



MOLECULAR ASPECTS

Molecular characterization of *Mycobacterium tuberculosis* isolates from Tanga, Tanzania: First insight of MIRU-VNTR and microarray-based spoligotyping in a high burden country



Abubakar S. Hoza ^{a, b, *}, Sayoki G. Mfinanga ^c, Irmgard Moser ^d, Brigitte König ^a

^a Department of Medical Microbiology and Epidemiology of Infectious Diseases, Medical Faculty, University of Leipzig, Liebig Str. 21, 04103 Leipzig, Germany

^b Department of Veterinary Microbiology and Parasitology, Sokoine University of Agriculture, P.O. Box 3019, Morogoro, Tanzania

^c National Institute for Medical Research, Muhimbili Medical Research Centre, P.O. Box 3436 Dar es Salaam, Tanzania

^d Friedrich Loeffler Institut, Institute of Molecular Pathogenesis, Naumburger Str. 96a, 07743 Jena, Germany

ARTICLE INFO

Article history:

Received 22 June 2015

Received in revised form

12 February 2016

Accepted 13 February 2016

Keywords:

Microarray-based spoligotyping

MIRU-VNTR

Mycobacterium tuberculosis

Primary health care facilities

Tanga-Tanzania

SUMMARY

Molecular typing of *Mycobacterium tuberculosis* (MTB) has greatly enhanced the understanding of the population structure of MTB isolates and epidemiology of tuberculosis (TB). To characterize prevalent genotypes of MTB, microarrays-based spoligotyping and mycobacterial interspersed repetitive unit-variable number of tandem repeats (MIRU-VNTR) were applied on 80 isolates collected from primary health care facilities in Tanga, North-eastern Tanzania. A total of 18 distinct spoligotypes were identified. The lineages by order of their predominance were EAI and CAS families (26.25%, 21 isolates each), LAM family and T super-family (10%, 8 isolates each), MANU family (3.75%, 3 isolates), Beijing family (2.5%, 2 isolates) and S family (1.25%, 1 isolate). Overall, sixteen (20%) strains could not be allocated to any lineage according to the SITVIT_WEB database. The allelic diversity (h) for specific MIRU-VNTR loci showed a considerable variation ranging from 0.826 of VNTR locus 3192 to 0.141 of VNTR locus 2059. The allelic diversity for 11 loci (VNTR 3192, 2996, 2165, 960, 4052, 424, 4156, 2531, 1644, 802 and 3690) exceeded 0.6, indicating highly discriminatory power. Seven loci (VNTR 2163b, 2401, 1955, 577, 4348, 2687 and 580) showed moderate discrimination ($0.3 \leq h \leq 0.6$), and three loci (VNTR3007, 154 and 2059) were less polymorphic. The present study suggests that the TB cases in Tanga might be caused by a diverse array of MTB strain families that may be indicative of a cosmopolitan population with frequent migration and travel. Microarray-based spoligotyping and MIRU-VNTR could be reliable tools in detecting different MTB genotypes in high burden settings.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Tuberculosis (TB) caused by bacteria belonging to the *Mycobacterium tuberculosis* complex (MTC), in humans primarily by *M. tuberculosis* (MTB), remains a major public health threat globally and a leading cause of morbidity and mortality in many sub-Saharan African countries. Co-infection with HIV/AIDS and the emergence of multi-drug resistant (MDR) MTB strains have worsened the situation [1]. Tanzania is among 22 countries with the highest burden of disease [2]. In 2013, the prevalence of bacteriologically confirmed

TB in Tanzania was 295 per 100,000 adult persons [3]. Notification of TB cases, has increased from about 61,838 cases in 2011 to 63,892 in 2012 which is about 3.3% increase [4]. This increase may be attributable to either increased TB/HIV co-infection [5] or reactivation of latent MTB infection or due to better diagnostic tools.

Molecular genotyping of MTC isolates has proven to be a valuable tool with great potential to significantly impact both individual clinical management and public health [6]. Techniques such as IS6110 restriction fragment length polymorphism (RFLP) typing [7], spoligotyping and mycobacterial interspersed repetitive unit - variable number tandem repeat (MIRU-VNTR) analysis [8,9], and more recently whole genome sequencing [10] have become powerful tools in understanding and predicting disease transmission dynamics. IS6110-RFLP is the gold standard among the MTB genotyping tools; however, the method is laborious and requires

* Corresponding author. Department of Veterinary Microbiology and Parasitology, Sokoine University of Agriculture, P.O. Box 3019, Morogoro, Tanzania. Tel.: +255 232603511/4; fax: +255 232604647.

E-mail address: abshoza@gmail.com (A.S. Hoza).

weeks for culturing the isolates since it demands large amounts of DNA and has poor inter-laboratory reproducibility hence rendering its use limited [11]. In addition, it is also less discriminatory among the MTB isolates having a low copy number (<5) of IS6110 [12].

PCR-based methods such as spoligotyping [13] and MIRU-VNTR typing [9] have enabled to overcome the drawbacks shown by the IS6110 fingerprinting. The methods are faster to perform and interpret, demand small quantities of genomic DNA and can easily be presented in digitalized format allowing for comparison between different laboratories [14]. Furthermore, availability of freely accessible web-based database for analysing data generated from these methods makes strains identification rather easier and quicker [1,15].

Spoligotyping is based on polymorphism among 43 spacers in the direct repeat (DR) locus among species of the MTC [13], whereas MIRU-VNTR is based on the number of repetitive units present in multiple defined loci [16]. The discriminatory power of this method is close to that of IS6110-based fingerprinting [12,17] and produces better resolution between strains with low IS6110 copy number [18]. Availability of an online tool for classification and analysis of strains of MTC http://tbinsight.cs.rpi.edu/run_tb_lineage.html [15] plus an international database at <http://www.pasteur-guade-loupe.fr/tb/spolddb4.htm> maintained at Pasteur Guadeloupe, allow comparison of spoligotypes of MTB [19].

Microarray-based spoligotyping offers a powerful high-throughput molecular typing alternative that is suitable for studying strain diversity in high burden populations and geographical areas to reveal epidemiological trends [20].

Only a limited number of studies have been conducted in Tanzania using modern molecular DNA fingerprinting techniques that are capable of directly tracing TB transmission routes. Information on important aspects of TB epidemiology, such as role of recent transmission in reactivation of latent cases is still lacking. Available data on the different spoligotype families of MTB strains in Tanzania are limited, and where available restricted to small geographical areas [21].

