Molecular epidemiology of tuberculosis in northeastern Ethiopia

Fikru Gashaw (✉ fikrug2012@gmail.com)

Kotebe Metropolitan University College of Natural and Computational Sciences <https://orcid.org/0000-0002-8962-9686>

Aboma Zewde

Ethiopian Public Health Institute

Endalkachew Tedla

Bikat Higher Diagnostic Laboratory

Biniam Wondale

Arbaminch University

Yalemtsehay Mekonnen

Addis Ababa University College of Natural Sciences

Berhanu Erko

Addis Ababa University Aklilu Lemma Institute of Pathobiology

Gobena Ameni

United Arab Emirates University

Research Article

Keywords: Tuberculosis, Molecular Epidemiology, Northeastern Ethiopia

Posted Date: July 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-682867/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Tuberculosis (TB) is a major public health threat causing the highest morbidity and mortality in Ethiopia. However, there is shortage of information on the species and strains of mycobacteria that cause TB in Ethiopia. The objective of this study is to investigate the diversity of *Mycobacterium tuberculosis* (*M. tb*) isolated from TB patients in northeastern Ethiopia.

Methods: A total of 384 smear positive pulmonary and extra-pulmonary TB cases were recruited on the basis of clinical examination and Ziehl-Neelsen staining for further investigation. Mycobacterial culturing, region of difference (RD) 9-based polymerase chain reaction (PCR), spoligotyping and mycobacterial interspersed repetitive unit-variable-number tandem repeat (MIRU-VNTR) typing were used for the identification the species and strains. Descriptive statistical analysis was used for the expression of the results while proportions were compared using χ^2 test. Statistical significance was considered when $p < 0.05$.

Results: Smear positive pulmonary tuberculosis accounted for 74.5% (286/384). The disease proportion was highest in age group of 18–37 years. Culture positivity was confirmed only in 29.2% (112/384). Speciation of the culture positive isolates using RD9 indicated that 77.7% (87/112) were *M. tuberculosis*. Further identification of the isolates using spoligotyping indicated 92.9% (104/112) interpretable spoligotyping patterns of which 20.2% (21/104) were grouped under 10 clustered patterns. Further grouping of the spoligotypes indicated that 86.5% (90/104) of the isolates as orphan types. Grouping of the isolates into the major lineages showed 52.9%, 27.9% and 19.2% of the isolates as Euroamerican, Indo-oceanic and East African Indian lineages, respectively. Typing of 69 isolates using 24-loci MIRU-VNTR typing gave interpretable results for 56 isolates each of them having distinct MIRU-VNTR profile and were considered as 56 different genotypes (strains).

Conclusion: *M. tuberculosis* was identified as the main cause of TB in northeastern Ethiopia and its strains were highly diversified.

Background

M. tuberculosis infects about a quarter of the global human populations and most of the cases were in the WHO regions of South-East Asia, Africa and the Western Pacific [1]. Human TB is thought to be originated in Africa and spread into Europe and Asia. It is later expanded to other geographical areas with two or three phylogenetically 'ancient' lineages staying in Africa [2].

Recent developments in DNA technology and molecular biology have led to methods for the rapid detection of mycobacterial DNA by nucleic acid amplification and characterization of the microbes. For such molecular determination of the *Mycobacterium* region of differentiation, spoligotyping and MIRU-VNTR are some commonly used tools to characterize, trace source of infection and arrest further transmission of the causative agent on time. The RDs harbor several important genes and virulence factors, and their presence or absence could help to identify species of different isolates in a particular

geographical region on an evolutionary time scale [3]. Spoligotyping is further used for simultaneous detection and molecular typing of the *Mycobacterium*. It is based on the visualization of the spacer DNA sequences in between the 36-bp direct repeats (DRs). The DR region in individual *M. tb* strains and in different members of the MTBC was identified and alignment of the spoligotype patterns was done to group the isolates according to their similarity into clades or strain families. The spoligotyping is further used for genotyping of the MTBC isolates and identify the circulating lineages with cluster formation and orphans at the targeted study area [4].

The MIRU-VNTR typing is the most advanced typing technique with more acceptance in comparison to spoligotyping sensitivity. It shows an adequate balance between variability and also has essential features to differentiate non-related isolates. The MIRU also has greater discriminatory performance as compared to spoligotyping with better performance when both are combined. It is further crucial for determination of recent transmission among the community [5]. The 24-locus sets have improved its discriminatory importance than the initial 12-locus or the 15-locus sets and is suggested as the current gold standard technique in molecular typing of the *Mycobacterium* [6-8].

In Ethiopia, most of the characterization technologies are regions of differentiation and spoligotyping while other forms of characterization are done outside the country. The MIRU-VNTR technology was established for the first time in the country and applied in this study using the *Mycobacterium* isolates from northeastern part of the country. Of the molecular characterization done in Ethiopia, the most studies were conducted in Northwest [9-12] where as it was rare from other parts of the country. Some of the studies found noble new Ethiopian phylogenetic lineages which were not reported elsewhere in the world. Scarcity of precise and instant *M. tb* detection technologies in the country is an additional concern. Nationwide status regarding molecular epidemiology of the disease is not yet well established. This needs further studies at different sites of the country to sum up and make countrywide determination at regional and zonal level for the bacterial strains. As far as our knowledge is concerned, no study has been conducted on the molecular characterization of the *Mycobacterium* in Oromia Special Zone and only limited investigation was done in the South Wollo Zone of this study settings.

