BMC Infectious Diseases

RESEARCH ARTICLE   Open Access

Molecular epidemiology of clinical *Mycobacterium tuberculosis* complex isolates in South Omo, Southern Ethiopia

Biniam Wondale[1,2]* , Kwon Keehwan[3], Girmay Medhin[1], Takele Teklu[1,4], Temesgen Mohammed[1], Samuel Tolosa[1], Aboma Zewude[1,5], Friehiwot Amsalu[6,7], Rembert Pieper[3] and Gobena Ameni[1,8]

Abstract

Background: Tuberculosis (TB) is caused by *Mycobacterium tuberculosis* complex (MTBC). Mapping the genetic diversity of MTBC in high TB burden country like Ethiopia is important to understand principles of the disease transmission and to strengthen the regional TB control program. The aim of this study was to investigate the genetic diversity of *Mycobacterium tuberculosis* complex (MTBC) isolates circulating in the South Omo, southern Ethiopia.

Methods: MTBC isolates ($N = 156$) were genetically analyzed using spacer oligotyping (spoligotyping) and mycobacterial interspersed repetitive unit-variable number of tandem repeat (MIRU-VNTR) typing. Major lineages and lineages were identified using MTBC databases. Logistic regression was used to correlate patient characteristics with strain clustering.

Results: The study identified Euro-American (EA), East-African-Indian (EAI), Indo-Oceanic (IO), Lineage_7/ Aethiops vertus, *Mycobacterium bovis* and *Mycobacterium africanum* major lineages in proportions of 67.3% (105/156), 22.4% (35/156), 6.4% (10/156), 1.9% (3/156), 1.3% (2/156) and 0.6% (1/156), respectively. Lineages identified were Delhi/CAS 23.9% (37/155), Ethiopia_2 20.6% (32/155), Haarlem 14.2% (22/155), URAL 14.2%(22/ 155), Ethiopia_3 8.4% (13/155), TUR 6.5% (10/155), Lineage_7/Aethiops vertus 1.9% (3/155), Bovis 1.3% (2/155), LAM 1.3% (2/155), EAI 0.6% (1/155), X 0.6% (1/155) and Ethiopia H$_{37}$Rv-like strain 0.6% (1/155). Of the genotyped isolates 5.8% (9/155) remained unassigned. The recent transmission index (RTI) was 3.9%. Orphan strains compared to shared types (AOR: 0.09, 95% CI: 0.04–0.25) were associated with reduced odds of clustering. The dominant TB lineage in pastoral areas was EAI and in non-pastoral areas was EA.

Conclusion: The epidemiological data, highly diverse MTBC strains and a low RTI in South Omo, provide information contributing to the TB Control Program of the country.

Keywords: *M. Tuberculosis*, Molecular epidemiology, Spoligotyping, MIRU-VNTR, South Omo

* Correspondence: biniamw2005@yahoo.com; biniam.wondale@amu.edu.et
[1]Aklilu Lemma Institute of Pathobiology, Addis Ababa University, PO. Box 1176, Addis Ababa, Ethiopia
[2]Department of Biology, Arba Minch University, PO. Box 21, Arba Minch, Ethiopia
Full list of author information is available at the end of the article

Background

The *Mycobacterium tuberculosis* complex (MTBC) constitutes a group of mycobacteria which are 99.9% similar at the nucleotide level and the causative agents for tuberculosis (TB) [11]. Globally, TB became the leading cause of death from an infectious disease [39]. Ethiopia stands 12th in the world and 4th in Africa among the high TB burden countries with 24,000 TB deaths and 165,000 new TB cases in 2018 [39]. The current prevalence of MDR/RR-TB in Ethiopia is 0.71 and 16% in new and previously treated TB cases, respectively [39].

Understanding the molecular epidemiology of TB is important for regional disease control. For instance, distinct strains may be linked to outbreaks [10], high virulence [42], emergency of drug resistance [44], disease progression [43], and can point to the geographic origin of a strain [20, 33] as well as identify new lineages [17, 28].

The South Omo Zone is an administrative unit in the southern Ethiopia bordering Kenya and South Sudan. The area is remote with a poor infrastructure and high population diversity with 16 different ethnic groups. Forty-two percent of South Omo's residents including 15 ethnicities have a pastoral life style. The facilities for health care and education are underdeveloped especially in the pastoral regions [4]. A previous study suggested that the prevalence of TB among pastoralists is higher than in other socio-economic groups in Ethiopia [24]. In depth, high resolution molecular epidemiological surveys are required to characterize the diversity of MTBC isolates in this remote pastoral region precisely.

Beginning in the 1990s a number of molecular genotyping techniques have evolved to differentiate MTBC at the species and strain levels [22]. Whole genome sequencing is ideal to identify a strain type, but technically and informatically demanding and too expensive to characterize regional MTBC diversity [22]. Widely used in TB research are spacer oligotyping (spoligotyping) and mycobacterial interspersed repetitive units – variable numbers of tandem repeat (MIRU-VNTR). Spoligotyping targets a single locus and has less discriminatory power but is simple and cost effective. MIRU-VNTR targets numerous loci with increased discriminatory power. International databases and data analysis tools were created for both genotyping methods [22].

Most MTBC genotyping studies in Ethiopia employed spoligotyping [3, 6, 7, 16, 19, 26, 43]. We are aware of only very few Ethiopian studies using spoligotyping and MIRU-VNTR simultaneously to profile MTBC strains [1, 9, 36, 37, 43]. New lineages such as the MTBC lineage lineage_7/Aethiops vertus [17, 28], Ethiopia_2 and Ethiopia_3 [37] were newly assigned. Most surveys were in geographic areas more accessible than South Omo. The objective of the current study is to examine the MTBC population structures in the latter region and compare the data to those of other Ethiopian regions and globally.

Methods

Human subject recruitment and specimen collections

As part of a study to investigate TB in South Omo using both molecular epidemiology and systems biology methods between 2014 and 2017, we established a human subject protocol and a consent form to recruit more than 2000 individuals from different ethnic groups and administrative districts. We discerned non-pastoral areas (Debub Ari, Semen Ari and part of Male) and pastoral areas (Bena Tsemay, Hamer, Dassenech, Selamago and part of Male). The minimum age for enrolment was 15 years. According to 2007 population census, 577,673 people (7.5% urban and 92.5% rural) lived in the area with equal proportions of men and women [14]. Basic demographic and clinical data were collected using a pre-structured questionnaire.

Local translators explained the study's objectives, risks and benefits to those individuals who were not able to communicate in the Amharic language. Those individuals who consented to participate in the study were asked to provide fine needle aspirate (FNA) and sputum samples to test for EPTB and PTB, respectively. Samples were stored at + 4 °C and -20 °C following collection and transferred to the sample processing and microbial culture laboratory in order to screen for growth of MTBC isolates. Individuals who were diagnosed with TB were educated on and received DOTS treatment according to the national TB management guideline [18].