The present study aimed at characterizing the causative agent from TB patients in Tanga, Tanzania, by using microarray-based spoligotyping together with conventional (i.e. agarose gel-based) MIRU-VNTR typing. This is the first study in which a recently developed microarray-based spoligotyping and MIRU-VNTR are employed in a high burden country. Findings of this study provide informative epidemiological data urgently needed for improved TB control programmes and for prediction of future epidemiological trends.

2. Materials and methods

2.1. Study population

Eighty MTB isolates collected from 372 new and recurrent TB patients during a cross-sectional study conducted in Tanga, Tanzania from November 2012 through January 2013 were eligible for this study. All patients with clinical signs and symptoms suggestive of TB self-referred to four primary health care facilities were recruited in the study. The facilities included Bombo Regional Referral Hospital, Makorora and Ngamiani health centres in Tanga and Muheza Designated District Hospital (MDDH) in Muheza. A Map of Tanga indicating the number of isolates from each site is provided as [Supplementary file S1](#). Demographic information and data were collected only after provision of informed consent.

2.2. Specimen and data collection

Structured questionnaires were administered to the patients attending four TB clinics who provided informed consent to the

study. Two sputum samples (one spot during the initial visit to the clinic and one early morning) were collected into small autoclavable wide mouth glass bottles. The specimens were examined by direct smear microscopy at the respective clinics using either Ziehl Neelsen (ZN) -stain (at Makorora and Ngamian health centres) or fluorescence stain (at Bombo referral hospital and MDHH). All morning sputum samples were then shipped to the mycobacteriology laboratory at the University Hospital Leipzig, Germany, for culture and molecular analysis.

2.3. Specimen processing and culture

Sputum specimens were digested and decontaminated using N-acetyl-L-cysteine-sodium hydroxide method [22] and were re-examined for the presence of acid-fast bacilli (AFB) by fluorescence stain in Leipzig. The isolates were cultured in BacT/Alert 3D liquid culture system (bioMérieux, Marcy-l'Étoile, France) and on Löwenstein-Jensen and Gottsacker slants (Artelt-ENCLIT GmbH, Wyhra, Germany). Gottsacker slopes contain sodium pyruvate for improved isolation yield of *Mycobacterium bovis*. Cultures were incubated at 37 °C for up to 8 weeks and confirmed as MTB by using line probe assay GenoType[®] MTBC; (Hain Lifescience, Nehren, Germany). Isolates included those from smear and culture positive patients (n = 54), as well as those from smear-negative but culture positive patients (n = 26).

2.4. Extraction of genomic DNA from MTB isolates

Mycobacterial cells from positive BacT/Alert bottles and from LJ and Gottsacker slopes were used for DNA extraction. Bacterial DNA was extracted from heat-inactivated AFB isolates. Briefly, a loopful of colony material was placed into a labelled screw-capped eppendorf tube containing 500 µl sterile distilled water or by taking 500 µl from a positive BacT/Alert bottle. Each specimen was incubated on a heat block at 95 °C for 20 min to inactivate the bacteria. Heat-killed cells were stored at 4 °C until analysis. DNA extraction was done by using a High Pure PCR Template Preparation Kit (Roche Diagnostics GmbH, Mannheim, Germany) according to manufacturer's instruction. Genomic DNA of the H37Rv strain and sterile distilled water were used as positive and negative controls, respectively, for all genotyping procedures.

2.5. Microarray-based spoligotyping

Spoligotyping of the MTB isolates was performed by using the newly developed microarray-based spoligotyping format of the Array Strip platform (Alere Technologies GmbH, Jena, Germany) [20]. Briefly, genomic DNA of MTB isolates was amplified using polymerase chain reaction (PCR) with the primers DRa (5'-biotin labelled) and DRb. The PCR products were then hybridized on Array Strips using hybridization kit (Alere) at 60 °C for 1 h and washing steps at 55 °C, otherwise following manufacturer's instructions. Recording of stained microarrays was done by using an Array Mate transmission reader (Alere). Signal intensities equal and higher than 0.3 (on a scale from 0 to 1.0) were considered positive for the respective probe [20]. The binary code data were automatically compared with SpolDB4.0 [19] and the updated version SITVITWEB [23] databases to identify concordant species and lineages.

2.6. MIRU-VNTR typing

To identify a suitable MIRU-VNTR locus panel for genotyping MTB isolates in this geographical area, 12 and 15 loci were chosen for analysing the 80 MTB isolates. MIRU-VNTR typing was

Table 1
Distribution of MTBC isolate by site.

Site	Smear results		Case history
	smear+/smear-		New/recurrent
Makorora HC	4/1		5/0
Ngamian HC	15/11		20/6
Bombo RH	8/6		14/0
MDDH	27/8		33/2
Total	54/26		72/8

HC; Health centre, RH, referral hospital; MDDH, Muheza Designated District hospital.

performed by PCR amplification of the selected MIRU loci using primers as described in the MIRU-VNTR standard protocol [16,24,25]. PCR mixture was prepared using the HotStar Taq DNA polymerase kit (Qiagen, Hilden, Germany). PCR products were analysed on 1.5% agarose gel against 100bp DNA ladder (PEQ lab Biotechnologie GmbH, Erlangen, Germany) in TAE buffer electrophoresis for 2 h at 100 constant voltage. The allele calling table provided in the MIRU-VNTR standard protocol [16] was applied to assign the number of alleles that correspond to the amplicon's size.

Results were entered into Ms Excel sheet in a digital format. MIRU-VNTRplus database (<http://www.miru-vntrplus.org/>) was used to identify the MTB strains by similarity search and phylogenetic tree, using a categorical coefficient of one and a distance cut-off of zero. The UPGMA dendrogram was constructed from the strains genotypes using the online MIRU-VNTRplus database [1]. The MIRU-VNTR allelic diversity (h) at a particular locus was calculated as follows: $h = 1 - \sum x_i^2 [n/n-1]$, where x_i is the

frequency of the i th allele at the locus and n is the number of isolates [26]. To determine the discriminatory power of the MIRU-VNTR for this patient population, the Hunter–Gaston discriminatory index (HGDI) was used [27,28]. The HGDI was calculated using the following formula:

$$HGDI = 1 - \frac{1}{N(N-1)} \sum_{j=1}^s nj(nj-1)$$

where, D is the numerical index of discrimination, N the total number of strains in the typing scheme, s is the total number of different strain types, and n_j is the number of strains belonging to the j th type.

2.7. Major lineages prediction

An online tool for classification and analysis of strains of MTC (http://tbinsight.cs.rpi.edu/run_tb_lineage.html) was used to determine the major lineages based on Conformal Bayesian Network (CBN) [15,29]. To determine the major lineages in our patient population, we used a combination of spoligotype and 12-loci of MIRU as a feature in the CBN method, since the complete data set for the 24 loci of MIRU was not available.