Methods

The study area

The study was conducted in Oromia Special Zone and South Wollo Zone of the Amhara Regional State, northeastern Ethiopia. A preliminary survey was made for 3 months in governmental health facilities of the seven districts in Oromia Special Zone; based on the availability of samples and transportation access, the data were continuously collected from Kemise and Bati Town health centers. However, due to a lower number of TB cases accessed from the special Zone, the study was extended to four neighboring health institutes of Dessie Town (Dessie Referral Hospital, Bikat Higher Diagnostic Laboratories, Dessie Health Center and Boru Meda Hospital) as additional study sites.

Study design and laboratory processing

An institution-based cross-sectional study was conducted from April 2015 to January 2017. Dry, translucent, leak-proof 50ml capacities of falcon tubes were used to collect a minimum of 3-5ml sputum samples. Disposable gloves and respiratory masks were used when samples were collected from suspected TB patients. For all study participants, the socio-demographic data was also recorded on the spot of sample collection.

Bacteriological examinations

Ziehl-Neelsen staining. Ziehl-Neelsen staining and direct microscopic examination for acid-fast bacteria was performed at sample collection sites. The positive samples were temporarily stored at -20 °C in the refrigerator of the health institutes and transported to Aklilu Lemma Institute of Pathobiology (ALIPB) using a cold chain at 4 °C for further laboratory processing [13-15].

Mycobacterial culturing. A stock of selective Lowenstein-Jensen (LJ) media, glycerol, pyruvate and a homogenized whole eggs was processed for the *Mycobacterium* culture. Both of the sputum and FNA samples were cultured following [16] procedures. The inoculated LJ slants were incubated aerobically at 37°C and monitored every week for the formation of mycobacterial colony until 8th week. The grown *Mycobacteria* colony was collected, heat killed and freezed whereas the weakly grown colonies were sub-cultured. Both, the heat killed and freezed isolates were kept at -80°C until molecular characterization was done using region of differentiation (RD), spoligotyping and MIRU-VNTR.

Molecular typing

Speciation of the isolates using RD9-based polymerase chain reaction

Heat killed cells and the genomes oligonucleotide primers (RD9 FW and RD9 REV) each at a concentration of 100 mM was used for the PCR amplification process. The reaction mixture was prepared and subjected to 35 cycles consisting of 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min in the PCR thermocycler (VWR, UK). Finally, the reaction mixture was maintained at 72°C for 10 min. The resulting PCR products were processed by 1% standard agarose gel electrophoresis at 110 V and 400 mA for 35 minutes. The gel was visualized using a UVP photodoc imaging system and the resulting bands were interpreted compared to positive (*M. tb* H37Rv and *M. bovis*) and negative (distilled water) controls [11,17].

Spoligotyping and its result interpretation. The DR region was amplified by PCR using oligonucleotide primers (DRa biotinylated at 5' end and DRbs) derived from the DR sequence [4]. The PCR was processed and hybridization of DNA was detected by the enhanced chemiluminescence detection liquid followed by exposure to X-ray film as described by the manufacturer (Hain Life Science Company). The autorad

spoligotyping results were checked visually by three experienced operators and the spacers were written in a binary format using uncapitalized English letters 'o' and 'n' indicating when the spacers are absent and present, respectively. The binary representation was converted to the octal code and entered to the international database SITVIT2 (Institute Pasteur de Guadeloupe) to determine and interpret the specific *M. tb* complex strain and SIT. The novel isolates which have not yet been described in the existing spoligotyping data base profile using SITVIT2 lists were considered as orphans [4].

MIRU-VNTR typing. In this study, performance and optimization of the protocol were done at ALIPB-AAU. The PCR amplification Master Mix was prepared for 26 PCR tubes of which 24 pairs of the primers [Additional file 1] were used for a single isolate (Table 1).

Twenty four µl of the Master Mix was distributed to 24 PCR tubes. Then 1µl of the respective 24-locus MIRU-VNTR primers were added to the tubes to make a final volume of 25µl for the amplification. Thermocycler was set with an enzyme activation step of 15 minutes at 95°C, followed by 40 cycles of 1 minute at 94°C for denaturation, 1 minute at 59°C for annealing, and 1 minute 30 seconds at 72°C for extension. Thereafter, the reactions were incubated for 10 minutes at 72°C for final extension/elongation. A positive control, H37Rv, and a negative control, sterile distilled water, were used in the study and the amplified PCR products were run in a gel-electrophoresis [6,18,19]. The 69 amplified PCR products of MGIT sub-culture positive isolates were electrophoresed to determine the size of amplicon. Electrophoresis was made on 1.8% (w/v) 300ml agarose gel with 15 µl ethidium bromide in 1X Tris Borate-EDTA (TBE) buffer run at 120 volts and 400 milliampere for 5 hours. Product sizes of each band were determined by comparing with the standard DNA ladder bands (100 bp and 50 bp) after photograph was taken under ultraviolet (UV) transilluminator. The numbers of MIRU-VNTR alleles were determined by inferring the size of bands with interpretation tables [19]. Then, the main phylogenetic predictions were facilitated using MIRU-VNTR-24 loci profiles into freely accessible, on-line strain identification databases of MIRU-VNTR*Plus* (<http://www.miru-vntrplus.org>). The isolate patterns were used to compare with the reference strains in the database for the assignment of MTBC species, lineages and genotypes. Phylogenetic dendrogram was constructed based on the neighbor-joining (NJ) clustering algorithms and minimum spanning tree (MST) analysis was also performed.