Generation of MTBC isolates

MTBC isolates were recovered from sputum and FNA samples for pulmonary TB (PTB) and extra-pulmonary TB (EPTB) cases, respectively. Mycobacterial cultures were performed at the Jinka Regional Laboratory (JRL). Samples were processed according to the modified Petroff's method by decontaminating the specimen with an equal volume of 4% NaOH for 15 min; the remaining volume was filled with phosphate buffer saline (PBS) and centrifuged at speed of 3000 g for 15 min. A drop of phenol red was added to a pellet as pH indicator and neutralized using 10% HCl. The neutralized pellet was inoculated onto two LJ media. One supplemented with glycerol and the other with pyruvate. Inoculated media were incubated at 37 °C for up to 8 weeks. Mycobacterial growth was monitored every week. Culture was considered negative after 8 weeks if no growth was observed. Positive colonies were further confirmed by Ziehl-Neelsen (ZN) staining. Heat treatment of mycobacterial isolates in dH_2O at 80 °C for 50 min was used for genomic DNA extraction without extensive DNA purification. Extracts were stored at -20 °C until they were used for molecular characterization.

Molecular characterization

DNA of mycobacterial isolates were primarily subjected to region of difference (RD) 9 typing [11]. Secondly, a single tube amplification method [40] was applied for genus typing for isolates which were not detected by RD9 typing. Thirdly, spoligotyping was performed for MTBC isolates as described by Kamerbeek et al., [23]. Laboratory results of spoligotyping were interpreted in binary format and lineages were assigned using an updated version of the SITVITWEB [13] and major lineages were assigned using the "TB insight" database. Isolates which have similar patter to those in the SITV IT database were assigned a Spoligo International Type (SIT) number. Isolates not assigned to SIT numbers were referred as "Orphans" spoligotype. Finally, MTBC DNA samples were subjected to MIRU-VNTR typing following an established procedure [35]. Laboratory results of MIRU-VNTR typing were interpreted using the MIRU-VNTR*plus* database (http://www.miru-vntrplus. org) to determine MTBC strain lineages and relatedness [2, 38]. A minimum spanning tree (MST) was constructed. Previously identified Ethiopian strains were assigned manually to Ethiopia_1 (Lineage 7), Ethiopia_2 and Ethiopia_3, which lack the spoligotype based spacer 4–24, 13 and 10–19, respectively [9, 17, 37]. The RD9 typing, genus typing and spoligotyping were performed at Aklilu Lemma Institute of Pathobiology (ALIPB), Addis Ababa University (AAU), Addis Ababa, Ethiopia whereas MIRU-VNTR typing was performed at J. Craig Venter Institute (JCVI), Maryland, USA.

Data analysis

Excel data were transformed into IBM SPSS version 20. The results were presented using descriptive statistics. The ability of genotyping methods to discriminate strains was calculated using Hunter-Gaston Discrimination Index (HGDI) [21]. The allelic diversity (h) based on MIRU-VNTR genotyping was calculated using HGDI and classified into relative discriminants based on previous [34] and newly proposed [27] index ranges. The recent transmission index (RTI) was calculated using a formula proposed by small et al. [33]. Isolates were said to be clustered if they have identical pattern based on spoligotyping and/or MIRU-VNTR. Clustering rate was determined using a formula (Nc-C)/N, where Nc = total number of isolates clustered, C = number of clusters, and N = total number of isolates [46]. The association between clustered strains and independent variables were computed using logistic regression models. The crude odds ratio (COR) and the adjusted odds ratio (AOR) were used to present the results. *P*-values of < 0.05 were considered statistically significant.

Results

Demographic and clinical characteristics of study participants

One thousand two hundred sputum and FNA samples of study participants were cultured for isolation of mycobacteria. Samples culture summary is presented in Table 1. In overall, 161 MTBC isolates were obtained upon RD9 and multiplex PCR based species and genus typing. Culture recovery rate vary from 2.5% (13/517) in community screened samples and 29.7% (33/111) in smear positive PTB samples from health facilities other than JGH to 67.3% (74/110) in smear positive PTB samples and 76.5% (26/34) in EPTB samples from JGH. Spoligotyping and MIRU-VNTR typing were carried out for a total of 156 isolates derived from 130 and 26 patients with clinical evidence of PTB and EPTB, respectively. Five samples were excluded due to insufficient quantities of DNA. Demographic and clinical characteristics of the corresponding 156 subjects are summarized in Table 2. The age range was 15–80 years with mean 32.9, median 30 and standard deviations 12.9. Female, PTB and new TB cases were 41.0% (64/156), 83.3% (130/156) and 95% (148/156), respectively.

Discriminatory power of genotyping methods and MIRU-VNTR allelic diversity

All South Omo MTBC isolates were genotyped using spoligotyping as well as MIRU-VNTR at the strain level, as presented in Tables 3 and 4. The discriminatory power of

Table 1 Culture summary of collected samples, South Omo, Ethiopia

Source	Samples from Health Facilities					Community > = 2 weeks cough in pastoral areas	Total
	JGH			Other than JGH[a]			
Sample type	Sputum Smear		FNA	Sputum Smear		Sputum	
	+ve	-ve		+ve	-ve		
Number of samples	110	404	34	111	24	517	1200
Recovered MTBC N (%)	74 (67.3)	14 (3.5)	26 (76.5)	33 (29.7)	1 (4.2)	13 (2.5)	161 (13.4)
Contamination N (%)	4 (3.6)	21 (5.2)	0 (0.0)	6 (5.4)	4 (16.7)	64 (12.4)	99 (8.3)

Abbreviations: *MTBC Mycobacterium tuberculosis* complex, *JGH* Jinka Genral Hospital; [a]Other than JGH includes- Health centers in D/Ari, BenaTsemay, Dassenech, Hammer, Male, Gnagatom and Selamago

Table 2 Demographic and clinical characteristics of study participants in molecular epidemiology, South Omo, Ethiopia

Characteristics	Residential areas									Unspecified	Total N (%)
	Non-pastoral (N = 70)			Pastoral (N = 67)							
	D/Ari N (%)	Jinka N (%)	[a]Other areas N (%)	B/Tsemay N (%)	Dassenech N (%)	Hamer N (%)	Male N (%)	Gnangatom N (%)	Selemago N (%)		
Sex											
Female	23 (48.9)	6 (37.5)	1 (14.3)	6 (42.9)	4 (44.4)	4 (26.7)	5 (41.7)	1 (50.0)	7 (46.7)	7 (36.8)	64 (41.0)
Male	24 (51.1)	10 (62.5)	6 (85.7)	8 (57.1)	5 (55.6)	11 (73.3)	7 (58.3)	1 (50.0)	8 (53.3)	12 (63.2)	92 (59.0)
Age in years											
15–29	19 (40.4)	11 (68.8)	2 (28.6)	7 (50.0)	4 (44.4)	2 (13.3)	7 (58.3)	1 (50.0)	11 (73.3)	6 (31.6)	70 (44.9)
30–44	19 (40.4)	3 (18.8)	5 (71.4)	3 (21.4)	1 (11.1)	9 (60.0)	2 (16.7)	1 (50.0)	3 (20.0)	13 (68.4)	59 (37.8)
45+	9 (19.1)	2 (12.5)	0 (0.0)	4 (28.6)	4 (44.4)	4 (26.7)	3 (25.0)	0 (0.0)	1 (6.7)	0 (0.0)	27 (17.3)
Type of TB											
PTB	37 (78.7)	13 (81.2)	3 (42.9)	13 (92.9)	9 (100.0)	15 (100.0)	10 (83.3)	2 (100.0)	9 (60.0)	19 (100.0)	130 (83.3)
EPTB	10 (21.3)	3 (18.8)	4 (57.1)	1 (7.1)	0 (0.0)	0 (0.0)	2 (16.7)	0 (0.0)	6 (37.5)	0 (0.0)	26 (16.7)
Medical history											
New cases	45 (95.7)	16 (100.0)	7 (100.0)	12 (85.7)	9 (100.0)	12 (80.0)	12 (100.0)	2 (100.00)	14 (93.3)	19 (0.0)	148 (94.9)
Retreat. Cases	2 (4.3)	0 (0.0)	0 (0.0)	2 (14.3)	0 (0.0)	3 (20.0)	0 (0.0)	0 (0.0)	1 (6.7)	0 (0.0)	8 (5.1)
Total	47 (30.1)	16 (10.3)	7 (4.5)	14 (9.0)	9 (5.8)	15 (9.6)	12 (7.7)	2 (1.3)	15 (9.6)	19 (12.2)	156 (100.0)