2.8. Ethical consideration

The study was reviewed and approved by the national ethical review committee with secretariat at the National Institute for Medical Research (NIMR), Dar es Salaam, Tanzania and provided with a certificate number NIMR/HQ/R.8a/Vol.IX/1401. All subjects

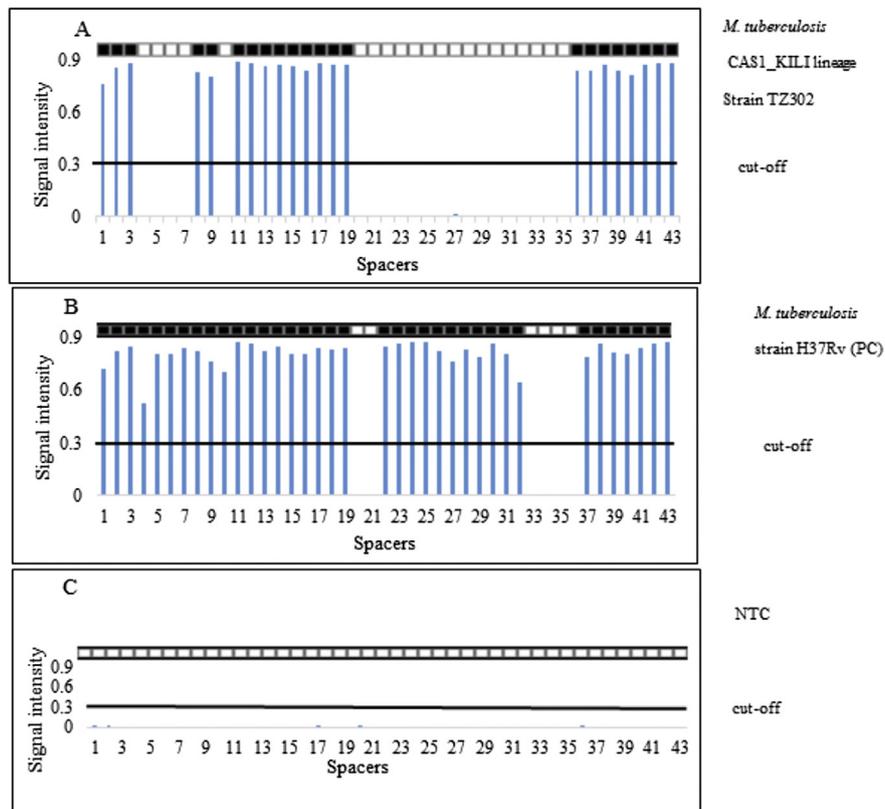


Figure 1. Illustration of Array Strip spoligotyping. A. Presentation of experimental output from Array Strip spoligotyping of *Mycobacterium tuberculosis* CAS1-KILI patient strain. B. Positive control *Mycobacterium tuberculosis* H37Rv strain. C. A nontemplate control (NTC). The Array Strip platform utilizes 4-by4-mm microarrays mounted on the bottom of reaction vessels that are arranged in strips of 8 and fit into the 96-well microtiter plate format.

Table 2
Most prevalent spoligotype families in Tanga according to the SITVIT_WEB database and their corresponding Major lineages predicted by CBN method.

SIT [*]	SITVIT_WEB [†]	CBN [‡]	N(%) [§]	Spoligotype pattern
8	EAI5	Indo-Oceanic	18 (22.5)	
21	CAS1_Kili	East-African Indian	16 (20.0)	
46	ND	Euro-American	6 (7.5)	
Orphan	ND	Euro-American	6 (7.5)	
53	T1	Euro-American	5 (6.25)	
4	ND	Euro-American	4 (5.0)	

ND, not defined.
^{*} SIT, spoligo-international-type from SITVIT_WEB database.
[†] Spoligotype families as assigned in the SITVIT_WEB database.
[‡] Major lineage predicted by CBN.
[§] Number of isolates with common SIT.

provided informed consent for the collection and study of their isolates.

3. Results

3.1. Description of the isolates

A total of 80 MTB isolates from 56 male and 24 female TB patients with a range 10–90 years (median age of 33) from four TB clinics were included in the study. Seventy-two of the patients were new cases, eight were recurrent cases; fifty-four of the subjects were smear-positive cases by direct smear microscopy, while 26 were smear-negative (Table 1). All 80 isolates were genotyped by microarray-based-spoligotyping and MIRU-VNTR typing.

3.2. Interpretation of the array results

The test results of the respective strains were plotted on a graph using signal intensity (i.e. presence or absence of a particular spacer) of each spacer for each particular spoligotype with cut-off point value of 0.3. The experimental output of the microarray-based spoligotyping of MTB for the representative predominant spoligotype lineage (CAS1_KILI) in the study population as well as the positive and negative controls are shown exemplarily in

Figure 1. We also compared the results from genotyping studies conducted previously in Tanzania using conventional membrane-based spoligotyping (Table 4).

3.3. Microarray-based spoligotyping results

Among the 80 isolates, the majority of strains belonged to the EAI and CAS lineages (26.25%, 21 isolates) each, LAM and T lineages comprised eight (10% each) strains, three (3.75%) were assigned to MANU lineage, two (2.5%) strains belonged to Beijing genotype and one (1.25%) strain was assigned to S-type. Sixteen (20%) isolates were not assigned to any lineage using the SITVIT_WEB database and were named as “not defined” in this study. The predominant spoligotype families in Tanga based on SITVIT_WEB database, their respective major lineages as predicted by CBN method are shown in Table 2, and information of all the 80 isolates is shown in Supplementary Table S2.