TB cases confirmed by the health personnel and those who fulfilled the inclusion criteria were included in the study. The samples (sputum and Fine needle aspirates (FNAs)) were collected on the spot by consenting participants 18 years and older. Those with severe TB who were unable to provide their sputum specimens were also excluded from the study.

Data analysis

Descriptive statistics was used to determine frequency and percentage. In spoligotyping, the reference data base available online <http://www.pasteur-guadeloupe.fr:8081/SITVITDemo/> was used to assign the shared international spoligotype numbers (SIT); but if SIT number was not found, the pattern was considered as 'orphan' type. An online tool Run TB-Lineage with a website of

http://tbinsight.cs.rpi.edu/run_tb_lineage.html was also used to identify family/clade, lineages and sublineages of the isolates. Spoligotypes consisting of more than one isolate were classified as clustered types while those with only one isolate were classified as singletons. For recent transmission index (RTI), cluster analysis was calculated using the formula $(nc - c)/n$, where nc is the total number of clustered patients, c is the number of clusters, and n as the total number of patients in the sample. The discriminatory power of each locus was also evaluated using the Hunter and Gaston Discriminatory Index [20].

Results

Tuberculosis infections and demographic characteristics of the study patients

A total of 384 TB cases (213 males and 171 females) were involved in the study. Both forms of TB were identified with smear positive pulmonary cases as more predominant (74.5%) than the EPTB. TB lymphadenitis was found to be the most prevalent (85.9%) form of EPTB with cervical adenopathy (75.3%) being the commonly existing cases. A diverse spoligotyping pattern was identified with 86.5% as not registered in the global spoligotyping database. A low proportion of the isolates (20.2%) was recognized in clustered forms by the spoligotyping. From the study, both modern and ancient Lineages were identified with the modern Euro-American Lineage as predominant. The 24-loci MIRU-VNTR showed all the isolates as orphan and highly diverse.

Most of the cases 64.3% (247/384) were recruited from the South Wollo Zone and the remaining subjects were from Oromia Special Zone and elsewhere. Of the identified TB cases, 96% (369/384) were recorded at the district level (Table 2).

Pulmonary TB cases accounted for 74.5% (286/384), and the overall prevalence of TB was highest (67.0%) in the 18–37 years age group.

Molecular epidemiology of tuberculosis

Speciation of the isolates using genomic deletion and spoligotyping

Deletion typing via region of differentiation was made in all 112 LJ-culture positive isolates with 77.7% had an intact RD9 (396 bp) and identified as *M. tuberculosis*. On the other hand, a total of 92.9% (104/112) of the isolates gave interpretable results for the formation of spoligotyping patterns. Ten clusters were identified in 20.2% (21/104) of the isolates while the remaining isolates were unique (singletons). Nine of the clusters had two isolates each while the remaining one cluster had three isolates with similar patterns. In these findings, the majority of the patients had different strains of *M.*

tuberculosis. Taking the number of cases with clustered genotypes into account, the RTI was calculated with a result of 0.12.

The study showed that there was a statistically significant difference in the proportion of clustering across the Oromia Special Zone and South Wollo Zone isolates (p -value = 0.000). In addition, all the isolates from the Oromia Special Zone were orphans (Table 3).

Families and lineages of the isolates

Identification of the strains using web-based SPOTCLUST databases indicated more than 73% (76/104) belong to the family T1, family33, H37Rv and CAS families accounted for 24.0% (25/104), 23.1% (24/104), 14.4% (15/104) and 11.5% (12/104), respectively. The major lineages identified by the study were: Lineage 1 or Indo-Oceanic (IO), Lineage 3 or East-African India (EAI) and Lineage 4 or Euro-American (EUA). Of these, the most predominant lineages was Lineage 4 followed by ancestral Lineage 1. Further characterization to the sub-lineage level showed a leading T sublineage from EUA genotype (Table 4).

There were 14 already recorded isolates as a distinct spoligotype patterns shared international types (SIT) namely SIT 149 (3 isolates), SIT 53 and 1378 (2 isolates each), SIT 1802, 47, 612, 1166, 1251, 1475 and 1547 (one isolate each). Majority of the isolates 86.5% (90/104) were found to be orphans [Additional file 2] with some spoligotyping patterns as shown in Table 5.