[a]Other areas, include Addis Ababa, Konso, Semen Ari and Sawula; D/Ari, Debub Ari; B/Tsemay, Bena Tsemay; Retreat. Cases, anti-TB retreatment cases

spoligotying and 24-loci MIRU-VNTR were 0.9700 and 0.9995, respectively. The analysis of spoligotypes resulted in 66 different patterns of which 24 clustered and 42 were unique. MIRU-VNTR analysis resulted in 149 different patterns of which six clustered and 143 were unique. The clustering rates were 57.7% ((114–24)/156) for spoligotyping and 3.9% ((12–6)/155) for the MIRU-VNTR analysis. The RTIs for South Omo were calculated to be 57.7% based on spoligotyping and 3.9% based on MIRU-VNTR data.

Allelic diversities (h) for each MIRU-VNTR locus are presented in Table 4. MIRU-VNTR loci are classified based on their allelic diversity and ability to discriminate among the isolates. Eleven MIRU-VNTR loci (424, 577, 802, 960, 1644, 1955, 2163b, 2347, 2996, 3192 and 4052) were highly discriminatory while 10 MIRU-VNTR loci (580, 2059, 2165, 2401, 2461, 2531, 3171, 3690, 4156 and 4348) were moderately discriminatory. The remaining three loci showed poor discriminatory values for the isolates under study. However, 11-loci are considered as

moderately discriminatory according to newly proposed discriminatory values [27].

Spoligotyping and MIRU-VNTR based identification of lineages and sub-lineages

The results of spoligotyping and MIRU-VNTR are presented in Tables 5 and 6. According to spoliogotyping results 76.3% (119/156) of the isolates belonged to 36 shared types (SIT numbers). The remaining, about one-qaurter of the isolates (37/156) were orphan strains. Euro-American (EA) was the most prevalent lineage with 67.3% (105/156) followed by East-African-Indian (EAI) with 22.4% (35/156) and Indo-Oceanic (IO) with 6.4% (10/156) of the isolates. Interestingly, EA was more prevalent in non-pastoral areas while EAI was more prevalent in pastoral areas (Fig. 1). Two *M. bovis* isolates were identified in pastoral areas.

According to SITVIT analysis, the identified lineages were T 30.1% (47/156), CAS1_Delhi 10.9% (17/156),

Table 3 Discriminatory power of genotyping methods

Typing Method	Different patterns (N) (%)	Clusters (N)	Cluster size (N)	Clustered isolates (N) (%)	Unique isolates (N) (%)	Clustering rate %	HGDI
Spoligotyping	66/156 (42.3)	24	2–17	114 (73.1)	42 (26.9)	57.7	0.97
MIRU-VNTR-24	149/155 (96.1)	6	2	12 (7.7)	143 (92.3)	3.9	0.9995
MIRU-VNTR −15	148/155 (95.5)	7	2	14 (9.0)	141 (91.0)	4.5	0.9994
MIRU-VNTR-12	102/155 (65.8)	23	2–8	76 (49.0)	79 (51.0)	34.2	0.9897
Spoligotyping + MIRU-VNTR	155/155 (100)	0	0	0	155 (100)	100	1

HGDI Hunter Gaston Discrimination index

Table 4 Occurrence of MIRU-VNTR alleles and allelic diversity

MIRU-VNTR			Allele No.[#]																*h*	Discriminant
24	15	12	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		Status
154		▓	1	7	135	6	2	2	1							1				0.239	Poor
424	▓		10	43	81	14	3	2	1	1										0.641	**High**
577	▓			38	25	84	6							1				1		0.623	**High**
580	▓	▓	12	129	3	5	1	3	2											0.301	Moderate
802	▓	▓		40	57	52	3	1	1	1										0.690	**High**
960	▓	▓	2		30	15	46	24	29	4	2		2	1						0.810	**High**
1644	▓	▓	12	35	74	25	6	1		1										0.688	**High**
1955	▓			2	17	77	54		1					1			2	1		0.623	**High**
2059		▓	2	17	126	7				1						1				0.318	Moderate
2163b	▓			20	57	6	30	8	10	15	2	3		1			3			0.797	**High**
2165	▓			7	82	56	4	3		1		1								0.585	Moderate
2347				26	45	79	3									1				0.627	**High**
2401	▓			2	53	9	90	1												0.546	Moderate
2461			1	7	117	23	5	1	1											0.408	Moderate
2531		▓			22	11	116	4	1										1	0.417	Moderate
2687		▓	2	137	13		1					1	1							0.213	Poor
2996	▓	▓		49	1	1	39	37	20	5	3									0.767	**High**
3007		▓		2	9	139	3				1					1				0.193	Poor
3171				2	13	115	21	2										1		0.419	Moderate
3192	▓	▓	1	2	46	76	26	3					1							0.647	**High**
3690	▓		2	5	3	103	27	5	6	3	1									0.527	Moderate
4052	▓			1	1	9	20	35	24	52	7	3		1		1				0.792	**High**
4156	▓		4	24	102	6	3	1	2	2				1		1	2	1		0.505	Moderate
4348		▓	2	6	115	28		3		1										0.418	Moderate

The MIRU-VNTR locus in the Mtb genome; [#] the number of alleles pertains to the frequency with which a distinct repeat unit was identified among the 155 isolates. High discriminant (h > 0.6), moderate discriminant (0.3 < = h < =0.6) and poor discriminant (h < 0.3) [34]. Shaded boxes show alleles used for the 15-loci MIRU-VNTR and the 12-loci MIRU-VNTR methods

Haarlem 10.3% (16/156), Ethiopian 1.9% (3/156), Turkey 1.9 (3/156), MANU 1.3% (2/156), LAM 0.6% (1/156), EAI1_SOM 0.6% (1/156), BOV 0.6% (1/156) and X 0.6% (1/156). Almost one-third of isolates (56/156) were not registered in updated version of SITVITWEB database (SITVIT2). According to MIRU-VNTR*plus* analysis result (S1 Fig, Table 5 and Table 6), the identified lineages were Delhi/CAS 23.9% (37/155), Ethiopia_2 20.6% (32/155), Haarlem and URAL 14.2% (22/155) each, Ethiopia_3 8.4% (13/155), TUR 6.5% (10/155), Ethiopia_1 1.9% (3/155), Bovis and LAM 1.3% (2/155) each, and EAI, X and Ethiopia H₃₇Rv-like strain 0.6% (1/155) each. Around 6% of the isolates were not assigned into lineages using MIRU-VNTRplus database.

Minimum spanning tree (MST) analysis

MST shows the relationship among strains and lineages based on MIRU-VNTR loci variant count. Five well-defined branches (Complex 1, 2, 3, 4 and 5) and four additional sub-branches (Complex 6, 7, 8 and 9) expanding from Complex 1 are featured in Fig. 2. Each complex largely consisted of distinct sub-lineages. Part of Complex 1 are the Ethiopia_2 sub-lineage and the additional sub-lineages H37Rv, Bovis, unidentified strains, and *M. canetti* (used as a reference). Complexes 2, 3, 4 and 5

are enriched in Ethiopian_3 and TUR, URAL, Delhi/CAS and Haarlem strains, respectively.

