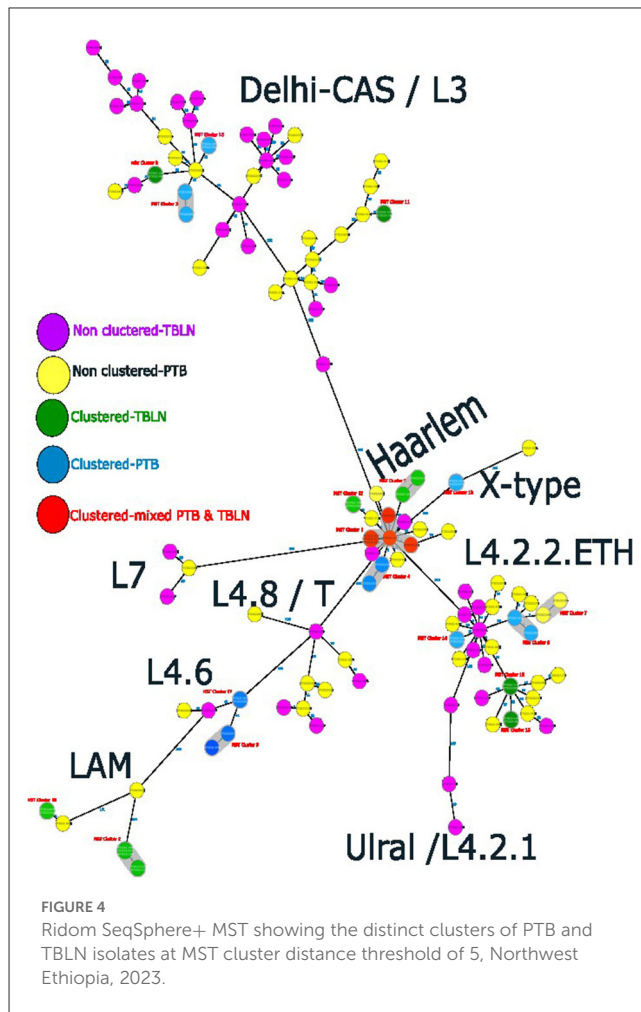
As shown in Figure 2, L4.2.2, which was classified as L4.2.ETH (Comas et al., 2015), and spoligotyping international type (SIT149) (Firdessa et al., 2013; Yimer et al., 2015), is considered a specialist isolate in Ethiopia, formed two major groups in the tree. The Haarlem/L4.1.2.1 sub-lineages were positioned in the middle of the tree and contained the biggest cluster (cluster size: 13 isolates). Unlike EA sub-lineages, the population structure of L3 and L7 was deep-rooted, complex, and less clustered (Figure 2).

The comparative phylogenetic tree of Mtb isolates obtained from TBLN and PTB is depicted in Figures 3, 4. Surprisingly, unlike previous reports, these figures showed distinct Mtb genotypic clusters among PTB and TBLN isolates in Northwest Ethiopia. While several identical Mtb genotypes (isolates differing with <5 alleles) cases were found from each of TBLN and PTB separately, a few identical Mtb isolates were identified from both PTB and TBLN (Figures 3, 4; Supplementary Table 3). A closer look at Figure 4; Supplementary Table 3 showed that PTB and TBLN shared one Haarlem genotype which is positioned centrally.

3.2. Transmission dynamics of *Mycobacterium tuberculosis*

At a threshold of five allelic distance, the cgMLST classified the 127 Mtb isolates into 84 cgMLST complex types, 84 unique and 18 clusters (Nc) containing 43 isolates (cluster size: 2–6 isolates). Thus, the overall CR and RTI were gauged at 33.8 and 19.7% (Figure 4; Supplementary Table 3), respectively. Furthermore, CR and RTI were calculated separately for PTB and TBLN. Hence the CR and RTI for PTB were 30 and 15% (number of isolates within the cluster = 18, Nc = 9, and N = 60), respectively. Similarly, the CR and RTI for TBLN were 31.1 and 18% (n = 19, Nc = 8, and N = 61), respectively. The second most surprising aspect of this study was the high RTI of TBLN (Figures 4, 5). Figure 5 depicts that the RTI of TBLN is unexpectedly high. The RTI is an indicator of epidemic severity, and it is the measure of the progression of clinical diseases from primary infection. The high RTI of TBLN implies that the TBLN is the result of recent transmission or reactivation from short latency. The higher proportion of TBLN among students (p > 0.05) was another evidence that the higher incidence of TBLN in Northwest Ethiopia is the result of recent transmission (Table 2).

Taken together, the Mtb isolates obtained from PTB and TBLN form distinct clusters which indicated the presence of strain genetic variation between the two clinical forms. Additionally, unlike the common belief that TBLN is mainly the result of reactivation of latent TB, this data demonstrated that TBLN is the result of recent transmission and/or rapid progression from short latency among immunocompetent adults.



The logistic regression analysis was computed using data from newly diagnosed PTB and TBLN cases (Group I: participants from Figure 1). The goodness of fit test, which is the measure of the ability of a model to generate high-quality predictions of the multivariable analysis, was assessed using the pseudo R² and the Hosmer and Lemeshow test. It was found that the p-value

TABLE 2 Logistic regression analysis TB forms with Mtb lineages and sub-lineages and socio-demographic factors, Northwest Ethiopia, 2023.