3.4. Major lineages prediction by CBN method

Among the 80 isolates predicted by the CBN method, majority of the strains belonged to the modern lineages, with Euro-American 33 (41.25%) being the most prevalent TB lineage, followed by East-African Indian 21 (26.25%) and East-Asian (Beijing) 2 (2.5%)

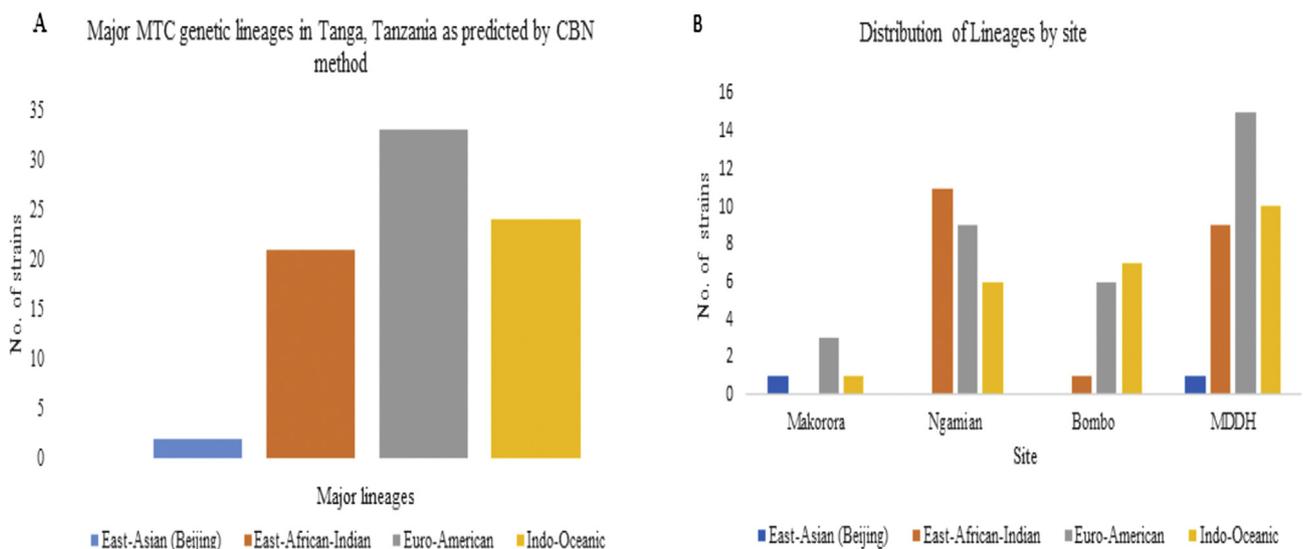


Figure 2. Major MTC lineages in Tanga, Tanzania as determined by CBN method. A shows predominant lineages in this geographical area. B shows lineages distribution by site.

Table 3
Proportion of first-line anti-TB drug resistance among different genotypes in Tanga.

SIT ^a	SITVIT_WEB ^b	No. [†]	% Resistance to:					
			INH	RMP	SM	EMB	PZA	MDR [‡]
46	ND	6	33.3	33.3	33.3	16.7	33.3	33.3
4	ND	4	50	50	50		50	50
42	LAM9	3	33.3	33.3	33.3		33.3	33.3
356	CAS	2	50	50		50		50

INH, isoniazid; RMP, rifampicin, SM, Streptomycin; EMB, Ethambutol; PZA, Pyrazinamide.

ND, not defined.

^a SIT from SITVIT_WEB database.

[†] Representing families annotated in SITVIT_WEB database.

[‡] Number of isolates.

[§] MDR, multidrug resistant, represent isolates resistant to at least isoniazid and rifampicin.

being the least prevalent. On the other hand 24 (30%) belonged to the ancestral lineages of the Indo-Oceanic lineage. Results of major lineages prediction and distribution by site are shown in Figure 2.

3.5. Drug resistance profile of the spoligotype families

All isolates were tested for their resistance to rifampicin (RMP), isoniazid (INH), streptomycin (SM), ethambutol (EMB) and pyrazinamide (PZA) by a proportion method using BaCT/ALERT 3D system [biomerieux, Marcy-l'Étoile, France]. Critical concentrations of 1 µg/ml for RMP, INH, and SM; 2 µg/ml for EMB and 200 µg/ml for PZA were respectively used as described before [30]. Results showed that the majority of the genotype families found in this population were susceptible to first-line anti-TB drugs. One of the three isolates assigned to LAM9 genotype was found to be a MDR case, and in addition was resistant to SM and PZA. Another resistant isolate was found in one of the two isolates designated to CAS genotype, which was both MDR case and EMB resistant. Other resistant strains were found among the isolates “not defined” in the SITVIT_WEB, with four isolates being MDR cases. The proportions of first-line anti-TB drugs among different genotypes are shown in Table 3.