Factors associated with strain clustering

Clustering of *M. tuberculosis* strains association with different factors were analyzed using spoligotyping pattern. Clustering was significantly associated with the residential area, the major lineage type and the SIT status before adjusted to confounding factors (Table 7). However, only being part of the orphan group compared to shared type (AOR: 0.19, 95% CI: 0.04–0.25) was significantly associated with reduced odds of clustering in the area under study after adjustment with confounding factors.

Discussion

This study was the first of its kind to analyze the MTBC population structure and transmission dynamics in the South Omo Zone, southern Ethiopia. The study included PTB and EPTB patients, and identified highly diverse lineages. The clustering rate/RTI was low in the study area. Logistic regression analysis showed that clustering of strains was associated with SIT status.

Health facilities other than JGH in the study area are in range of 14 to 250 km from Jinka town where JGH and the Regional Laboratory are located. Due to feasibility, samples were stored in health facilities at -20 °C from

Table 5 Major lineages and lineages of MTBC strains concordant with existing SIT numbers

SIT	Octal code	Binary code	Major lineage (TB insight)	Lineage (SITVITWEB)	Lineage (MIRU-VNTR*plus*)	Isolates N (%)
25	703777740003171		EAI	CAS1_DELHI	Delhi/CAS	3 (1.9)
26	703777740003771		EAI	CAS1_DELHI	Delhi/CAS	5 (3.2)
37	777737777760771		EA	T3	Ethiopia_2	17 (10.9)
41	777777404760771		EA	Turkey	TUR	3 (1.9)
42	777777607760771		EA	LAM9	LAM	1 (0.6)
47	777777774020771		EA	H1	Haarlem	7 (4.5)
48	777777777413731		IO	EAI1_SOM	EAI	1 (0.6)
53	777777777760771		EA	T1	Haarlem	1 (0.6)
53	777777777760771		EA	T1	Ethiopia_2	3 (1.9)
53	777777777760771		EA	T1	TUR	1 (1.3)
53	777777777760771		EA	T1	Ethiopia_3	1 (1.3)
53	777777777760771		EA	T1	URAL	1 (0.6)
53	777777777760771		EA	T1	Unidentified	1 (0.6)
54	777777777763771		IO	MANU2	TUR	1 (0.6)
73	777737777760731		EA	T	Ethiopia_2	5 (3.2)
78	777777777760711		EA	T1	Ethiopia_3	1 (0.6)
118	777767777760771		EA	T1	Unidentified	1 (0.6)
149	777000377760771		EA	T3_ETH	Ethiopia_3	10 (6.4)
289	703777740003571		EAI	CAS1_DELHI	Delhi/CAS	3 (1.9)
334	577777777760771		EA	T1	TUR	1 (0.6)
336	777776777760731		EA	X1	X	1 (0.6)
374	777777776000771		EA	Unknown	URAL	1 (0.6)
442	777737377760771		EA	T	Ethiopia_2	1 (0.6)
764	777757777720771		EA	H3	Haarlem	2 (1.3)
777	777777777420771		EA	Ural_1	URAL	8 (5.1)
910	700000007177771		Lineage_7	Ethiopian	Ethiopia_1	1 (0.6)
952	603777740003771		EAI	CAS1_Delhi	Delhi/CAS	5 (3.2)
1153	772700000003631		IO	Unknown	Unidentified	2(1.3)
1200	703777747777771		EAI	Unknown	Delhi/CAS	5 (3.2)
1547	777727777760771		EA	T3	Ethiopia_2	1 (0.6)
1606	603777740000771		EAI	Unknown	Delhi/CAS	7 (4.5)
1729	700000004177771		Lineage_7	Ethiopian	Ethiopia_1	2 (1.3)
523	777777777777771		IO	Manu_ancestor		1 (0.6)
2027	777777777000731		EA	Unknown	URAL	1 (0.6)
2364	703601740003771		EAI	CAS1_Delhi	Delhi/CAS	1 (0.6)
2501	777777774220771		EA	H3	Haarlem	1 (0.6)
3134	777737377720771		EA	H3	Haarlem	6 (3.8)
3136	760000001760771		EA	Unknown	Ethiopia_2	3 (1.9)
3411	777002377760771		EA	T1	Unidentified	1 (0.6)
3749	402773777757600		M. bovis	BOV	Bovis	1 (0.6)
3862	737777777760700		EA	T1	$H_{37}Rv$	1 (0.6)
Total						**119 (76.3)**

Abbreviations: *EA* Euro-American, *EAI* East-African-Indian, *IO* Indo-oceanic

a week to 3 weeks. The variation in culture recovery rate of MTBC isolates in this study possibly associated with sample storage conditions. There was continuous electric interruption in the Zone which affects the storage temperature which could compromise the viability of MTBC in the sample. JGH had its own backup generator that might be the reason for better recovery rate for samples from JGH. In connection, the overall culture recovery rate in this study is less than previous studies in Ethiopia [47, 48].
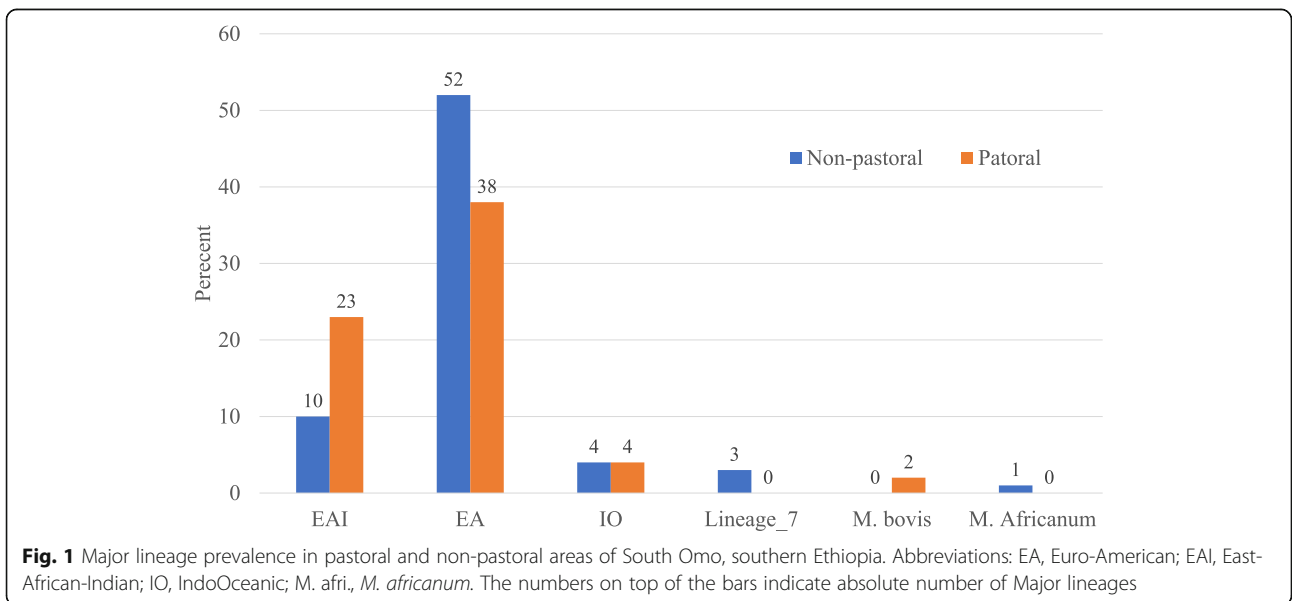
In contrast to previous reports in the study area [8, 41], the number of EPTB cases in this study was low. From personal observations, the low number of EPTB cases in this study might be due to lack of skilled pathologist to take FNA samples whereas in previous studies relied on clinical symptoms. Spoligotyping and MIRU-