Variables	TB forms		P-value of COR	AOR (95%CI) P-value	
	TBLN	PTB			
Age groups (year)	≥65	3 (4.5)	0 (0)	1.00	–
	15–64	54 (81.8)	41 (100)	0.99	–
	1–14	9(13.6)	0 (0)	1	–
Gender	Female	33 (50.0)	19 (46.3)	0.713	0.698 (0.220–2.214) 0.541
	Male	33 (50.0)	22 (53.7)	1	1
Residency	Rural	38 (58.5)	12 (29.3)	0.004	3.508 (1.085–11.40) 0.034
	Urban	27 (41.5)	29 (70.7)	1	1
Marriage	Single	24 (38.7)	13 (40.6)	0.611	1.103 (0.199–6.120) 0.911
	Married	30 (48.4)	13 (40.6)	0.387	0.436 (0.097–1.963) 0.279
	Divorced and widowed	8 (12.9)	6 (18.8)	1	1
Current job	Pre-schooled children	2 (3.1)	0 (0)	0.99	–
	Student	20 (31.3)	5 (15.6)	0.039	0.217 (0.038–1.232) 0.085
	HW/farmer	24 (37.5)	8 (25.0)	0.07	0.907 (0.186–4.414) 0.903
	HRE	8 (12.5)	9 (28.1)	0.858	1.089 (0.215–5.521) 0.918
	Job-finder	10 (15.6)	10 (31.3)	1	1
Sub-lineages	EA	7 (10.6)	7 (16.7)	1.00	0.955 (0.137–6.642) 0.963
	L4.2.2.ETH	10 (15.2)	6 (14.3)	0.455	0.885 (0.128–6.124) 0.902
	Delhi- CAS	23 (34.8)	9 (21.4)	0.115	0.196 (0.036–1.075) 0.061
	Haarlem	9 (13.6)	8 (19.0)	0.858	0.939 (0.154–5.730) 0.946
	LAM	7 (10.6)	2 (4.8)	0.172	0.102 (0.006–1.619) 0.106
	Other	10 (15.2)	10 (23.8)	1	1

Bold values are either statistically significant or are higher within TBLN despite statistical insignificance.

of pseudo R^2 and Hosmer and Lemeshow test were 0.54 and 0.797, respectively. Additionally, the percent correct classification of the model became 80.6%. These results indicated that the model was fairly good. Hence, L4.2.2.ETH with AOR: 7.42, 95%CI: 1.050–52.456 and a p -value of 0.045 and Haarlem with AOR: 27.37, 95%CI: 3.337–224.507 and a p -value of 0.002 demonstrated significant association with clustering. The two forms of TB (PTB vs. TBLN) did not show a significant association with clustering ($p > 0.05$). Moreover, other variables, such as gender, age, residency, and marital status, were not associated with Mtb clustering (Table 3).

4. Discussion

A few studies used the WGS technique to determine the genomic diversity and transmission dynamics of Mtb in Ethiopia (Comas et al., 2015; Yimer et al., 2016; and Welekidan et al., 2021). However, the aims of these studies were different from each other and also different from the present study. This study assessed and compared the genomic diversity and transmission dynamic of Mtb isolates obtained from PTB and TBLN cases in Northwest Ethiopia.

The WGS data unveiled a high genetic diversity of the Mtb population, L4 being the predominant (100/146, 68.5%),

followed by L3 (29.45%). While the LAM and Haarlem isolates showed a low rate of SNP, insertion, substitution, and deletion, the L3 and L7 isolates demonstrated a high rate of mutations (Supplementary Table 1). The difference might be due to the nature (sympatric vs. allopatric) and timing of the pathogens–host association. When the specific host and Mtb coevolve together for a long, the host presents constant metabolic supplies. Thus, the metabolic genes of the Mtb become rendered useless and deleted (Ochman and Moran, 2001; Moran, 2002) as a way to cellular economization (Martínez-Cano et al., 2015).

The comparative Mtb genetic diversity showed similar lineages and sub-lineages, but it showed distinct genomic clusters among the two clinical forms (PTB vs. TBLN). The clusters became mixed of TBLN and PTB isolates at a higher threshold of allelic difference (>5 allelic difference). Faksri et al.'s (2018) study found genetic variants of Mtb that are commonly found in TB meningitis patients compared to PTB cases. To further support our finding, we re-evaluated the TBLN geographic distributions mapped in the previous studies at Africa (Mekonnen et al., 2019b), Ethiopia (Mekonnen et al., 2019b), and Amhara Regional State level (Mekonnen et al., 2022). Figures 6A, B shows the geographic pattern of TBLN similar to Mtb geographical structuring. Furthermore, we assessed the Mtb lineages and sub-lineages distribution among countries that had TBLN reports

TABLE 3 Logistic regression for clustering status with socio-demographic factors and Mtb sub-lineages, Northwest Ethiopia, 2023.

Variables		Clustering status		P-value of COR	AOR (95% CI) P-value
		Unique N (%)	Clustered N (%)		
Age category (year)	1–14	5 (55.6)	4 (44.4)	0.99	–
	15–64	55 (57.9)	40 (42.1)	0.99	–
	≥65	3 (100)	0 (0)	1	1
Gender	Female	32 (61.5)	20 (38.5)	0.587	1.43 (0.413–4.985) 0.570
	Male	31 (56.4)	24 (43.6)	1	1
TB types	PTB	38 (57.6)	28 (42.4)	0.841	0.81 (0.233–2.847) 0.748
	TBLN	25 (59.5)	17 (40.5)	1	1
Residential status	Urban	33 (58.9)	23 (41.1)	0.911	0.51 (0.139–1.914) 0.322
	Rural	30 (60)	20 (40)	1	1
Marital status	Single	17 (45.9)	20 (54.1)	0.477	0.68 (0.113–4.047) 0.668
	Married	26 (60.5)	17 (39.5)	0.826	0.77 (0.147–3.994) 0.752
	Divorced and widowed	8 (57.1)	6 (42.9)	1	1
Current work	Pre-school	0 (0)	2 (100)	0.99	–
	Student	12 (48)	13 (52)	0.894	1.35 (0.223–8.208) 0.743
	HW/farmer	21 (65.5)	11 (34.4)	0.266	0.22 (0.034–1.410) 0.110
	HRE	10 (58.8)	7 (41.2)	0.592	0.83 (0.158–4.362) 0.826
	Job-finder	10 (50)	10 (50)	1	1
Mtb sub-lineage	EA	11 (78.6)	3 (21.4)	0.809	1.29 (0.160–10.434) 0.811
	L4.2.2.ETH	7 (43.8)	9 (56.3)	0.061	7.42 (1.050–52.456) 0.045
	CAS	26 (81.3)	6 (18.8)	0.592	1.54 (0.270–8.736) 0.628
	Haarlem (4.1.2.1)	3 (17.6)	14 (82.4)	0.001	27.37 (3.337–224.507) 0.002
	LAM	1 (11.1)	8 (88.9)	0.007	–
	Other	15 (75)	5 (25)	1	1
Total	63 (58.3)	45 (41.7)			