Mycobacterial interspersed repetitive unit variable number tandem repeat

Of the 69/112 MGIT sub-culture positive isolates, 56 had valid amplification products for the 24-loci MIRU-VNTR while the remaining 13 isolates had either incomplete or negative MIRU-VNTR profiles. The locus band was absent for one of the 24 loci (MIRU-VNTR locus 4052) with the oligos sequence FW(5'-AACGCTCAGCTGTCGGAT-3') and REV(5'-CGGCCGTGCCGCCAGGTCCTTCCCGAT-3') in the gel for all isolates. The MIRU-24 locus of different *M. tb.* strains was amplified by PCR and separated by agarose gel electrophoresis. Each strain had a different allele number ranging from 0-9 repeats. Highly diverse genotypes were displayed with all the valid patterns as unique and no clustered isolates were detected. The discriminatory efficiency of 24-loci MIRU-VNTR and a combination of both the MIRU-VNTR and spoligotyping in this study was found as the highest with HGDI as 1.000 in each case. It was higher than that of spoligotyping discriminatory power (HGDI = 0.996). The minimum spanning tree (MST) analysis determined the evolutionary relationship among the strains using MIRU-VNTR_{plus} data and none of the strains form a distinct complex [Additional file 3].

Based on the 56 isolates under study for both spoligotyping and MIRU-VNTR analysis, one of the isolates was not valid for the spoligotyping patterns. The 24-loci MIRU-VNTR genotyping showed all the isolates as unique and no recent transmission among the TB cases. The SITVIT analysis for the remaining isolates identified a diverse sub-lineages as T (20%), CAS1-Delhi (16.4%), Manu3 (12.8%), CAS1-kili (9.1%), Manu2 (7.3%), CAS (5.5%), H1 (5.5%), T3-ETH (5.5%), H3 (3.6%), H (3.6%) and the remaining small

proportion Manu3, EAI1-SOM, H3-Ural-1, Manu-ancestor, T-Tuscany and Turkey were 1.8% each. Based on 24-loci MIRU-VNTR *plus* analysis, the identified lineages were Delhi/CAS (32.1%), Ethiopia_3 (26.8%), Unknown (19.6%), Haarlem and NEW-1 4 (7.1% each), Ural (5.4%) and EAI (1.8%) (Fig. 1). The result of MIRU-VNTR showed that isolates with similar spoligotype pattern (clustered isolates) had unique copies (singletons) under the 24-loci MIRU-VNTR analysis in the study. All isolates in the study were new types to MLVA MtbC15 and 30 isolates were new types to MLVA MtbC9. These new types were submitted and confirmed to the MIRU-VNTR *plus* nomenclature database.

Discussion

This is the first molecular epidemiologic study of MTBC strains from Oromia Special Zone in Amhara region with an additional characterizations of the *Mycobacterium* isolates from South Wollo using region of differentiation, spoligotyping and 24-loci MIRU-VNTR typing. Importantly, the MIRU-VNTR technique was established for the first time in the country at ALIPB-AAU using gel electrophoresis and identified all the characterized isolates as orphan. The study also found pulmonary TB cases as the most dominant TB types.

Culture positivity, pulmonary and extra-pulmonary tuberculosis

The lower proportion of culture positivity in this study is close to the study in Addis Ababa [21]; this might be due to delayed culturing time and electric interruption when the specimens were preserved in the refrigerator at the sample collection sites and a long distance travel from temporarily storage sample collection site to ALIPB where the specimens were cultured. These factors increase the chance of bacterial death in the collected sputum samples. Perhaps it could also be our expectation that bacterial culture positivity might decrease when the samples were stored in the refrigerator for a long duration than instant culturing.

The extent of pulmonary and extra-pulmonary TB culture positivity in this study is lower than that of a study in India [22]. A study from different sites in Ethiopia also reported greater culture positivity of both clinically manifested smear positive pulmonary 79% (753/953) and extra-pulmonary 38% (456/1198) TB [15]. Likewise, a study report from northwestern Ethiopia and Addis Ababa showed greater culture positivity of PTB than EPTB [23, 24]. In all instances, there is lower culture positivity of EPTB than pulmonary ones; this might be due to cytological suspicion of the specimen by pathologist unlike to the detection of disease causative organism itself as in the case of PTB. Moreover, the suspected *Mycobacterium* cellular infection could be paucibacillary which decreases the sensitivity of diagnostic test in EPTB.

The findings of this study agree with the global report that the proportion of EPTB is higher; this might be due to diagnostic challenges including shortage of pathologists to identify and treat the cases on time in most health institutions. There were no pathologists to diagnose EPTB in all governmental health institutes including the referral hospital where this study was conducted. Because of this, all the

suspected cases were referred to a single private diagnostic laboratory (Bikat Higher Diagnostic Laboratory) and remained as the diagnostic challenge of the area.

The EPTB has different manifestations based on the organs being attacked and its intent of dissemination in the body. Similar to other relevant studies in the country, this finding also revealed lymph nodes as the leading organs affected. In fact, the percentage of their infection rate differs as cervical, auxiliary, inguinal, supra-clavicular, sub-mandibular and anterior neck lymph nodes [15, 23]. The higher infection of lymph node is similar to the study reported from Germany [25] but the most common sites involved were bones/joints and lymph nodes in United States of America [26], whereas the genitourinary system and skin were the common sites of infection reported from Hong Kong [27]. Such differences might be attributable to either host or pathogen related factors as well as access to patient sample collection in the clinical settings.