Table 6 Major lineages and lineages of orphan MTBC strains

SIT	Octal code	Binary code	Major lineage (TB insight)	Lineage (SITVITWEB)	Lineage (MIRU-VNTR*plus*)	Isolates N (%)
Orphan-1	000003437540771		EA	Unknown	Unidentified	1 (0.6)
Orphan-2	001777777420731		EA	Unknown	URAL	2 (1.3)
Orphan-3	400003435540771		EA	Unknown	URAL	1 (0.6)
Orphan-4	400022237560771		EA	Unknown	Haarlem	1 (0.6)
Orphan-6	400343777776600		M. bovis	Unknown	Bovis	1 (0.6)
Orphan-7	500027437560771		EA	Unknown	Haarlem	1 (0.6)
Orphan-8	501767400142771		EAI	Unknown	URAL	1 (0.6)
Orphan-9	501777777423771		IO	Unknown	URAL	1 (0.6)
Orphan-10	540000001740661		M. afri.	Unknown	Ethiopia_2	1 (0.6)
Orphan-11	577737377720771		EA	Unknown	Haarlem	1 (0.6)
Orphan-12	577767777400771		IO	Unknown	Delhi/CAS	1 (0.6)
Orphan-13	603777740000271		EAI	Unknown	Delhi/CAS	1 (0.6)
Orphan-14	603777740000671		EAI	Unknown	Delhi/CAS	1 (0.6)
Orphan-15	603777757560771		EA	Unknown	Delhi/CAS	1 (0.6)
Orphan-17	703777744203771		EAI	Unknown	Delhi/CAS	2 (1.3)
Orphan-18	717777777763771		EA	Unknown	Unidentified	1 (0.6)
Orphan-21	763777777563771		EA	Unknown	Delhi/CAS	1 (0.6)
Orphan-24	772702000003631		IO	Unknown	Unidentified	2 (1.3)
Orphan-25	777002000020771		EA	Unknown	Ethiopia_3	1 (0.6)
Orphan-27	777347777777771		IO	Unknown	URAL	1 (0.6)
Orphan-28	777727377720771		EA	Unknown	Haarlem	1 (0.6)
Orphan-29	777737277760771		EA	Unknown	Ethiopia_2	1 (0.6)
Orphan-30	777737337720771		EA	Unknown	Haarlem	1 (0.6)
Orphan-32	777737777000731		EA	Unknown	URAL	2 (1.3)
Orphan-33	777777404560771		EA	Unknown	TUR	1 (0.6)
Orphan-34	777777444760771		EA	Unknown	TUR	3 (1.9)
Orphan-35	777777647760701		EA	Unknown	LAM	1 (0.6)
Orphan-38	777777777420571		EA	Unknown	URAL	2 (1.3)
Orphan-39	401777777420731		EA	Unknown	URAL	1 (0.6)
Orphan-40	603777700017771		EAI	Unknown	Delhi/CAS	1 (0.6)
Total						**37 (23.7)**

Abbreviations CBN Conformal Bayesian network, *EA* Euro-American, *EAI* East-African-Indian, *IO* IndoOceanic, *M. afri.*, *M. africanum*

Fig. 1 Major lineage prevalence in pastoral and non-pastoral areas of South Omo, southern Ethiopia. Abbreviations: EA, Euro-American; EAI, East-African-Indian; IO, IndoOceanic; M. afri., *M. africanum*. The numbers on top of the bars indicate absolute number of Major lineages

Table 7 Association of different factors with clustering of *M. tuberculosis* complex

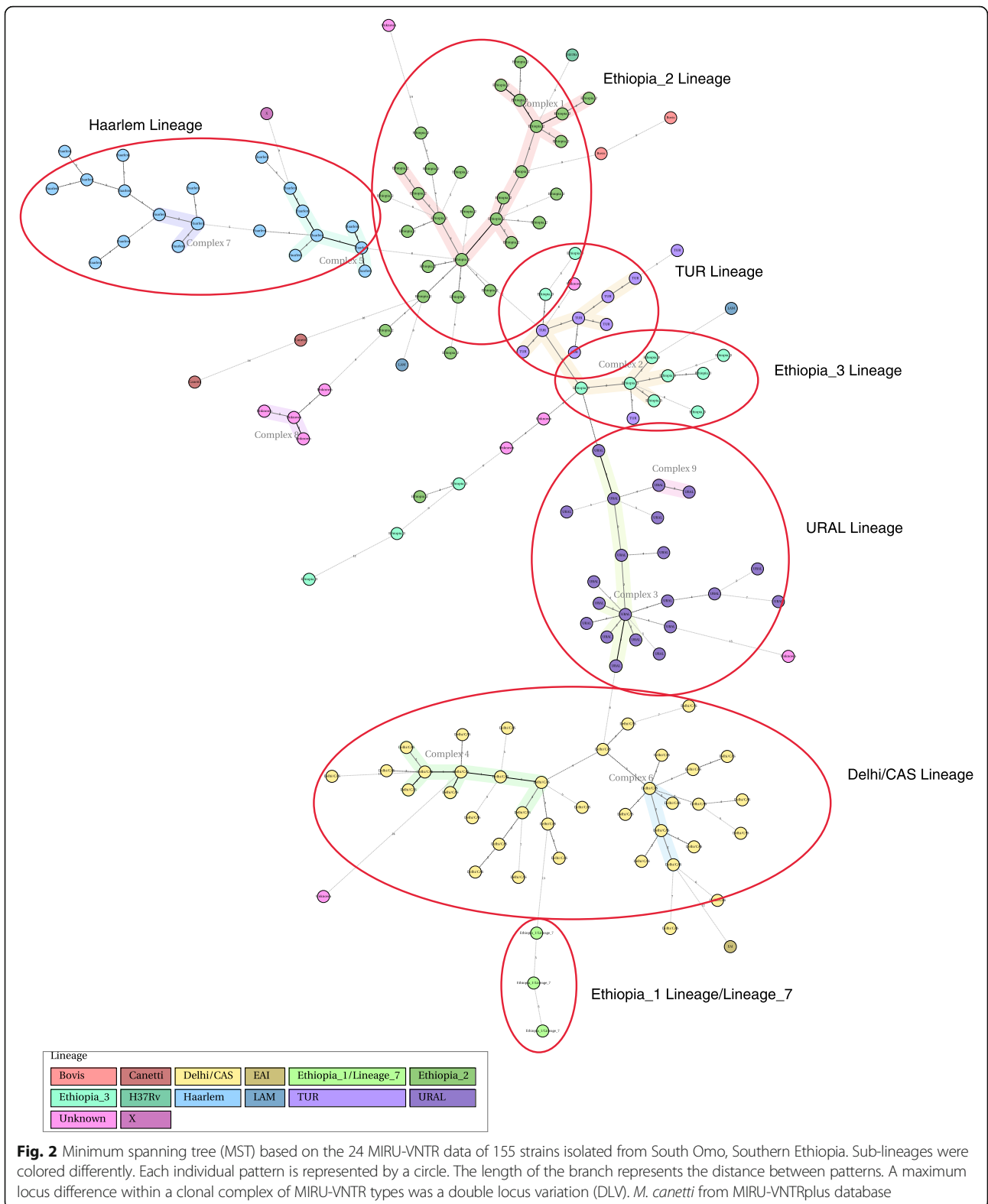| Characteristics	Total N (%)	Clustered N (%)	COR (95% CI)	P value	AOR (95% CI)	P value
Sex						
Female	64 (41.0)	46 (71.9)	1.00		NA	
Male	92 (59.0)	68 (73.9)	1.11 (0.54–2.27)	0.778		
Age in years						
15–29	70 (44.9)	55 (78.6)	1.00		NA	
30–44	59 (37.8)	41 (69.5)	0.62 (0.28–1.38)	0.241		
> =45	27 (17.3)	18 (66.7)	0.55 (0.20–1.46)	0.227		
TB Type						
PTB	130 (83.3)	92 (70.8)	1.00		NA	
EPTB	26 (16.7)	22 (84.6)	2.27 (0.73–7.04)	0.155		
Medical history						
New cases	148 (94.9)	107 (72.3)	1.00		NA	
Retreat. Cases	8 (5.1)	7 (87.5)	2.68 (0.32–22.48)	0.363		
Residential area						
Non-pastoral	70 (51.1)	48 (68.6)	1.00		1	
Pastoral	67 (48.9)	56 (83.6)	**2.33 (1.03–5.30)**	**0.043**	2.29 (0.82–6.35)	0.112
Major lineages[a]						
EAI	35 (22.9)	30 (85.7)	1.00		1.00	
EA	105 (68.6)	78 (74.3)	0.48 (0.17–1.37)	0.170	0.59 (0.16–2.22)	0.439
IO	10 (6.5)	4 (40.0)	**0.11 (0.02–0.54)**	**< 0.006**	0.24 (0.03–1.83)	0.169
Lineage_7	3 (2.0)	2 (66.7)	0.33 (0.03–4.40)	0.404	0.22 (0.01–3.50)	0.110
SIT status						
Shared type	119 (76.3)	101 (84.9)	1.00		1.00	
Orphan	37 (23.7)	13 (35.1)	**0.10 (0.04–0.22)**	**< 0.001**	**0.09 (0.04–0.25)**	**< 0.001**
Lineage[a]						
Delhi/CAS	37 (26.6)	30 (81.1)	1.00		NA	
Haarlem	22 (15.8)	16 (72.7)	0.62 (0.18–2.17)	0.456		
URAL	22 (15.8)	15 (68.2)	0.50 (0.15–1.69)	0.264		
Ethiopia_2	32 (23.0)	28 (87.5)	1.63 (0.43–6.19)	0.470		
TUR	10 (7.2)	7 (70.0)	0.54 (0.11–2.65)	0.452		
Ethiopia_3	13 (9.4)	11 (84.6)	1.28 (0.23–7.14)	0.776		
Ethiopia_1	3 (2.2)	2 (66.7)	0.47 (0.04–5.90)	0.556		