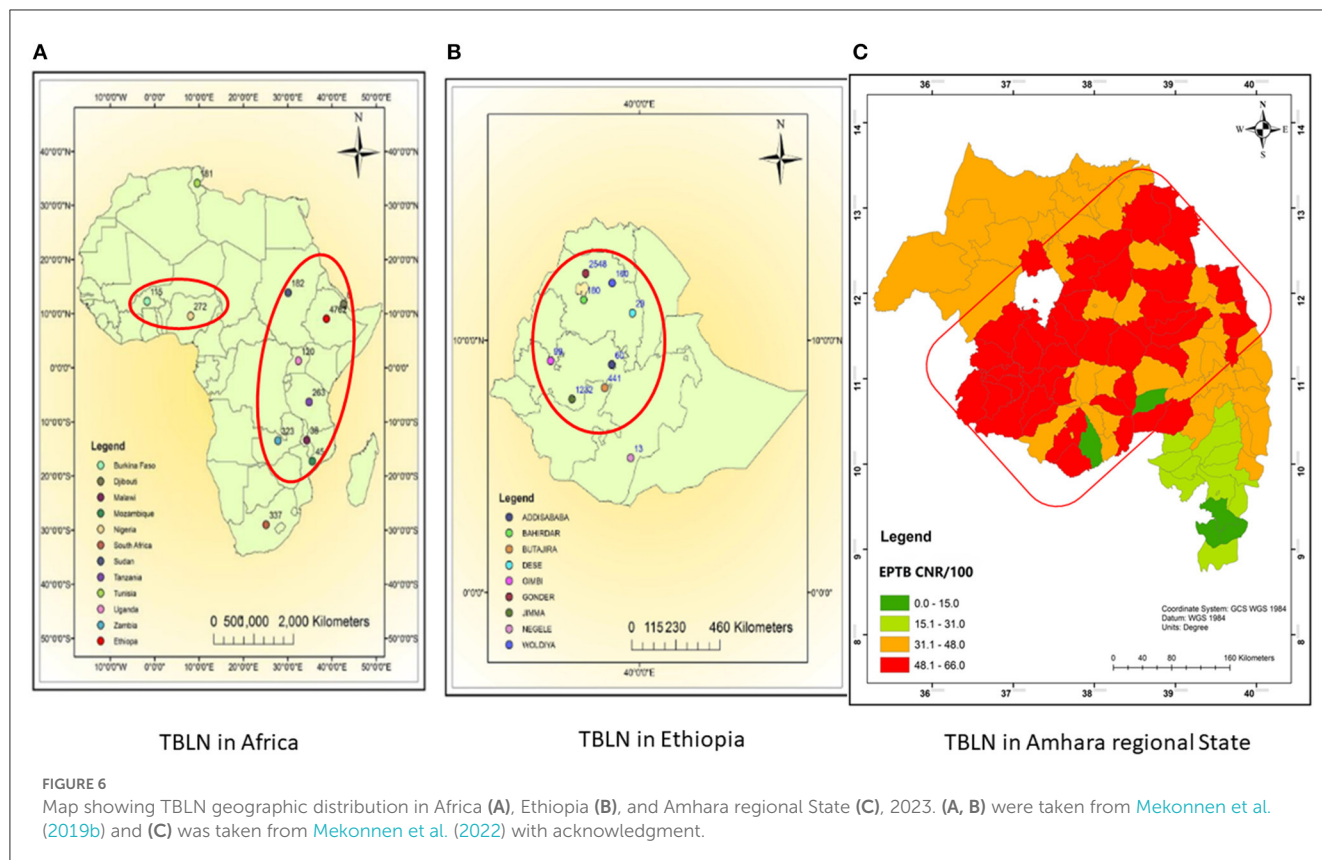
Mtb, *Mycobacterium tuberculosis*; COR, crude odds ratio; AOR, adjusted odds ratio; PTB, pulmonary tuberculosis; TBLN, Tuberculosis lymphadenitis; HRE, high-risk employee; HW, housewife; EA.ETH, Euro-American lineage restricted to Ethiopia; LAM, Latin America- Mediterranean; T, Tuscany; Delhi-CAS, Delhi-Central Asia; Other, T, X, Ural, Uganda. Bolded variables are statistically significant variables.