The largest number of TB cases in Dessie Town compared with other districts might be due to higher population density in the Town. This is in agreement with the WHO report that revealed the prevalence of TB cases as considerably higher in urban areas than in rural areas [28]. In addition, access and better diagnosis with more proximity to the health institutes could be another factor to find out the higher number of cases in Dessie Town. Male to female ratio of this finding is the same to bacteriologically confirmed pulmonary TB patients of WHO report for Ethiopia with M:F ratio as 1.2 [1]. Such greater proportion of male patients to females might be due to biological, social and economic activities to contact with many people. The disease was also more common within age ranges of 18 - 37 years which could be due to their active movement for economic engagement leading to greater risk of exposure.

Molecular epidemiology of tuberculosis

The detection of *M. tb* from culture positive samples in this study using RD9 was lower than the findings from many other studies. A study from Western [29], Northeastern [14], central [30] and Northwestern [12] Ethiopia reported *M. tb* detection proportion of 97.1% in the first study and 100% in all the rest. On the contrary, there was also a report from Addis Ababa [31] that showed RD9 detection typing for *M. tb* with a lower proportion of 47.7% (41/86) than this finding.

The variation of spoligotype clustering in this study in comparison to many other reports might be due to the differences in geographical study settings. There were higher clustering rates reported from most studies in Ethiopia [30,32,33]. Similarly, studies outside of the country also reported greater proportion of clusters [34-37] than the overall clustering rate of this study 20.2% (21/104) which might be due to the poor recovery rate of smear positive bacteria on LJ-culture. In fact, this finding is close to 18.8% (6/32) [14] and 23% (6/26) [38] report from Dessie and Addis Ababa, respectively. The findings of all isolates in Oromia Special Zone as Orphan imply that they were not registered in the database as there was no more study from the area so far. The lower proportion of clustering rates in this study could imply that the *Mycobacterium* infections were from unrelated sources or they might be caused due to latent

tuberculosis activation. The predominance families T1, family33, H37Rv and CAS is in line with a systematic review reported from the country [39] as well as a research finding [40].

The higher proportion of modern strains Lineage 3 (East-African Indian) and Lineage 4 (Euro-American) of the *Mycobacterium* than the ancient Lineage 1 (Indio Oceanic) was in agreement with a study from Dessie 71.4% (20/28) [14]. This greater proportion of the modern strains to the ancient strains could be due to the recent expansion of tuberculosis in Ethiopia compared with the ancient Indio Oceanic ones which are more common in populations living around the Indian Ocean.

The higher proportion of T sublineage followed by CAS1-Delhi in the present study was compatible with a study reported from Addis Ababa [41]. Higher proportion of T3-ETH sublineage implied its greatest transmission rate in the area. Unlike other studies [12,30], it was also found that lower number of isolates (2 or 3) with the same spoligotyping patterns were found in a single cluster of this finding. In fact, the less number of isolates in a cluster agrees with a study report from Gambella region, Southwest Ethiopia [42]. The higher proportion of orphans could be due to limitation of reports to the spoligotype database. On the other hand, the leading shared strain type frequently reported from Ethiopia namely SIT149 [39,43-45] was also detected at greater rate in this finding.

The highest discriminatory power of the 24-loci MIRU-VNTR to spoligotyping agrees with other similar studies [30,46] and it can be used alone or in combination with spoligotyping for the discrimination purpose. Missing some of the 24-loci in this study might be due to amplification errors in the PCRs as similar to the validation of 24-Locus Variable-Number Tandem-Repeat Typing for *M. tuberculosis* [8].

The lowest clustering rate by spoligotyping and absence of clustering at all in MIRU-VNTR of this study is an indication of minor or no recent transmission of the *Mycobacterium* in the area, which implies that the disease is mainly due to endogenous reactivation of the latent TB infection. Similarly, other studies also showed that the clustering rate of MIRU-VNTR is less than that of spoligotyping [32,35,47,48]. Those clustering rate differences in different study areas might be due to the differences in the geography, population density and socio-economic diversity [49].

In agreement with previous studies, the predominance of lineage 3 (Delhi/CAS; 32.1%) using MIRU-VNTR 24-loci genotyping showed its wide distribution throughout the country [10,50,51]. On the contrary, it was not the predominant strain rather H37Rv like and Ethiopia_3 were the most common sub-lineages in studies from prisoners and communities in Southern, Southwestern and Southeastern Ethiopia [52] and Eastern Ethiopia [32], respectively. This revealed that the overall predominant lineage of *M. tb* across the country varies.

Conclusions

The findings in this study show that all of the causative agents were *M. tuberculosis* with pulmonary TBs at a greater proportion. The majority of the isolates were singletons rather than clustered forms suggesting that lower recent transmission but evidence for more of endogenous reactivation. Family T1

and family33 were the most frequently infecting families with Euro-American lineage as the pre-dominant. The study also identified that the majority of the isolates were orphan. All the isolates typed by MIRU-VNTR 24-loci were singletons implying the greatest genetic diversity. Thus, the presence of a large number of orphan isolates needs further investigation using more numbers of isolates to report and register in the international database. Identifying the status of genetic diversities could also help to strengthen TB prevention and control programs in the study area.