[a]Sub-categories having only singletons or clustered (based on strain clustering data using spoligotype patterns) were excluded from the regression analysis; NA: Not applicable; some abbreviations sued here were defined in the legend of Table 2

VNTR are recommended methods for the profiling of MTBC isolates [23, 35]. Both genotyping methods in this study were in range of highly discriminant [27, 34]. MIRU-VNTR has higher discriminatory value than spoligotyping as shown here and in earlier studies [1, 9, 36, 37].

Spoligotyping of South Omo MTBC isolates resulted in a clustering rate of 57.7%%. This rate agrees with a previous study in Gambella, Southwest Ethiopia [3] which is geographically proximate to the present study site. The rate is lower than a national survey [19] and that of studies in Addis Ababa [26], Northwest Ethiopia [37], Eastern Ethiopia [7], and central Ethiopia [6]. But it was a study in Western Ethiopia [16]. The MIRU-VNTR clustering rate was 3.9%. The rate is lower than other studies in Ethiopia [1, 9, 36, 37, 43] and higher than a Chinese study [15]. Such variability in clustering rate among studies could be due to differences in geography, population density, ethnicity and socio-economic diversity [31]. The low clustering rate in our study could also be associated with low culture recovery rate of samples which make potential isolates from the study population not to be genotyped and/or presence of low TB transmission in South Omo due to geographic expanse which disfavor the transmission as a result of very less crowdedness in the community.

Fig. 2 Minimum spanning tree (MST) based on the 24 MIRU-VNTR data of 155 strains isolated from South Omo, Southern Ethiopia. Sub-lineages were colored differently. Each individual pattern is represented by a circle. The length of the branch represents the distance between patterns. A maximum locus difference within a clonal complex of MIRU-VNTR types was a double locus variation (DLV). *M. canetti* from MIRU-VNTRplus database

Most MIRU-VNTR alleles in this study were highly and moderately discriminant based on the allelic diversity (h) which is an indirect indicator of the sample representativeness of the study population [34]. Values of h more than 0.8 and less than 0.1 are unsuitable for genotyping [30]. In the present study, all MIRU-VNTR

loci were suitable for genotyping of the isolates except locus 960 with an *h* value of 0.810.

This study identified six major lineages (EA, EAI, IO, lineage_7, *M. bovis* and *M. africanum*). Four isolates were identified using TB insight database as *M. africanum* which is known to be localized in West Africa. However, three of the four isolates were re-identified using updated version of SITVITWEB as Ethiopian and considered as Lineage_7/Ethiopia_1 [9, 17, 37] but one isolate remained unknown. The population in the area is endogenous which make the plausibility of an unknown isolate to be *M. africanum* less probable. In general, this might imply the need of updating TB Insight database with newly generated MTBC data from the horn of Africa, particularly Ethiopia.

EA is the most dominant lineage in the world [32], ranging in Ethiopia from 32.5% near the border to South Sudan [3] to 86.8% in central Ethiopia [6]. This might highlight the introduction of EA lineages from abroad through the capital city and their expansion to the peripheral areas. In addition, the existence of EA in high number in contrast to Ethiopian lineage, Lineage_7, in the country might indicate the high transmission ability of EA. The contribution of *M. bovis* for TB was low which is supported by data from other studies [7, 17]. While larger clinical studies are needed, our data suggested that the role of *M. bovis* as a causative agent of TB in pastoral area presumably linked to contact with infected cattle and, consumption of raw milk and meat [5].

According to the updated version of SITVITWEB database, the T lineage was predominant in this study. This lineage accommodates MTBC strains which do not have phylogeographic specificity [12]. Among the T sub-lineages, T3 (42.2%) was the most dominant one in this study. It is also called Ethiopia_2 [9] which is followed by T3-ETH, also called Ethiopia_3 [9]. These isolates supposed to be phylogeographically specific to Ethiopia including well defined Ethiopian lineage also called lineage_7 [17]. The CAS1_Delhi lineage was the second predominant lineage in the study area which was followed by Haarlem. The Haarlem is believed to descend from the European continent [12]. The predominance of the T lineage in Ethiopia and the CAS lineage in Tanzania [25] and Kenya [29] supports the notion of enrichment of MTBC strains in certain geographies. The Turkey lineage was present in the study area which is believed to be specific to Turkey [45]. This might be associated with presence of Turkey investors in the South Omo. In addition, Ural_1, Manu, Bov, EAI_SOM, and X lineages were identified in less frequency in the area. It is plausible that the observed phylogeographic diversity has linked to considerable international tourism in South Omo, Ethiopia.