(Mekonnen et al., 2019a). As such, TBLN was mainly reported from West and East African countries, having a complex MTBC population genetic structure and high prevalence of Mtb Genotypes that have narrow host ranges (specialist genotypes). For instance, in West Africa, *L5/6/Maf* and *L4.6.2/LAM-CAM* are dominant specialist lineages and sub-lineages, respectively. Similarly in East Africa, *L4.6/SIT37*, *L4.2.2/SIT149*, *L4.10/SIT53*, *L4.6.1/T2_Uganda*, and several CAS sub-lineages are prevalent (Mekonnen et al., 2019a). Hence, the high incidence rate of TBLN in Ethiopia and other Eastern and Western African countries is likely the result of Mtb genomic variation. Negrete-Paz et al. (2021) also demonstrated intra-(sub)-lineage differences among TB clinical phenotypes. They briefly stated that strains of a sub-lineage vary in terms of immune response and virulence. Hence, these variations perhaps affect disease severity and clinical presentation (Negrete-Paz et al., 2021). Tuberculous lymphadenitis is the less severe form of TB (Hasan et al., 2009). It is known that long-established host–parasite coevolution leads to benign disease with

insidious onset (May and Anderson, 1983). Further computational co-evolution model intuition could prove this scenario if it works for TBLN.

Unlike other zones of Amhara Regional State shown in Figure 6C, the low incidence rate of TBLN in the North Shewa Zone (Figure 6C light green color) could be due to environmental and host genetic differences besides Mtb genomics. Unlike other zones of Amhara, North Shewa is high land with cold weather.

Perhaps the other most compelling finding of this study is the higher CR and RTI in TBLN cases which are even slightly higher compared with PTB cases ($p > 0.05$). This is direct evidence showing that the high incidence rate of TBLN in the Amhara region is not just a result of the reactivation of latent TB. Rather, the finding indicated the rapid progression of clinical illness from primary infection or short latency. The high incidence rate of TBLN in younger age people, such as students (Table 2), was another evidence of the short latency nature of TBLN in Amhara Regional State. This rapid clinical progression of TBLN from primary



infection was not due to immunodeficiency. This is because the immunohematological values of TBLN are relatively higher among TBLN compared with PTB (Mekonnen et al., 2023). Rather, it might be due to immunological tolerance resulting from sympatric-host pathogen association (Seal et al., 2021).

Due to the poor health system and socio-demographic factors, unrecognized extensive TB transmission might have taken place in Ethiopia for a long period. Such hidden extensive micro epidemics could probably push the co-evolution toward immune tolerance and then TBLN clinical phenotype. Host controls Mtb infections by activating innate and adaptive cellular immunities. Immune response against Mtb causes immunopathology in both the host and the pathogen. Hence, the host's ability to invest in self-toxic immune responses might have a limit. Beyond this threshold, the host switches from active resistance to tolerance strategy. Furthermore, tolerance to invading pathogens also has a threshold. Hence, the host's immune response strategy against Mtb is perhaps fine-tuned by the fitness effects of both the host and Mtb. This ability of the host for maintaining a balanced immunity and homeostasis during infection results in less severe forms of TB, such as TBLN, smear-negative PTB, lower cavitation, and negative chest X-ray finding (Seal et al., 2021). Taken together, the insights gained from this study may be of assistance to re-evaluate the current prevailing dogma which says- "no strain difference between PTB and TBLN isolates."

L4.2.2.ETH forms two major branches containing a mix of clustered and unique strains (Figure 2). The L4.2.2.ETH topology in the cgMLST and SNP phylogeny is similar to Comas et al.'s (2015) phylogenetic tree. The phylogenetic position of L4.2.2.ETH is close

to the Ural family, indicating their introduction into Ethiopia from Central Asian states (Mokrousov, 2012). Compared with L4.6, the arrival of L4.2 into Ethiopia is likely a recent phenomenon and possibly linked with the major north-south human migration during the reign of the Queen of Sheba (Comas et al., 2015). The LAM (L4.3.4.1 & L4.3.4.2) sub-lineages showed low deletion and substitution rates in this study. LAM is considered a generalist sub-lineage that is found in over 47 countries and highly prevalent in Europe with high genetic diversity. Its global success was likely driven by European migration and colonization (Stucki et al., 2016). The other highly clustered sub-lineage was Haarlem/L4.1.2.1 which is found in over 49 countries (Stucki et al., 2016). This sub-lineage is estimated to be introduced into Ethiopia at the time of early Portuguese contact with Ethiopia in the 16th century or later (Comas et al., 2015).

Lineage 3 is the second leading genotype found in this study, and this report is in line with previous studies (Tessema et al., 2013; Biadlegne et al., 2015; Mekonnen et al., 2019c; and Welekidan et al., 2021). There is a high genomic substitution, deletion, or SNP in L3 similar to L7. Our phylogenetic tree showed that the genetic diversity of L3 is very complex and likely shaped by several independent evolutionary events or multiple entries into Ethiopia (Figure 3). The L3 contains both specialist (isolates restricted in some countries including Ethiopia) and generalist genotypes (Comas et al., 2015; Tulu and Ameni, 2018). With a 90.2% probability, South Asia was predicted as the origin of all L3 strains with multiple (at least four) independent introductions into East and North Africa (Shuaib et al., 2022). L7 was restricted to Ethiopia and branched off around the time of initial human

migrations out of Africa (Comas et al., 2015). Collectively, except L7, the other (sub) lineages followed the human migration; out of Africa (northern route via Egypt) (Pagani et al., 2015) and back migration into Africa in the different demographic histories (Pankhurst, 1964; Molinaro et al., 2019). Taken together, the population genomic structure of Mtb in Ethiopia is likely driven by long and stable host-Mtb reciprocal evolutionary changes (May and Anderson, 1983; Freschi et al., 2021) and consecutive entry of new variants via trade, missionaries, and human migration (Curtis, 2008; Comas et al., 2015). In addition to that, the poverty and high population density might have contributed to local adaptation and expansion of newly imported Mtb.