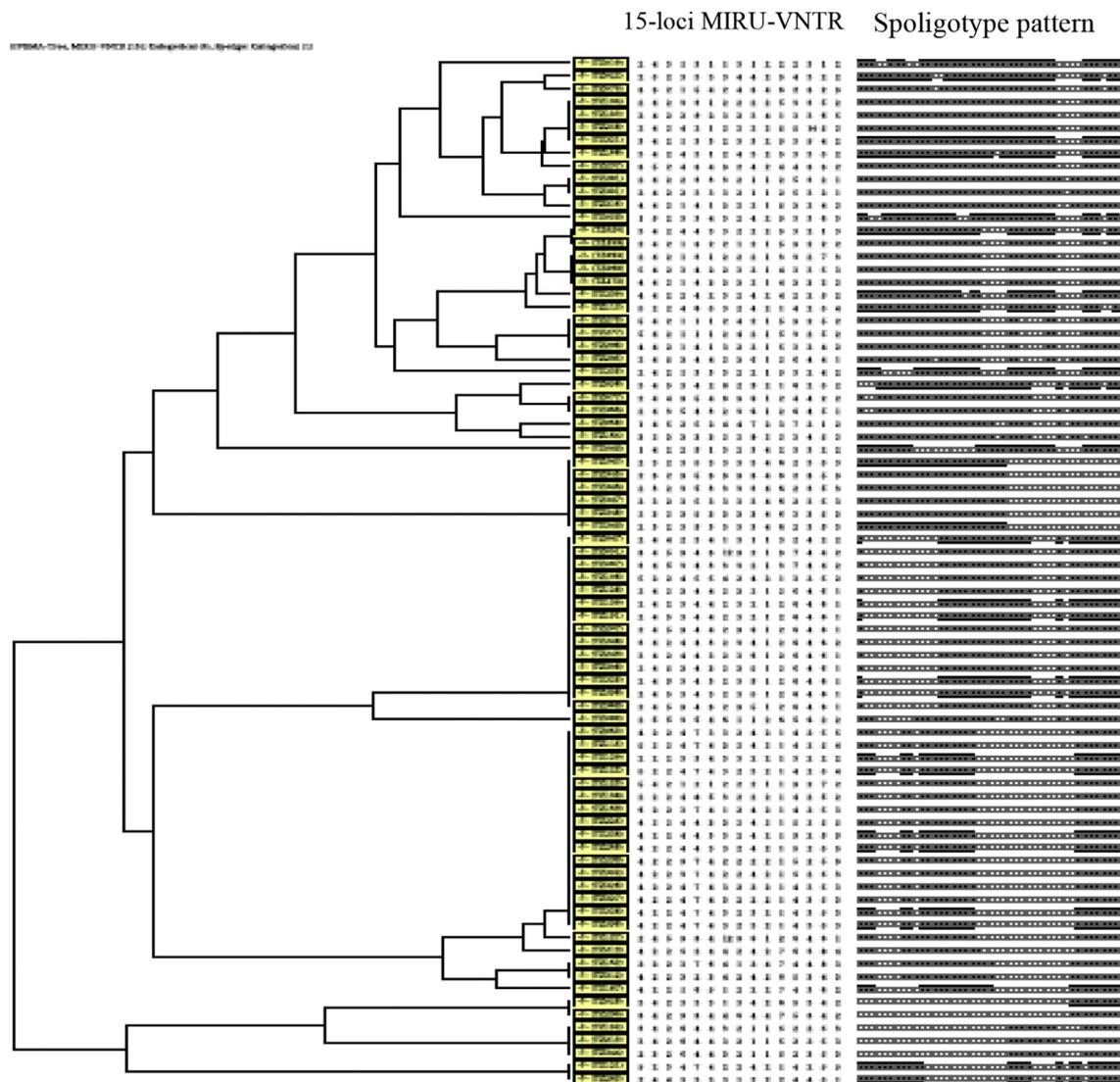


Figure 3. This dendrogram was generated by using unweighted pair group method with arithmetic averages (UPGMA) algorithm in the MIRU-VNTRplus database (<http://www.miru-vntrplus.org/>). The tree was built on 80 *M. tuberculosis* clinical isolates from Tanga based on 15 loci MIRU-VNTR genotypes and Microarray-based spoligotyping. The order of MIRU loci is as follows, left to right: 424, 577, 580, 802, 960, 1644, 1955, 2163b, 2165, 2401, 2996, 3192, 3690, 4052 and 4156.

3.6. MIRU-VNTR typing

In order to evaluate and determine the most suitable loci for genotyping the *M. tuberculosis* isolates in Tanga, we analysed 12 and 15 loci panels of MIRU-VNTR. The allelic diversity (h) calculated of 80 MTB isolates for specific MIRU-VNTR loci showed a considerable variation ranging from 0.826 of VNTR locus 3192 to 0.141 of VNTR locus 2059. The allelic diversity for 11 loci (VNTR 3192, 2996, 2165, 960, 4052, 424, 4156, 2531, 1644, 802 and 3690) exceeded 0.6, indicating that they are highly discriminating. Seven loci (VNTR 2163b, 2401, 1955, 577, 4348, 2687 and 580) showed moderate discrimination ($0.3 \leq h \leq 0.6$), three loci (VNTR3007, 154 and 2059) were even less polymorphic (Table 5).

Due to low discriminatory power of the 12-loci MIRU-VNTR set seen for this patient population, the 15 loci MIRU-VNTR set with the overall HGDI of 0.9889 and 21.3% clustering rate was deemed suitable for genotyping MTB isolates in this patient population.

VNTR locus 3192 was the most discriminatory locus with seven alleles; one to seven repeats of this locus were observed with allele number three being the most common found in 19 isolates. VNTR locus 2996 exhibited in eight alleles was the second most discriminatory locus with allele number one being the most common found in 21 isolates followed by locus 2165 exhibited in eight alleles, with allele two being the most common found in 24 isolates. VNTR locus 960 exhibited in nine alleles, with allele four being most common, observed in 28 isolates. The least variable locus was VNTR 580 exhibited in two to five alleles, with allele two being the most common found in 60 isolates. Other loci with their respective discriminatory indices are shown in Table 5.

Of the 80 *M. tuberculosis* strains genotyped, a total of 58 different VNTR genotypes were identified. Of these, 46 (79.3%) were unique (i.e. only for a single strain) and 34 strains were clustered into 12 (20.7%) clusters of two to six representatives. The largest cluster observed comprised of six isolates all from the same location whereas other clusters comprised between two to four strains. The UPGMA dendrogram was constructed using 15 loci MIRU-VNTR (Figure 3).

4. Discussion

In the present study, we report an insight into the MTB strains isolated in Tanga, Tanzania by combining microarray-based spoligotyping and MIRU-VNTR typing method. To the best of our knowledge, this is the first study on the genetic diversity of MTB in a high burden setting where the two approaches have been employed.

The results obtained reveal that the population structure of MTB isolates in Tanga appears to be heterogeneous, as 18 known spoligotype families based on the SITVIT_WEB database were obtained from 80 MTB isolates as depicted in Supplementary Table S2. Similar families have been reported in variable proportions in other countries bordering Tanzania [31–35], indicating a geographical widespread of this group of spoligotypes in the region. These findings are also in agreement with those of other studies conducted in Tanzania which reported the predominance of these families [21,36,37].

The predominance of the CAS1_KILI in this area and other parts of the country as previously reported [36] may be enhanced by virulence, transmissibility and/or specific adaptation to a host population. Increased reporting of CAS1_KILI strain in Tanzania is, however, not clear, whether it is a result of chance of this strain to spread following introduction or is due to its phylogeographical specificity for Tanzania. Six (37.5%) of the “not defined” isolates in the SITVIT_WEB had unique patterns and were predominant in Muheza District and was the most clustered family suggesting a recent transmission.

Strain families previously reported in Tanzania but not reported in this study include the Haarlem genotype [21,37] and the X genotypes [21,36–38]. We detected a high proportion of T family isolates as found in previous studies [21,37]. Furthermore, we report sub-lineage EAI6_BGD1 which was reported in a recent study [21] and sub-lineage EAI1_SOM [21,37]. We also report a relative high proportion of MANU family as compared to those reported previously [36,37] (Table 4).

The presence of MANU family in this geographical area suggests the predominance of this family in northern part of the country, as Tanga lies in the north-eastern corner of Tanzania and the previous MANU families were detected in Kilimanjaro and Serengeti ecosystem in northern Tanzania as well [36,37]. The diverse distribution of strains observed in this and other studies in Tanzania so far underscore the need to map the distribution of MTC genotypes in the country.

Generally, the results of major lineages prediction by CBN method suggest that 70% of TB infections in this geographical area are due to the modern lineages, whereas 30% of TB infections is due to the ancestral lineages mainly of Indo-oceanic lineage. This suggests that modern *M. tuberculosis* strains contribute significantly to TB infections in Tanga. Modern *M. tuberculosis* strains represent epidemic strains that were introduced into the same geographical regions more recently as a consequence of the world-wide spread of the tuberculosis epidemic [19]. Over the past decades, Tanzania has experienced a tremendous improvement in infrastructure development and expanding local and international trade, which

Table 4
Comparison of prevalence of MTC families based on microarray-based spoligotyping (this study) and in other studies in Tanzania.