When we look at MIRU-VNTR*plus* based lineage in the present study, Delhi/CAS was the predominat one which is in agreement with previous studies in Ethiopia [9, 36, 37, 43]. Ethiopia_2 is the predominant Ethiopia specific lineage followed by the Ethiopia_3 and lineage_7, similar to a previous study in geographic proximity, Southwestern Ethiopia [36]. But in studies at far distance, in Northwestern Ethiopia [9, 43], the predominant lineage was lineage_7 followed by the Ethiopia_3 and Ethiopia_2 lineages. These findings indicate that the distribution of Ethiopia specific lineages differ moderately from area to area within the country localities. This information is useful for the country's TB Control Program. Almost 6% of isolates in this study were not assigned into lineages which requires further study and introduction into the genotype database. The relationship among lineages in the MST based on MIRU-VNTR loci was in agreement with similar studies conducted in Ethiopia [9, 37].

Based on the generated data from spoligotyping and MIRU-VNTR, it is possible to say that these two methods can complement each other. But they have different precisions. For instance, MIRU-VNTRplus can identify 37 Delhi/CAS, 22 Haarlem and 22 Ural lineages whereas SITVITWEB can only identify 17 CAS1_Delhi, 16 Haarlem and eight Ural_1. This differences probably associated with the algorithm used by such databases. Finally, from all isolates of South Omo in this study, MIRU-VNTRplus and SITVITWEB didn't assign nine and 56 isolates into lineages, respectively.

The multivariate logistic regression analysis in this study showed none of the variables had association with strain clustering except SIT shared status. Orphan strains were less likely to cluster compared to shared strains which implies that shared strains have higher transmission rate compared to orphan strains in the study area.

We contend that the number of genotyped isolates is sufficient for a primary representation of South Omo's MTBC population structure, assessment of clustering rates and RTIs. However, having less culture recovery rate from samples other than JGH in this study limited our ability to identify clusters and RTIs more comprehensively.

Conclusion

The MTBC strains derived from TB patients in South Omo were highly diverse while the RTI was low in this marginalized region, as compared to other studies in high TB burden countries. Nonetheless, the genotyping data are useful as an input to map the population structure of MTBC in Ethiopia and to the TB Control Program in the pastoral region.

Wondale et al. BMC Infectious Diseases   (2020) 20:750

Page 11 of 12

Supplementary information

> **Additional file 1.** Phylogenetic tree for 155 isolates was constructed based on 24 locus MIRU-VNTR. The dendrogram was calculated using neighbor-joining clustering algorithms.

Abbreviations
AAU: Addis Ababa University; ALIPB: Aklilu Lemma Institute of Pathobiology; AOR: Adjusted odds ratio; CI: Confidence interval; COR: Crude odds ratio; DOTS: Direct observed treatment short course; EA: Euro-American; EAI: East-African-Indian; EPTB: Extrapulmonary tuberculosis; FNA: Fine needle aspirate; HGDI: Hunter-Gaston Discriminatory Index; JCVI: J. Craig Venter Institute; JRL: Jinka Regional Laboratory; IO: Indo-Oceanic; LJ: Lowenstein Jensen; MIRU-VNTR: Mycobacterial interspersed repetitive unit-variable number of tandem repeat; MTBC: *Mycobacterium tuberculosis* complex; MST: Minimum spanning tree; PCR: Polymerase chain reaction; PTB: Pulmonary tuberculosis; RD: Region difference; RTI: Recent transmission index; TB: Pulmonary tuberculosis; ZN: Ziehl-Neelsen

Authors' contributions
BW participated in the study conception, design, data collection, laboratory work, data analysis, data interpretation and manuscript writing. KK participated in its laboratory work, data analysis, data interpretation and manuscript writing. GM participated in its design, data analysis, data interpretation and manuscript writing. TT and FA participated in its data collection and manuscript writing. TM, ST and AZ participated in its laboratory work and manuscript writing. RP and GA participated in the study conception, design, data interpretation and manuscript writing. The authors have read and approved the manuscript.

Availability of data and materials
The data generated or analysed during this study are not publicly available due to it may compromise patient anonymity but are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
The study obtained ethical clearance from ALIPB, AAU Institutional Review Board (ALIPB/IRB/22-B/2012/13) and from the National Research Ethics Committee (Ref No:3.10/785/07). Permission to conduct the study was also obtained from the South Omo Zone Health Department. Written informed consent was obtained from each of the study participants and from their parent or guardian for those who were less than 18 years. Finally, the J. Craig Venter Institutional Review board also approved the study under protocol number 2014–198.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Aklilu Lemma Institute of Pathobiology, Addis Ababa University, PO. Box 1176, Addis Ababa, Ethiopia. [2]Department of Biology, Arba Minch University, PO. Box 21, Arba Minch, Ethiopia. [3]J. Craig Venter Institute, Rockville, MD, USA. [4]Department of Immunology and Molecular Biology, University of Gondar, PO. Box 196, Gondar, Ethiopia. [5]Ethiopian Public Health Institute, P.O.Box 1242, Addis Ababa, Ethiopia. [6]Jinka General Hospital, Jinka, Ethiopia. [7]Department of General Surgery, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia. [8]Department of Veterinary Medicine, College of Food and Agriculture, United Arab Emirates University, P.O.Box 15551, Al Ain, UAE.

References

1. 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.
2. Allix-Beguec C, Harmsen D, Weniger T, Supply P, Niemann S. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. J Clin Microbiol. 2008;46:2692–9.
3. Asebe G, Worku A, Zewude A, Tafess K, Ameni G. Molecular characterization of *Mycobacterium tuberculosis* complex in Gambella region, Southwest Ethiopia. Afr J Microbiol Res. 2016;10:8.
4. AU. Policy Framework for Pastoralism in Africa: Securing, Protecting and Improving the Lives, Livelihoods and Rights of Pastoralist Communities Addis Ababa. Ethiopia: Africa Union; 2010.
5. Ayele WY, Neill SD, Zinsstag J, Weiss MG, Pavlik I. Bovine tuberculosis: an old disease but a new threat to Africa. Int J Tuberculosis Lung Dis. 2004;8:924–37.
6. Bedewi Z, Worku A, Mekonnen Y, Yimer G, Medhin G, Mamo G, et al. Molecular typing of *Mycobacterium tuberculosis* complex isolated from pulmonary tuberculosis patients in Central Ethiopia. BMC Infect Dis. 2017;17:184.
7. Belay M, Ameni G, Bjune G, Couvin D, Rastogi N, Abebe F. Strain diversity of *Mycobacterium tuberculosis* isolates from pulmonary tuberculosis patients in Afar pastoral region of Ethiopia. Biomed Res Int. 2014;2014 https://doi.org/10.1155/2014/238532.
8. Berg S, Schelling E, Hailu E, Firdessa R, Gumi B, Erenso G, et al. 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.
9. Biadglegne F, Merker M, Sack U, Rodloff AC, Niemann S. 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. 2015;10:e0137865.
10. Bifani PJ, Mathema B, Liu Z, Moghazeh S, Shopsin B, Tempalski B, et al. Identification of a w variant outbreak of *Mycobacterium tuberculosis* via population-based molecular epidemiology. JAMA. 1999;282:2321–7.
11. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc Natl Acad Sci U S A. 2002;99:3684–9.
12. Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj SA, et al. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol. 2006;6:23.
13. Couvin D, David A, Zozio T, Rastogi N. Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the *Mycobacterium tuberculosis* genotyping database. Infect Genet Evol. 2019;72:31–43 PMID: 30593925.
14. CSA. Summary and statistical report of the 2007 population and housing census- population size by age and sex: Federal Democratic Republic of Ethiopia, population census commission, with support from UNFPA; 2008.
15. Dai GM, Zhang ZG, Ding PJ, Zhang Q, Wang L, Wang LX, et al. Differences in the population of genetics of Mycobacterium tuberculosis between urban migrants and local residents in Beijing, China. Chin Med J. 2013;126:4066–71.
16. Disassa H, Tafess K, Worku A, Ameni G. A preliminary study on molecular characterization of Mycobacterium tuberculosis in Benishangul Gumuz region, Western Ethiopia. B M R J. 2015;10:1–10.
17. Firdessa R, Berg S, Hailu E, Schelling E, Gumi B, Erenso G, et al. Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. Emerg Infect Dis. 2013;19:460–3.
18. FMOH. Guidelines for clinical and programmatic management of TB, TB/HIV and leprosy in Ethiopia. Addis Ababa: FMOH; 2013.