Declarations

List of abbreviations

AAU: Addis Ababa university; ALIPB: Aklilu Lemma institute of pathobiology; CAS: Central Asian; CBN: Conformal Bayesian network; DNA: Deoxyribonucleic acid; DRs: Direct repeats; EAI: East-African India; EPTB: Extra-pulmonary tuberculosis; EUA: Euro-American; FNA: Fine needle aspirate; H: Haarlem; HGDI: Hunter Gaston discriminatory index; IO: Indo-Oceanic; KBBN: Knowledge-based Bayesian network; LAM: Latin American; LJ: Lowenstein-Jensen; *M. tb*: *Mycobacterium tuberculosis*; MGIT: Mycobacteria growth indicator tube; MIRU-VNTR: Mycobacterial interspersed repetitive unit-variable-number tandem repeat; MST: Minimum spanning tree; MTBC: *Mycobacterium tuberculosis* complex; PCR: Polymerase chain reaction; PTB: Pulmonary tuberculosis; RD: Region of difference; RTI: Recent transmission index; SIT: Shared international types; T: Tuscany; TB: Tuberculosis; UV: Ultraviolet; WHO: World health organization.

Ethics approval and consent to participate

Ethical clearance was obtained from the Institute Review Board of College of Natural and Computational Sciences, Addis Ababa University (Ref. No. CNSDO/392/07/15) later renewed in July 2019 (CNSDO/668/11/2019). A letter of permission was also obtained from the health bureaus of the study sites and a written informed consent from study participants.

Consent for publication

Not applicable

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by funds from AAU thematic research project A/C 0162230106072100101. In addition, it was also supported in part by the NIH/ Fogarty International Center Global Infectious Diseases grant D43TW009127 through H3 Africa.

Authors' contributions

FG, GA, BE, and YM conceived the idea and designed the study. FG, GA, BE, YM and ET coordinated sample collection. FG, GA, BW, AZ and ET carried out the laboratory analysis. FG, GA and BW did data analysis and interpretation. FG drafted the manuscript, and BE, BW, GA and YM reviewed and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank the zonal and districts health bureau of Oromia Special Zone and South Wollo Zone in Amhara region who gave us permission to conduct the research under their health institutes. All health institutes of the sample collection sites and health personels are duly acknowledged for their primary identification and diagnosis of TB cases. We also greatly thank all study participants for being part of this study. We would like to acknowledge Lisa Sthreshley for buying laboratory materials and processing its transportation from the USA to Ethiopia (ALIPB) specifically to establish the new diagnostic technique (MIRU-VNTR) in Ethiopia. All TB laboratory workers of ALIPB and the staff members are duly acknowledged for their technical support, without which this study would have not been completed. We appreciate Hannah Nicol an English language specialist at Emory University for her constructive comments on the writings of this paper.

References

1. WHO. Global tuberculosis report. Geneva: World health organization; 2019.
2. Comas I, Coscolla M, Luo T, Borrell S, Holt K, Kato-Maeda M. Out-of-Africa migration and Neolithic co expansion of *Mycobacterium tuberculosis* with modern humans. Nat Genet. 2013;45:1176–82.
3. Pinsky B, Banaei N. Multiplex Real-Time PCR Assay for Rapid Identification of *Mycobacterium tuberculosis* Complex Members to the Species Level. J Clin Microbiol. 2008;46:2241–46.
4. Kamerbeek J, Schouls L, Kolk A, Agterveld M, Soolingen D, Suijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol. 1997;35:907-14.
5. Rozo-Anaya J, Ribón Y. Molecular tools for *Mycobacterium tuberculosis* genotyping. Rev salud pública. 2010;12: 510-21.
6. Supply P, Allix C, Lesjean S. "Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*," J Clin Microbiol.

2006;44: 4498-510.

7. Iwamoto T, Yoshida S, Suzuki K, Tomita M, Fujiyama R, Tanaka N, 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.

8. Beer J, Akkerman O, Schürch A, Mulder A, Werf T, Zanden A, et al. Optimization of Standard In-House 24-Locus Variable-Number Tandem-Repeat Typing for *Mycobacterium tuberculosis* and Its Direct Application to Clinical Material. J Clin Microbiol. 2014;52: 1338-42.

9. Tessema B, Beer J, Emmrich F, Sack U, Rodloff AC. Analysis of gene mutations associated with isoniazid, rifampicin and ethambutol resistance among *Mycobacterium tuberculosis* isolates from Ethiopia. BMC Infect Dis. 2012;12:37.

10. Tessema B, Beer J, Merker M, Emmrich F, Sack U, Rodloff AC. Molecular epidemiology and transmission dynamics of *Mycobacterium tuberculosis* in Northwest Ethiopia: new phylogenetic lineages found in Northwest Ethiopia. *BMC Infect Dis.* **2013;13**:131.

11. Debebe T, Admassu A, Mamo G, Ameni G. Molecular characterization of *Mycobacterium tuberculosis* isolated from pulmonary tuberculosis patients in Felege Hiwot Referral Hospital, northwest Ethiopia. J Microb Immunol Infect. 2014; 47:333-38.

12. Alelign A, Petros B, Ameni G. Smear positive tuberculosis and genetic diversity of *M. tuberculosis* isolates in individuals visiting health facilities in South Gondar Zone, northwest Ethiopia. PLoS One. 2019;14:e0216437.