Family	(Eldholm et al., 2006) [*]	(Kibiki et al., 2007) [*]	(Mfinanga et al., 2014) [*]	(Mbugi et al., 2014) [†]	Present study [‡]
EAI	25 (17.0)	13 (10.0)	49 (10.1)	25 (11.7)	21 (26.3)
CAS	52 (35.4)	49 (37.7)	195 (40.0)	55 (25.7)	21 (26.3)
LAM	33 (22.4)	30 (23.2)	84 (17.5)	38 (17.8)	8 (10.0)
T	21 (14.3)	16 (12.3)	58 (11.9)	52 (24.3)	8 (10.0)
Beijing	7 (4.8)	7 (5.4)	33 (6.8)	8 (3.7)	2 (2.5)
S	3 (2.0)	0	4 (0.8)	1 (0.5)	1 (1.3)
MANU	0	3 (2.3)	0	1 (0.5)	3 (3.7)
X	1 (0.7)	0	0	1 (0.5)	0
Haarlem	0	0	10 (2.1)	6 (2.8)	0
Not defined	5 (3.4)	12 (9.2)	52 (10.7)	25 (11.7)	16 (20.0)
Total	147	130	487	214	80

^{*} Membrane-based spoligotyping.

[†] Membrane-based spoligotyping and 24 loci typing MIRU-VNTR.

[‡] Microarray-based spoligotyping and 15 loci MIRU-VNTR typing (this study).

Table 5
Allelic diversity of 80 MTC strains in Tanga using the 12 and 15 loci panels of MIRU-VNTR.

VNTR locus	Alias	No. of isolates in respective copy											HGDI			
		1	2	3	4	5	6	7	8	9	10	11				
15-loci MIRU panel	424	Mtub04	3	34	16	17	9	1								0.729
	577	ETRC		22	7	47	4									0.576
	580	MIRU04		60	4	3	13									0.412
	802	MIRU40	13	5	35	22	2	3								0.709
	960	MIRU10	2	4	18	28	13	2	11	1	1					0.787
	1644	MIRU16	11	6	30	28	4	1								0.719
	1955	Mtub21	4	19	46	7		3		1						0.609
	2163b	QUB11b	4	40	29	5		2								0.619
	2165	ETRA	5	24	16	19	3	9	1		3					0.804
	2401	Mtub30	32	37		11										0.615
	2996	MIRU26	21	20	4	4	17	9	4	1						0.813
	3192	MIRU31	2	11	19	17	13	15	3							0.826
	3690	Mtub39		29	28	20	1	1				1				0.692
	4052	QUB26		10		10	36	14	5	5						0.737
	4156	QUB4156	12	32	24	5	7									0.725
12-loci MIRU panel	154	MIRU02	6	73	1											0.164
	580	MIRU04		60	4	3	13									0.412
	960	MIRU10	2	4	18	28	13	2	11	1	1					0.787
	1644	MIRU16	11	6	30	28	4	1								0.719
	2059	MIRU20	6	74												0.141
	2531	MIRU23					27	31	8	5		5	4			0.725
	2687	MIRU24	56	24												0.453
	2996	MIRU26	21	20	4	4	17	9	4	1						0.813
	3007	MIRU27	2	6	68	4										0.272
	3192	MIRU31	2	11	19	17	13	15	3							0.826
4348	MIRU39	7	27	46											0.555	
802	MIRU40	13	5	35	22	2	3								0.709	

HGDI; Hunter-Gaston discriminatory index

permit movement of people from within and outside the country. Consequently, this might have played an important role in the introduction of new *M. tuberculosis* strains in this region.

Spoligotyping, be it microarray-based or membrane-based, is a rapid diagnostic method most suitable for studying TB disease in high burden countries. However, an important component for technology transfer, especially for resource-poor settings like Tanzania, will be the cost/output ratio required to introduce a new technique (microarray-based high-throughput) compared to the established one (membrane-based low throughput). Microarray-based method has a potential of: (i) producing results within a single working day (high turnaround time), (ii) ease of operation and use, since no need of handling membrane in a dot blot manifold and developing a chemi-luminescence film in a darkroom, (ii) data are automatically processed using an online database, and (iv) relatively low cost (~5–10 euros per assay). Given its upper hand advantages, microarray-based method could be used as a relatively inexpensive “first screen” genotyping of MTC, and still remain an informative assay that could be suited to resource-poor countries. Moreover,

digital-numerical data output makes it less prone to interpretation errors, in addition allows easy interpretation of data (Figure 1).

MIRU-VNTR typing has been used in epidemiology studies, and has shown adequate stability in tracking recent transmission and distinguishing relapses and reinfections [9]. Different VNTR typing sets have been used. In the present study, we evaluated a 15 loci panel for its ability to discriminate the different MTB genotypes in Tanga. Many loci had high discriminatory index with VNTR locus 3192 being the most discriminatory, while only VNTR locus 580 showed low discriminatory index ($h = 0.412$) (Table 5).

The high genetic diversity (i.e. HGDI) of 0.989 in this study is relatively close to those observed in Kawempe, Uganda (0.996) [33] and in Ndola, Zambia (0.998) [32], which are similar settings in endemic countries bordering Tanzania, in the north and south, respectively. The diversity of MTB isolates in Tanga population may be related to reactivation of latent MTB infection, and/or may be due to increased human population movement. Consequently, this may predispose this population to possible introduction of potentially new different strains [39].

The standard MIRU 15 loci set has been recommended as the standard for routine molecular epidemiology of TB, including outbreak investigations and population-based transmission studies [40]. As shown in this study, the 15 loci panel offers a better discriminatory power in this patient population. Therefore, we recommend that this 15 loci panel should be used as a first-line panel for genotyping MTB isolates in this setting, especially for routine epidemiological investigation, as it may be more cost effective than the full set of 24 loci. It should also be noted that not all 24 loci are required for MTB genotyping in any given situation [16,41], as the number of loci required depends on the lineage known to be prevalent in the investigated area. Furthermore, based on the allelic diversity of individual MIRU-VNTR locus, different combinations of MIRU-VNTR loci offering high resolution and cost effective analyses for the different MTC lineages need to be sought [41].

However, it is important to note that the accuracy of the phylogenetic grouping by MIRU-VNTR is more exact than that of spoligotyping but depends on the number of loci included in the analysis and classification errors are reduced when analysing 24 loci [42]. Strains with ambiguous spoligotype signatures like the unknown family isolates in our study belonging to the Euro-American lineages may require a more discriminatory marker such as the 24 loci.