5. Strength and limitations

This WGS study gave high-resolution information about the role of Mtb genotypes in the clinical presentation of TB, genomic diversity, and transmission dynamics. However, it has some limitations. This study was carried out at the institutional level and included patients coming from different zones. Thus, clustered isolates did not mean they are epidemiologically linked, rather, it might be due to the widespread availability of the genotypes in the regional state. Similarly, the unique isolates might not be the result of reactivation, rather it might have been linked to cases in the community where the patient came. Hence, there might be a sampling bias. Furthermore, the sample size was small which precluded us from the in-depth assessment of the socio-demographic-related variables.

6. Conclusion and recommendations

TBLN and PTB showed no difference at lineage and sub-lineage levels. However, the isolates from the two clinical forms formed a distinct cluster, indicating the presence of pathogen genomic factors. The higher rate of clustering and RTI among TBLN confirmed the rapid progression of TBLN from primary infection and/or short latency. Hence, the high incidence rate of TBLN in Amhara Regional State is likely shaped by Mtb genetic diversity and rapid progression from primary infection. Comparative community-based studies containing data from the environment and host genome could give a detailed view. Additionally, the characterization of lineage-independent genetic components might give further insight.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NCBI, PRNJA975069.

Ethics statement

The study was approved by research and Ethical Review Committee of Science College of Bahir Dar University, with reference number PGRCSV/111/2012. Informed written consent was obtained from each participant before data collection. This

study was conducted in accordance with the Declaration of Helsinki. All the information obtained from the study subjects were coded to maintain confidentiality. The patients/participants provided their written informed consent to participate in this study.

Author contributions

DM: substantial contributions to the conception of the work, data collection, laboratory works, analyses, interpretation of data for the work, and drafting of the work. AM and EN: substantially contributed to the conception, methodology, validation, monitoring of the work, secured partial fund for the work, and revising it critically for important intellectual content. BA: substantially contributed to the bioinformatics analysis and revised it critically for important intellectual content. AAA and AB: substantially contributed to supervision, monitoring of the laboratory work, and revising it critically for important intellectual content. EA and SH-L: participated in the DNA extraction, data analysis and interpretation, and revising the draft critically for important intellectual content. CJ: prepared the sequencing library, substantially contributed to the data analyses, interpretation of data for the work, and revised the draft critically for important intellectual content. AA: substantially contributed to the conception, methodology, and revising the draft critically for important intellectual content. LH-L: monitoring and coordinating the DNA extraction and WGS, substantially contributed to the bioinformatics, and revising the draft critically for important intellectual content. All authors critically revised and approved the version to be published.

Funding

The sample collection was funded by the Institute of Biotechnology, Bahir Dar University through the EN mega project. The Mtb culture and identification-related lab supply were supported by Amhara Public Health Institute, Bahir Dar Ethiopia. The whole-genome sequencing (WGS) and publication fee was covered by the National Center of Microbiology, Institute of Carlos III, Madrid, Spain. International Federation for Clinical Chemistry (IFCC) gave financial support to DM through the IFCC Professional Scientific Exchange Programme (PSEP) for 3-month WGS laboratory work.

Acknowledgments

We would like to express our gratitude to the Institute of Biotechnology and the Institute of Carlos III for administrative support. We thank the Amhara Public Health Institute Mycobacteriology Laboratory Staff, Felege Hiwote Referral Hospital DOTS clinic physicians, and Pathology laboratory Team, DOTS clinic staff of Han, Bahir Dar, and Shum-Abo Health Centers for their support in the data collection. We also thank

the Institute of Carlos III Mycobacteriology Laboratory staff: Martha and Rackel for their support during DNA extraction and WGS. We are extremely grateful to International Federation for Clinical Chemistry (IFCC) for giving a 3-month travel grant scholarship for DM for conducting the whole-genome sequence of *Mycobacterium tuberculosis* isolates at the National Center of Microbiology, Majadahonda, Spain through IFCC Professional Scientific Exchange Programme (PSEP).