13. Yang Z, Kong Y, Wilson F, Foxman B, Fowler A, Marrs C. Identification of Risk Factors for Extra-pulmonary Tuberculosis. Clin Infect Dis. 2004;38:199-205.

14. Birhanu T, Belay M, Ameni G. Molecular Characterization of *Mycobacterium tuberculosis* Complex Isolated from Tuberculous Lymphadenitis Patients at Dessie Private Hospitals, Northern Ethiopia. JBAH. 2014; 4:125-32.

15. Berg S, Schelling E, Hailu E, Firdessa R, Gumi B, Erenso G. 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. 2015;15:112.

16. Tripathi K, Tripathi PC, Nema S, Shrivastava AK, Dwiwedi K, Dhanvijay AK. Modified Petroff's Method: an Excellent Simplified Decontamination Technique in Comparison with Petroff's Method. Int J Recent Trends Sci Technol. 2014;10:2249-8109.

17. Rajender KR, Kauser F, Srinivas S, Zanetti S, Sechi LA, Ahmed N. Analysis of Genomic Downsizing on the Basis of Region-of-Difference Polymorphism Profiling of *Mycobacterium tuberculosis* Patient Isolates

Reveals Geographic Partitioning. J Clin Microbiol. 2005;43:5978-82.

18. Cowan L, Mosher L, Diem L, Massey J, Crawford J. Variable-Number Tandem Repeat Typing of *Mycobacterium tuberculosis* Isolates with Low Copy Numbers of IS6110 by Using Mycobacterial Interspersed Repetitive Units. J Clin Microbiol. 2002; 40:1592–602.
19. Supply P. Multilocus Variable Number Tandem Repeat Genotyping of *Mycobacterium tuberculosis* Technical guide. Institute de Biologie/Institute Pasteur de Lille. 2005; 5-73.
20. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol. 1988;26:2465-66.
21. Adane K, Ameni G, Bekele S, Abebe M, Aseffa A. Prevalence and drug resistance profile of *Mycobacterium tuberculosis* isolated from pulmonary tuberculosis patients attending two public hospitals in East Gojjam Zone, northwest Ethiopia. BMC Public Health. 2015;15:572.
22. Sinha P, Gupta A, Prakash P, Anupurba S, Tripathi R, Srivastava G. Differentiation of *Mycobacterium tuberculosis* complex from non-tubercular mycobacteria by nested multiplex PCR targeting IS6110, MTP40 and 32kD alpha antigen encoding gene fragments. BMC Infect Dis. 2016; 16:123.
23. Korma W, Mihret A, Hussien J, Anthony R, Lakew M, Aseffa A. Clinical, molecular and drug sensitivity pattern of mycobacterial isolates from extra-pulmonary tuberculosis cases in Addis Ababa, Ethiopia. BMC Infect Dis. 2015;15:456.
24. Fanosie A, Gelaw B, Tessema B, Tesfay W, Admasu A, Yitayew G. *Mycobacterium tuberculosis* Complex and HIV Co-Infection among extrapulmonary tuberculosis suspected cases at the university of Gondar hospital, Northwestern Ethiopia. PLoS One. 2016; doi:10.1371/journal.pone.0150646.
25. Forssbohm M, Zwahlen M, Loddenkemper R, Rieder H. Demographic characteristics of patients with extrapulmonary tuberculosis in Germany. Eur Respir J. 2008;31:99–105.
26. Peto H, Pratt R, Harrington T, LoBue P, Armstrong L. Epidemiology of extrapulmonary tuberculosis in the United States, 1993–2006. Clin Infect Dis. 2009;49:1350–7.
27. Noertjojo K, Tam C, Chan S, Chan-Yeung M. Extra-pulmonary and pulmonary tuberculosis in Hong Kong. Int J Tuberc Lung Dis. 2002;6:879–86.
28. WHO. Global tuberculosis report. Geneva: World health organization; 2015.
29. Disassa H, Tafess K, Worku A, Ameni G. A Preliminary Study on Molecular Characterization of *Mycobacterium tuberculosis* in Benishangul Gumuz Region, Western Ethiopia. BMRJ. 2015;10:1-10.
30. Bedewi Z, Worku A, Mekonnen Y, Yimer G, Medhin G, Mamo G. Molecular typing of *Mycobacterium tuberculosis* complex isolated from pulmonary tuberculosis patients in central Ethiopia. BMC Infect Dis.