The fact, that 26/80 (32.5%) patients were smear-negative suggest that a substantial number of patients may be missed out if we continue relying on insensitive smear microscopy for TB screening (Table 1). Establishing culture facilities at least at regional levels may help detecting such patients. This underscores need to consider smear-negative patients when conducting future surveillance programmes in the country.

The findings of this study suggest that TB cases in Tanga might be caused by a diverse array of MTB strains an indication of a cosmopolitan population with frequent migration and travel. Furthermore, the dominant genotypes such as CAS1_KILI and EAI5 may have been present in this population for an extended period or there is an ongoing transmission of closely related strains. Although these strains are not representative of all strains in these areas, they provide preliminary insight into the strain diversity of MTB genotypes in Tanga. The emergence of new subtype strain families identified may well confirm the dynamic nature of this population.

In conclusion, this study provides a preliminary insight into the MTB genotypes circulating in Tanga and that microarray-based spoligotyping and MIRU-VNTR 15 loci panel could be reliable tools in detecting different genotypes in settings like Tanga. However, we propose a large sample size with a long recruitment period for the whole country, in order to provide a detailed population structure of MTB circulating in Tanzania, and an informative epidemiological data needed for improved TB control programmes. We also call for strengthening efforts on early case finding through enhanced public health education campaigns and provision of accessible diagnostic services with enhanced treatment compliance in the country.

Acknowledgement

This study was partly supported by Integrated Control of Neglected Zoonoses (ICONZ FP7-221948) project and Self-diagnostics (BGADL-0224) project, University of Leipzig. We are grateful to the technical support from Gesine Kauth and Uta Brommer of the Friedrich Loeffler Institute (FLI, Jena), Ms. Elizabeth Kraftschek of the Department of Medical Microbiology and Epidemiology of Infectious Diseases, University Hospital Leipzig for her cooperation. We thank Dr. Chola Shamputa for his advice on spoligotyping and MIRU-VNTR analysis, and Dr. Mathias Merker of

Molecular Mycobacteriology, Research Centre Borstel for his assistance on MIRU-VNTR analysis.

Funding: None.

Ethical approval: Not required.

Competing interests: The authors declare that there is no conflict of interests regarding the publication of this paper.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.tube.2016.02.002>.

References

- [1] Allix 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.
- [2] WHO. Global tuberculosis report. Geneva, Switzerland: WHO Press; 2014.
- [3] NTLP. The first national tuberculosis prevalence survey in the United Republic of Tanzania final report. 2013.
- [4] NTLP. National tuberculosis and leprosy programme annual report. Dar es Salaam, Tanzania: National TB and Leprosy Programme; 2011.
- [5] WHO. Global tuberculosis control Surveillance, Planning, financing. 2007. Geneva Switzerland.
- [6] Allix C, Walravens K, Saegerman C, Godfroid J, Supply P, Fauville-Dufaux M. Evaluation of the epidemiological relevance of variable-number tandem-repeat genotyping of *Mycobacterium bovis* and comparison of the method with IS6110 restriction fragment length polymorphism analysis and spoligotyping. *J Clin Microbiol* 2006;44:1951–62.
- [7] Van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM, Small PM. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993;31:406–9.
- [8] van Deutekom H, Supply P, de Haas PEW, Willery E, Hoijing SP, Locht C, Coutinho RA, van Soolingen D. Molecular typing of *Mycobacterium tuberculosis* by mycobacterial interspersed repetitive unit-variable-number tandem repeat analysis, a more accurate method for identifying epidemiological links between patients with tuberculosis. *J Clin Microbiol* 2005;43:4473–9.
- [9] Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Locht C. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J Clin Microbiol* 2001;39:3563–71.
- [10] Roetzer A, Diel R, Kohl TA, Rückert C, Nübel U, Blom J, Wirth T, Jaenicke S, Schuback S, Rüschergerdes S, Supply P, Kalinowski J, Niemann S. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS Med* 2013;10:e1001387.
- [11] Stavrum R, Mphahlele M, Ovreås K, Muthivhi T, Fourie PB, Weyer K, Grewal HMS. High diversity of *Mycobacterium tuberculosis* genotypes in South Africa and preponderance of mixed infections among ST53 isolates. *J Clin Microbiol* 2009;47:1848–56.
- [12] Barlow REL, Gascoyne-Binzi DM, Gillespie SH, Dickens A, Qamer S, Hawkey PM. Comparison of variable number tandem repeat and IS6110-restriction fragment length polymorphism analyses for discrimination of high- and low-copy-number IS6110 *Mycobacterium tuberculosis* isolates. *J Clin Microbiol* 2001;39:2453–7.
- [13] Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 1997;35:907–14.
- [14] Shamputa IC, Lee J, Allix-Béguec C, Cho E, Lee J, Rajan V, Lee EG, Min JH, Carroll MW, Goldfeder LC, Kim JH, Kang HS, Hwang S, Eum S-Y, Park SK, Lee H, Supply P, Cho S-N, Via LE, Barry CE. Genetic diversity of *Mycobacterium tuberculosis* isolates from a tertiary care tuberculosis hospital in South Korea. *J Clin Microbiol* 2010;48:387–94.
- [15] Shabbbeer A, Cowan LS, Ozcaglar C, Rastogi N, Vandenberg SL, Yener B, Bennett KP. TB-Lineage: an online tool for classification and analysis of strains of *Mycobacterium tuberculosis* complex. *Infect Genet Evol* 2012;12:789–97.
- [16] Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, Savine E, van Deutekom HHP, Roring S, Bifani P, Kurepina N, Kreiswirth B, Sola C, Rastogi N, Vatin V, Guttierrez MC, Fauville M, Niemann S, Skuce R, Kremer K, Locht C, van Soolingen D. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 2006;44:4498–510.
- [17] Hawkey PM, Smith EG, Evans JT, Monk P, Bryan G, Mohamed HH, Bardhan M, Pugh RN. Mycobacterial interspersed repetitive unit typing of *Mycobacterium*