Conflict of interest

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

References

- Abel, L., El-Baghdadi, J., Bousfiha, A. A., Casanova, J.-L., and Schurr, E. (2014). Human genetics of tuberculosis: a long and winding road. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130428. doi: 10.1098/rstb.2013.0428
- Adhikari, B. (2022). History and evolution of tuberculosis and global health. *Lancet Infect. Dis.* 22, 462. doi: 10.1016/S1473-3099(22)00157-8
- Ameni, G., Vordermeier, M., Firdessa, R., Aseffa, A., Hewinson, G., Gordon, S. V., et al. (2011). *Mycobacterium tuberculosis* infection in grazing cattle in central Ethiopia. *Vet. J.* 188, 359–361. doi: 10.1016/j.tvjl.2010.05.005
- Asres, A., Jerene, D., and Deressa, W. (2019). Delays to anti-tuberculosis treatment initiation among cases on directly observed treatment short course in districts of southwestern Ethiopia: a cross sectional study. *BMC Infect. Dis.* 19, 1–9. doi: 10.1186/s12879-019-4089-x
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Barberis, I., Bragazzi, N. L., Galluzzo, L., and Martini, M. (2017). The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus. *J. Prev. Med. Hyg.* 58, E9.
- Berg, S., Schelling, E., Hailu, E., Firdessa, R., Gumi, B., Erenso, G., et al. (2015). Investigation of the high rates of extrapulmonary tuberculosis in Ethiopia reveals no single driving factor and minimal evidence for zoonotic transmission of *Mycobacterium bovis* infection. *BMC Infect. Dis.* 15, 112. doi: 10.1186/s12879-015-0846-7
- Biadlegne, F., Merker, M., Sack, U., Rodloff, A. C., and Niemann, S. (2015). Tuberculous lymphadenitis in Ethiopia predominantly caused by strains belonging to the Delhi/CAS lineage and newly identified Ethiopian clades of the *Mycobacterium tuberculosis* complex. *PLoS ONE* 10, e0137865. doi: 10.1371/journal.pone.0137865
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bomanji, J., Sharma, R., Mittal, B. R., Gambhir, S., Qureshy, A., Begum, S. M., et al. (2020). PET/CT features of extrapulmonary tuberculosis at first clinical presentation: a cross-sectional observational 18F-FDG imaging study across six countries. *Eur. Respir. J.* 55. doi: 10.1183/13993003.01959-2019
- Coll, F., Mc Nerney, R., Guerra-Assunção, J. A., Glynn, J. R., Perdigão, J., Viveiros, M., et al. (2014). A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* 5, 1–5. doi: 10.1038/ncomms5812
- Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., et al. (2013). Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat. Genet.* 45, 1176–1182. doi: 10.1038/ng.2744
- Comas, I., Hailu, E., Kiros, T., Bekele, S., Mekonnen, W., Gumi, B., et al. (2015). Population genomics of *Mycobacterium tuberculosis* in Ethiopia contradicts the virgin soil hypothesis for human tuberculosis in Sub-Saharan Africa. *Curr. Biol.* 25, 3260–3266. doi: 10.1016/j.cub.2015.10.061
- Coscolla, M., Gagneux, S., Menardo, F., Loiseau, C., Ruiz-Rodriguez, P., Borrell, S., et al. (2021). Phylogenomics of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. *Microb. Genome* 7, 000477. doi: 10.1099/mgen.0.000477
- Curtis, M. C. (2008). *Changing Settlement Patterns in the Aksum-Yeha Region of Ethiopia: 700 BC-AD 850*. Boston, MA: JSTOR.
- de Jong, B. C., Antonio, M., and Gagneux, S. (2010). *Mycobacterium africanum*-review of an important cause of human tuberculosis in West Africa. *PLoS Negl. Trop. Dis.* 4, 0000744. doi: 10.1371/journal.pntd.0000744
- Deveci, H. S., Kule, M., Kule, Z. A., and Habesoglu, T. E. (2016). Diagnostic challenges in cervical tuberculous lymphadenitis: a review. *North. Clin. Istanbul* 3, 150. doi: 10.14744/nci.2016.20982
- Ejo, M., Torrea, G., Uwizeye, C., Kassa, M., Girma, Y., Bekele, T., et al. (2021). Genetic diversity of the *Mycobacterium tuberculosis* complex strains from newly diagnosed tuberculosis patients in Northwest Ethiopia reveals a predominance of East-African-Indian and Euro-American lineages. *IJID* 103, 72–80. doi: 10.1016/j.ijid.2020.11.129
- Faksri, K., Xia, E., Ong, R. T., Tan, J. H., Nonghanphithak, D., Makhao, N., et al. (2018). Comparative whole-genome sequence analysis of *Mycobacterium tuberculosis* isolated from tuberculous meningitis and pulmonary tuberculosis patients. *Sci. Rep.* 8, 4910. doi: 10.1038/s41598-018-23337-y
- Firdessa, R., Berg, S., Hailu, E., Schelling, E., Gumi, B., Erenso, G., et al. (2013). Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. *EID* 19, 460. doi: 10.3201/eid1903.120256
- Freschi, L., Vargas Jr, R., Husain, A., Kamal, S. M., Skrahina, A., Tahseen, S., et al. (2021). Population structure, biogeography and transmissibility of *Mycobacterium tuberculosis*. *Nat. Commun.* 12, 6099. doi: 10.1038/s41467-021-26248-1
- Gagneux, S. (2018). Ecology and evolution of *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* 16, 202–213. doi: 10.1038/nrmicro.2018.8
- Ganchua, S. K. C., White, A. G., Klein, E. C., and Flynn, J. L. (2020). Lymph nodes—The neglected battlefield in tuberculosis. *PLoS Pathog.* 16, e1008632. doi: 10.1371/journal.ppat.1008632
- Garby, J. L., Johnston, J. C., Sui, S. J. H., Cook, V. J., Shah, L., Brodtkin, E., et al. (2011). Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *NEJM* 364, 730–739. doi: 10.1056/NEJMoa1003176
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Hasan, Z., Cliff, J. M., Dockrell, H. M., Jamil, B., Irfan, M., Ashraf, M., et al. (2009). CCL2 responses to *Mycobacterium tuberculosis* are associated with disease severity in tuberculosis. *PLoS ONE* 4, e8459. doi: 10.1371/journal.pone.0008459
- Herath, S., and Lewis, C. (2014). Pulmonary involvement in patients presenting with extra-pulmonary tuberculosis: thinking beyond a normal chest x-ray. *J. Prim. Health Care* 6, 64–68. doi: 10.1071/HC14064
- Homolka, S., Projahn, M., Feuerriegel, S., Ubben, T., Diel, R., Nübel, U., et al. (2012). High resolution discrimination of clinical *Mycobacterium tuberculosis* complex strains based on single nucleotide polymorphisms. *PLoS ONE* 7, e39855. doi: 10.1371/journal.pone.0039855
- Iwnetu, R., Van Den Hombergh, J., Woldeamanuel, Y., Asfaw, M., Gebrekirstos, C., Negussie, Y., et al. (2009). Is tuberculous lymphadenitis over-diagnosed in Ethiopia? Comparative performance of diagnostic tests for mycobacterial lymphadenitis in a high-burden country. *Scand. J. Infect. Dis.* 41, 462–468. doi: 10.1080/00365540902897697

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1211267/full#supplementary-material>