31. Negesse A, Belay M, Medhin G, Ayalew S, Mihret A, Legesse M. Prevalence of pulmonary tuberculosis among patients presenting with cough of any duration in Addis Ababa, Ethiopia. 2019; doi: <http://dx.doi.org/10.1101/622464>.
32. Mekonnen A, Merker M, Collins JM, Addise D, Aseffa A, Petros B, et al. Molecular epidemiology and drug resistance patterns of *Mycobacterium tuberculosis* complex isolates from university students and the local community in Eastern Ethiopia. PLoS One. 2018;13: e0198054.
33. Molina-Moya B, Agonafir M, Blanco S, Dacombe R, Gomgnimbou MK, Spinasse L, et al. Microbead-based spoligotyping of *Mycobacterium tuberculosis* from Ziehl-Neelsen-stained microscopy preparations in Ethiopia. Scientific reports. 2018; 8:3987.
34. Kisa O, Tarhan G, Gunal S, Albay A, Durmaz R, Saribas Z. Distribution of Spoligotyping Defined Genotypic Lineages among Drug-Resistant *Mycobacterium tuberculosis* Complex Clinical Isolates in Ankara, Turkey. PLoS One. 2012;7:e30331.
35. Ahmed M, Mohammed S, Nasurallah H, Ali M, Couvin D, Rastogi N. Snapshot of the genetic diversity of *Mycobacterium tuberculosis* isolates in Iraq. Int J Mycobacteriol. 2014;3:184-96.
36. Ali R, Trovato A, Couvin D, Al-Thwani A, Borroni E, Dhaer F, et al. Molecular Epidemiology and Genotyping of *Mycobacterium tuberculosis* Isolated in Baghdad. BioMed Res Int. 2014;2014:580981.
37. Carvalho R, Vasconcellos S, Issa M, Soares F, Mota P, Araújo F, et al. Molecular Typing of *Mycobacterium bovis* from Cattle Reared in Midwest Brazil. PLoS One. 2016;11:e0162459.
38. Chemed A, Mihret A, Abebe T, Worku T, Ameni G. Genotyping of *Mycobacterium tuberculosis* isolated from pulmonary tuberculosis patients among people living with HIV in Addis Ababa: Cross sectional study. J Clin Tuberc Other Mycobact Dis. 2018;12:34–37.
39. Tulu B, Ameni G. Spoligotyping based genetic diversity of *Mycobacterium tuberculosis* in Ethiopia: a systematic review. BMC Infect Dis. 2018;18:140.
40. Mihret A, Bekele Y, Aytenew M, Assefa Y, Abebe M, Wassie L, et al. Diversity of *Mycobacterium tuberculosis* Isolates from New Pulmonary Tuberculosis Cases in Addis Ababa, Ethiopia. Tuberc Res Treat. 2012;doi:10.1155/2012/892079.
41. Zewdie O, Mihret A, Ameni G, Worku A, Gemechu T, Abebe T. Molecular typing of mycobacteria isolated from tuberculous lymphadenitis cases in Addis Ababa, Ethiopia. Int J Tuberc Lung Dis. 2016;20:1529–34.
42. Asebe G, Dissasa H, Teklu T, Gebreegizeabhe G, Tafese K, Ameni G. Treatment outcome of Tuberculosis Patients at Gambella Hospital, Southwest Ethiopia: Three-year Retrospective Study. J Infect

Dis Ther. 2015;3:2 <http://dx.doi.org/10.4172/2332-0877.1000211>.

43. Getahun M, Ameni G, Kebede A, Yaregal Z, Hailu E, Medihn G, et al. Molecular typing and drug sensitivity testing of *Mycobacterium tuberculosis* isolated by a community-based survey in Ethiopia. BMC Public Health. 2015;15:751.
44. Tilahun M, Ameni G, Desta K, Zewude A, Yamuah L, Abebe M, 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.
45. Haile B, Tafess K, Zewude A, Yenew B, Siu G, Ameni G. Spoligotyping and drug sensitivity of *Mycobacterium tuberculosis* isolated from pulmonary tuberculosis patients in the Arsi Zone of southeastern Ethiopia. New Microbes and New Infect. 2020;33:100620.
46. Shi J, Zheng D, Zhu Y, Ma X, Wang S, Li H, 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.
47. Bouklata N, Supply P, Jaouhari S, Reda Charof R, Seghrouchni F, Sadki K, 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.
48. Caleffi-Ferracioli KR, Baldin VP, Noguti EN, Siqueira VL, Scodr RB, Pavan FR, et al. Molecular characterization of *Mycobacterium tuberculosis* and *Mycobacterium bovis* isolates by *Enterobacterial Repetitive Intergenic Consensus*-PCR. Braz J Pharm Sci. 2018; 54:e17554.
49. Pareek M, Evans J, Innes J, Smith G, Hingley-Wilson S, Loughheed KE, et al. Ethnicity and mycobacterial lineage as determinants of tuberculosis disease phenotype. Thorax. 2013;68: 221-29.
50. Yimer SA, Norheim G, Namouchi A, Zegeye ED, Kinander W, Tønjum T, et al. *Mycobacterium tuberculosis* lineage 7 strains are associated with prolonged patient delay in seeking treatment for pulmonary tuberculosis in Amhara Region, Ethiopia. J Clin Microbiol. 2015;53:1301–09.
51. Tadesse M, Abebe G, Bekele A, Bezabih M, Rijk P, Meehan C, et al. The Predominance of Ethiopian Specific *Mycobacterium tuberculosis* Families and Minimal Contribution of *Mycobacterium bovis* in Tuberculous Lymphadenitis Patients in Southwest Ethiopia. Infect Genet Evol. 2017;55:251-59.
52. Ali S, Beckert P, Haileamlak A, Wieser A, Pritsch M, Heinrich N, et al. Drug resistance and population structure of *M. tuberculosis* isolates from prisons and communities in Ethiopia. BMC Infect Dis. 2016;16:687.

Tables

Figures

Supplementary Files

- Additionalfiles.doc
- Table.doc