- tuberculosis* compared to IS6110-based restriction fragment length polymorphism analysis for investigation of apparently clustered cases of tuberculosis. *J Clin Microbiol* 2003;41:3514–20.
- [18] Cowan LS, Mosher L, Diem L, Massey JP, Crawford JT. 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] Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajj SA, Allix C, Aristimuño L, Arora J, Baumanis V, Binder L, Cafrune P, Cataldi A, Cheong S, Diel R, Ellermeier C, Evans JT, Fauville Dufaux M, Ferdinand S, de Viedma GD, Garzelli C, Gazzola L, Gomes HM, Guttierrez MC, Hawkey PM, van Helden PD, Kadival GV, Kreiswirth BN, Kremer K, Kubin M, Kulkarni SP, Liens B, Lillebaek T, Ho ML, Martin C, Mokrousov I, Narvskaja O, Ngeow YF, Naumann L, Niemann S, Parwati I, Rahim Z, Rasoloforazanamparany V, Rasolonavalona T, Rossetti ML, Singh UB, Somoskovi A, Skuce RA, van Soolingen D, Streicher EM, Suffys PN, Tortoli E, Tracevska T, Vincent V, Victor TC, Warren RM, Yap SF, Zaman K, Portaels F, Rastogi N, Sola C. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol* 2006;6:23.
- [20] Ruettger A, Nieter J, Skrypnik A, Engelmann I, Ziegler A, Moser I, Monecke S, Ehrlich R, Sachse K. Rapid spoligotyping of *Mycobacterium tuberculosis* complex bacteria by use of a microarray system with automatic data processing and assignment. *J Clin Microbiol* 2012;50:2492–5.
- [21] Mfinanga SG, Warren RM, Kazwala R, Ngadaya E, Kahwa A, Kazimoto T, Kimaro G, Mfaume S, Timothy C, Egwaga S, Streicher EM, van Pittius, Mørkve O, Cleaveland S. Genetic profile of *Mycobacterium tuberculosis* and treatment outcomes in human pulmonary tuberculosis in Tanzania. *Tanzan J Hlth Res* 2014;16.
- [22] Kent PT, Kubica GP. Public health mycobacteriology: a guide for the level III laboratory. Atlanta, Ga: U.S. Dept. of Health and Human Services, Public Health Service, Centers for Disease Control; 1985.
- [23] Demay C, Liens B, Burguière T, Hill V, Couvin D, Millet J, Mokrousov I, Sola C, Zozio T, Rastogi N. SITVITWEB – a publicly available international multi-marker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect Genet Evol* 2012;12:755–66.
- [24] Supply P, editor. MIRU VNTR typing: the new international standard for TB molecular epidemiology; 2010.
- [25] Supply P. Multilocus variable number tandem repeat genotyping of *Mycobacterium tuberculosis*. 2005. Technical Guide.
- [26] Selander RK, Caugant DA, Ochman H, Musser JM, Gilmour MN, Whittam TS. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol* 1986;51:873–84.
- [27] 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.
- [28] Hunter PR. Reproducibility and indices of discriminatory power of microbial typing methods. *J Clin Microbiol* 1990;28:1903–5.
- [29] Aminian M, Shabbeer A, Bennett KP. A conformal Bayesian network for classification of *Mycobacterium tuberculosis* complex lineages. *BMC Bioinforma* 2010;11(Suppl. 3):S4.
- [30] Hoza AS, Mfinanga SG, König B. Anti-TB drug resistance in Tanga, Tanzania: a cross sectional facility-base prevalence among pulmonary TB patients. *Asian Pac J Trop Dis* 2015;8:907–13.
- [31] Asimwe B. Molecular characterization of *Mycobacterium tuberculosis* complex from Kampala, Uganda. Doctoral degree. Stockholm: Sweden & Kampala; 2008. Uganda.
- [32] Mulenga C, Shamputa IC, Mwakazanga D, Kapata N, Portaels F, Rigouts L. Diversity of *Mycobacterium tuberculosis* genotypes circulating in Ndola, Zambia. *BMC Infect Dis* 2010;10:177.
- [33] Nabyonga L, Kateete DP, Katabazi FA, Odong PR, Whalen CC, Dickman KR, Moses JL. Determination of circulating *Mycobacterium tuberculosis* strains and transmission patterns among pulmonary TB patients in Kawempe municipality, Uganda, using MIRU-VNTR. *BMC Res Notes* 2011;4:1–6.
- [34] Ogara TD, Githui W, Kikui G, Okari J, Asiko V, Wangui E, Jordaam AM, van Helden PD, Streicher ME, Victor TC. Diversity of *Mycobacterium tuberculosis* strains in Nairobi, Kenya. *Afr J Health Sci* 2012;20:82–90.
- [35] Asimwe BB, Joloba ML, Ghebremichael S, Koivula T, Kateete DP, Katabazi FA, Pennhag A, Petersson R, Kallenius G. DNA restriction fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolates from HIV-seropositive and HIV-seronegative patients in Kampala, Uganda. *BMC Infect Dis* 2009;9:1–9.
- [36] Kibiki GS, Mulder B, Dolmans Wil MV, de Beer Jessica L, Boeree M, Sam N, van Soolingen D, Sola C, van der Zanden AGM. *M. tuberculosis* genotypic diversity and drug susceptibility pattern in HIV-infected and non-HIV-infected patients in northern Tanzania. *BMC Microbiol* 2007;7:51.
- [37] Mbugi EV, Katala BZ, Siame KK, Keyyu JD, Kendall SL, Dockrell HM, Streicher EM, Michel AL, Rweyemamu MM, Warren RM, Matee MI, van Helden PD. Genetic diversity of *Mycobacterium tuberculosis* isolated from tuberculosis patients in the Serengeti ecosystem in Tanzania. *Tuberculosis* 2014;1–9.
- [38] Eldholm V, Matee M, Mfinanga SGM, Heun M, Dahle UR. A first insight into the genetic diversity of *Mycobacterium tuberculosis* in Dar es Salaam, Tanzania, assessed by spoligotyping. *BMC Microbiol* 2006;6:76.
- [39] Glynn JR, Alghamdi S, Mallard K, McNeerney R, Ndlovu R, Munthali L, Houben RM, Fine PEM, French N, Crampin AC. Changes in *Mycobacterium tuberculosis* genotype families over 20 years in a population-based study in Northern Malawi. *PLoS ONE* 2010;5:e12259.
- [40] Asante-Poku A, Nyaho MS, Borrell S, Comas I, Gagneux S, Yeboah-Manu D. Evaluation of customised lineage-specific sets of MIRU-VNTR loci for genotyping *Mycobacterium tuberculosis* complex isolates in Ghana. *PLoS ONE* 2014;9:e92675.
- [41] Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS ONE* 2009;4:e7815.
- [42] Cowan SL, Diem L, Monson T, Wand P, Temporado D, Oemig VT, Crawford TJ. Evaluation of a two-step approach for large-scale, prospective genotyping of *Mycobacterium tuberculosis* isolates in the United States. *J Clin Microbiol* 2005;43:688–95.