- Katoh, K., Misawa, K., Kuma, K. i., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucl. Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kohl, T. A., Diel, R., Harmsen, D., Rothgänger, J., Walter, K. M., Merker, M., et al. (2014). Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. *J. Clin. Microbiol.* 52, 2479–2486. doi: 10.1128/JCM.00567-14
- Kohl, T. A., Harmsen, D., Rothgänger, J., Walker, T., Diel, R., and Niemann, S. (2018a). Harmonized genome wide typing of tubercle bacilli using a web-based gene-by-gene nomenclature system. *EBioMedicine* 34, 131–138. doi: 10.1016/j.ebiom.2018.07.030
- Kohl, T. A., Utpatel, C., Schleusener, V., De Filippo, M. R., Beckert, P., Cirillo, D. M., et al. (2018b). MTBseq: a comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex isolates. *PeerJ* 6, e5895. doi: 10.7717/peerj.5895
- Maiden, M. C., Van Rensburg, M. J. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A., et al. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat. Rev. Microbiol.* 11, 728–736. doi: 10.1038/nrmicro3093
- Martínez-Cano, D. J., Reyes-Prieto, M., Martínez-Romero, E., Partida-Martínez, L. P., Latorre, A., Moya, A., et al. (2015). Evolution of small prokaryotic genomes. *Front. Microbiol.* 5, 742. doi: 10.3389/fmicb.2014.00742
- May, R. M., and Anderson, R. M. (1983). Epidemiology and genetics in the coevolution of parasites and hosts. *Proc. R. Soc. Lond. B Biol. Sci.* 219, 281–313. doi: 10.1098/rspb.1983.0075
- Meehan, C. J., Moris, P., Kohl, T. A., Pečerska, J., Akter, S., Merker, M., et al. (2018). The relationship between transmission time and clustering methods in *Mycobacterium tuberculosis* epidemiology. *EBioMedicine* 37, 410–416. doi: 10.1016/j.ebiom.2018.10.013
- Mekonnen, D., Derbie, A., Abeje, A., Shumet, A., Kassahun, Y., Nibret, E., et al. (2019a). Genomic diversity and transmission dynamics of *M. tuberculosis* in Africa: a systematic review and meta-analysis. *IJTLD* 23, 1314–1326. doi: 10.5588/ijtld.19.0127
- Mekonnen, D., Derbie, A., Abeje, A., Shumet, A., Nibret, E., Biadlegne, F., et al. (2019b). Epidemiology of tuberculous lymphadenitis in Africa: a systematic review and meta-analysis. *PLoS ONE* 14, e0215647. doi: 10.1371/journal.pone.0215647
- Mekonnen, D., Derbie, A., Chanie, A., Shumet, A., Biadlegne, F., Kassahun, Y., et al. (2019c). Molecular epidemiology of *M. tuberculosis* in Ethiopia: a systematic review and meta-analysis. *Tuberculosis* 118, 101858. doi: 10.1016/j.tube.2019.101858
- Mekonnen, D., Munshae, A., Nibret, E., Derbie, A., Abeje, A., Feleke, B. E., et al. (2022). Tuberculosis case notification rate mapping in Amhara Regional State, Ethiopia: four years retrospective study: Tuberculosis. *Ethiop. Med. J.* 60, 3–12.
- Mekonnen, D., Nibret, E., Munshae, A., Derbie, A., Zenebe, Y., Tadese, A., et al. (2023). Comparative serum lipid and immunohematological values among adult pulmonary tuberculosis and tuberculous lymphadenitis cases and their association with sputum bacilli load and time to culture positivity in Northwestern Ethiopia. *Lipids Health Dis.* 22, 56. doi: 10.1186/s12944-023-01821-3
- Modi, A., Vai, S., Caramelli, D., and Lari, M. (2021). “The Illumina sequencing protocol and the NovaSeq 6000 system,” in *Bacterial Pangenomics. Methods in Molecular Biology*, Vol 2242, eds A. Mengoni, G. Bacci, and M. Fondi (New York, NY: Springer), 15–42. doi: 10.1007/978-1-0716-1099-2_2
- Mokrousov, I. (2012). The quiet and controversial: ural family of *Mycobacterium tuberculosis*. *Infect. Genet. Evol.* 12, 619–629. doi: 10.1016/j.meegid.2011.09.026
- Molinario, L., Montinaro, F., Yelmen, B., Marnetto, D., Behar, D. M., Kivisild, T., et al. (2019). West Asian sources of the Eurasian component in Ethiopians: a reassessment. *Sci. Rep.* 9, 18811. doi: 10.1038/s41598-019-55344-y
- Moran, N. A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108, 583–586. doi: 10.1016/S0092-8674(02)00665-7
- Negrete-Paz, A. M., Vázquez-Marrufo, G., and Vázquez-Garcidueñas, M. S. (2021). Whole-genome comparative analysis at the lineage/sublineage level discloses relationships between *Mycobacterium tuberculosis* genotype and clinical phenotype. *PeerJ* 9, e12128. doi: 10.7717/peerj.12128
- Ochman, H., and Moran, N. A. (2001). Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* 292, 1096–1099. doi: 10.1126/science.1058543
- Pagani, L., Schiffls, S., Gurdasani, D., Danecek, P., Scally, A., Chen, Y., et al. (2015). Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. *AJHG* 96, 986–991. doi: 10.1016/j.ajhg.2015.04.019
- Palomino, J. C., Leão, S. C., and Ritacco, V. (2007). *Tuberculosis 2007; From Basic Science to Patient Care*. Amedeo Challenge. Available online at: <https://books.google.com.et/books?id=enUOkAEACAAJ>
- Pankhurst, R. (1964). The trade of Northern Ethiopia in the nineteenth and early twentieth centuries. *J. Ethiop. Stud* 2, 49–159.
- Polesky, A., Grove, W., and Bhatia, G. (2005). Peripheral tuberculous lymphadenitis: epidemiology, diagnosis, treatment, and outcome. *Medicine* 84, 350–362. doi: 10.1097/01.md.0000189090.52626.7a
- Qian, X., Nguyen, D. T., Lyu, J., Albers, A. E., Bi, X., and Graviss, E. A. (2018). Risk factors for extrapulmonary dissemination of tuberculosis and associated mortality during treatment for extrapulmonary tuberculosis. *Emerg. Microbes Infect.* 7, 1–14. doi: 10.1038/s41426-018-0106-1
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <https://www.R-project.org/>
- Sakula, A. (1982). Robert Koch: centenary of the discovery of the tubercle bacillus, 1882. *Thorax* 37, 246–251. doi: 10.1136/thx.37.4.246
- Seal, S., Dharmarajan, G., and Khan, I. (2021). Evolution of pathogen tolerance and emerging infections: a missing experimental paradigm. *Elife* 10, e68874. doi: 10.7554/eLife.68874
- Shuaib, Y. A., Utpatel, C., Kohl, T. A., Barilar, I., Diricks, M., Ashraf, N., et al. (2022). Origin and global expansion of *Mycobacterium tuberculosis* complex lineage 3. *Genes* 13, 990. doi: 10.3390/genes13060990
- Stamatidakis, A. (2006). RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446
- Stucki, D., Brites, D., Jeljeli, L., Coscolla, M., Liu, Q., Trauner, A., et al. (2016). *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat. Genet.* 48, 1535–1543. doi: 10.1038/ng.3704
- Tadesse, M., Abebe, G., Bekele, A., Bezabih, M., de Rijk, P., Meehan, C. J., et al. (2017). The predominance of Ethiopian specific *Mycobacterium tuberculosis* families and minimal contribution of *Mycobacterium bovis* in tuberculous lymphadenitis patients in Southwest Ethiopia. *Infect. Genet. Evol.* 55, 251–259. doi: 10.1016/j.meegid.2017.09.016
- Tanaka, M. M., and Francis, A. R. (2005). Methods of quantifying and visualising outbreaks of tuberculosis using genotypic information. *Infect. Genet. Evol.* 5, 35–43. doi: 10.1016/j.meegid.2004.06.001
- Tessema, B., Beer, J., Merker, M., Emmrich, F., Sack, U., Rodloff, A. C., et al. (2013). Molecular epidemiology and transmission dynamics of *Mycobacterium tuberculosis* in Northwest Ethiopia: new phylogenetic lineages found in Northwest Ethiopia. *BMC Infect. Dis.* 13, 131. doi: 10.1186/1471-2334-13-131
- Tulu, B., and Ameni, G. (2018). Spoligotyping based genetic diversity of *Mycobacterium tuberculosis* in Ethiopia: a systematic review. *BMC Infect. Dis.* 18, 1–10. doi: 10.1186/s12879-018-3046-4
- Walker, T. M., Ip, C. L., Harrell, R. H., Evans, J. T., Kapatai, G., Dedicat, M. J., et al. (2013). Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect. Dis.* 13, 137–146. doi: 10.1016/S1473-3099(12)70277-3
- Welekidan, L. N., Yimer, S. A., Skjerve, E., Dejene, T. A., Homberset, H., Tønjum, T., et al. (2021). Whole genome sequencing of drug resistant and drug susceptible *Mycobacterium tuberculosis* isolates from Tigray region, Ethiopia. *Front. Microbiol.* 12, 743198. doi: 10.3389/fmicb.2021.743198
- WHO (2022). *Global Tuberculosis Report 2022*.
- Yew, W. W., and Lee J. (1995). Pathogenesis of cervical tuberculous lymphadenitis: pathways to anatomic localization. *Tuber. Lung Dis.* 76, 275–276. doi: 10.1016/S0962-8479(05)80019-X
- Yimer, S. A., Namouchi, A., Zegeye, E. D., Holm-Hansen, C., Norheim, G., Abebe, M., et al. (2016). Deciphering the recent phylogenetic expansion of the originally deeply rooted *Mycobacterium tuberculosis* lineage 7. *BMC Evol. Biol.* 16, 1–10. doi: 10.1186/s12862-016-0715-z
- Yimer, S. A., Norheim, G., Namouchi, A., Zegeye, E. D., Kinander, W., Tønjum, T., et al. (2015). *Mycobacterium tuberculosis* lineage 7 strains are associated with prolonged patient delay in seeking treatment for pulmonary tuberculosis in Amhara Region, Ethiopia. *J. Clin. Microbiol.* 53, 1301–1309. doi: 10.1128/JCM.03566-14