19. 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.

20. Hermans PWM, Messadi F, Guebrexabher H, Van Soolingen D, De Haas PEW, Heersma H, et al. Analysis of the population structure of *Mycobacterium tuberculosis* in Ethiopia, Tunisia, and the Netherlands: usefulness of DNA typing for global tuberculosis epidemiology. J Infect Dis. 1995;171:1504–13.

21. 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;26:2465–6.

22. Jagielski T, Minias A, Van Ingen J, Rastogi N, Brzostek A, Zaczek A, et al. Methodological and clinical aspects of the molecular epidemiology of *Mycobacterium tuberculosis* and other mycobacteria. Clin Microbiol Rev. 2016;29:239–90.

23. Kamerbeek J, Schouls L, Kolk A, Van Agterveld M, Van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol. 1997;35:907–14.

24. Kebede AH, Alebachew Z, Tsegaye F, Lemma E, Abebe A, Agonafir M, et al. The first population-based national tuberculosis prevalence survey in Ethiopia, 2010-2011. Int J Tuberc Lung Dis. 2014;18:635–9.

25. Mbugi EV, Katale BZ, Siame KK, Keyyu JD, Kendall SL, Dockrell HM, et al. Genetic diversity of *Mycobacterium tuberculosis* isolated from tuberculosis patients in the Serengeti ecosystem in Tanzania. Tuberculosis. 2015;95:170–8.

26. Mihret A, Bekele Y, Loxton AG, Jordan AM, Yamuah L, Aseffa A, et al. Diversity of *Mycobacterium tuberculosis* isolates from new pulmonary tuberculosis cases in Addis Ababa, Ethiopia. Tuberc Res Treat. 2012;2012: 892079.

27. Mokrousov I. Revisiting the Hunter Gaston discriminatory index: note of caution and courses of change. Tuberculosis. 2017;104:20–3.

28. Nebenzahl-Guimaraes H, Yimer SA, Holm-Hansen C, De Beer J, Brosch R, Van Soolingen D. Genomic characterization of *Mycobacterium tuberculosis* lineage 7 and a proposed name: 'Aethiops vetus'. Microbial Genomics. 2016; 2:e000063.

29. Ogaro TD, Githui W, Kikuvi G, Okari J, Asiko V, Wangui E, et al. Diversity of *Mycobacterium tuberculosis* strains in Nairobi, Kenya. Afr J Health Sci. 2012;20:9.

30. Pan XL, Zhang CL, Nakajima C, Fu J, Shao CX, Zhao LN, et al. A quantitative and efficient approach to select MIRU-VNTR loci based on accumulation of the percentage differences of strains for discriminating divergent *Mycobacterium tuberculosis* sublineages. Emerg Microbes Infect. 2017;e68:6.

31. Pareek M, Evans J, Innes J, Smith G, Hingley-Wilson S, Lougheed KE, et al. Ethnicity and mycobacterial lineage as determinants of tuberculosis disease phenotype. Thorax. 2013;68:221–9.

32. Reed MB, Pichler VK, Mcintosh F, Mattia A, Fallow A, Masala S, et al. Major *Mycobacterium tuberculosis* lineages associate with patient country of origin. J Clin Microbiol. 2009;47:1119–28.

33. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N Engl J Med. 1994;330:1703–9.

34. Sola C, Filliol I, Legrand E, Lesjean S, Locht C, Supply P, et al. Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. Infect Genet Evol. 2003;3:125–33.

35. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rusch-Gerdes S, Willery E, et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol. 2006;44:4498–510.

36. Tadesse M, Abebe G, Bekele A, Bezabih M, De Rijk P, Meehan CJ, 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;55:251–9.

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

38. Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D. MIRU-VNTRplus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. Nucleic Acids Res. 2010;38:W326–31.

39. WHO. Global tuberculosis report; 2019. https://www.who.int/tb/publications/global_report/en/, 2019. Accessed 6 Apr 2020.

40. Wilton S, Cousins D. Detection and identification of multiple mycobacterial pathogens by DNA amplification in a single tube. PCR Methods Appl. 1992; 1:269–73.

41. Wondale B, Medihn G, Teklu T, Mersha W, Tamirat M, Ameni G. A retrospective study on tuberculosis treatment outcomes at Jinka general hospital, southern Ethiopia. BMC Res Notes. 2017;10:680.

42. Yimer SA, Birhanu AG, Kalayou S, Riaz T, Zegeye ED, Beyene GT, et al. Comparative proteomic analysis of *Mycobacterium tuberculosis* lineage 7 and lineage 4 strains reveals differentially abundant proteins linked to slow growth and virulence. Front Microbiol. 2017;8:795.

43. Yimer SA, Norheim G, Namouchi A, Zegeye ED, Kinander W, Tonjum 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;53:1301–9.

44. Zhou Y, Van Den Hof S, Wang S, Pang Y, Zhao B, Xia H, et al. Association between genotype and drug resistance profiles of *Mycobacterium tuberculosis* strains circulating in China in a national drug resistance survey. PLoS One. 2017;12:e0174197.

45. Zozio T, Allix C, Gunal S, Saribas Z, Alp A, Durmaz R, et al. Genotyping of *Mycobacterium tuberculosis* clinical isolates in two cities of Turkey: description of a new family of genotypes that is phylogeographically specific for Asia minor. BMC Microbiol. 2005;5:44.

46. Meehan CJ, Moris P, Kohl TA, Pečerska J, Akter S, Merker M, et al. The relationship between transmission time and clustering methods in Mycobacterium tuberculosis epidemiology. EBioMedicine. 2018;37:410–6.

47. Tilahun M, Ameni G, Desta K, Zewude A, Yamuah L, Abebe M, Aseffa A. 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(2):e0193083.

48. Bruchfeld J, Aderaye G, Palme IB, Bjorvatn B, Britton S, Feleke Y, et al. Evaluation of Outpatients with Suspected Pulmonary Tuberculosis in a High HIV Prevalence Setting in Ethiopia: Clinical, Diagnostic and Epidemiological Characteristics. Scand J Infect Dis. 2002;34:331–7.

Publisher's Note