7. Glossary

AOR, Adjusted odds ratio; APHI, Amhara Public Health Institute; BWA-MEM, Burrows–Wheeler Aligner–Maximal Exact Match; CAS, Central Asia; cgMLST, core genome multilocus sequence typing; COR, crude odds ratio; DRTB, drug-resistant TB; EA.ETH, Euro-American lineage restricted to Ethiopia; EA, Euro-American; EAI, East African Indian; EPM, enhanced PCR mix; GATK, Genome Analysis Toolkit; HRE, high-risk employee; HW, housewife; L, lineages; LAM, Latin America–Mediterranean; LJ, Löwenstein–Jensen medium; Mtb,

Mycobacterium tuberculosis; MDR-TB, multidrug-resistant TB; MGIT, Mycobacterium growth indicator tube; MRCA, most recent common ancestor; MSA, multiple sequence alignment; MTBC, *Mycobacterium tuberculosis complex*; NR, Neolithic revolution; PTB, pulmonary tuberculosis; RaxML, randomized accelerated maximum likelihood; SDS, sodium dodecylbenzenesulfonate; SEA, South East Asia; SNP, single-nucleotide polymorphism; T, Tuscany; TB, tuberculosis; TBL, terminal branch lengths; TBLN, *Tuberculosis lymphadenitis*; WGS, whole-genome sequencing; WHO, World Health Organization; XDR-TB, extensively drug-resistant TB.