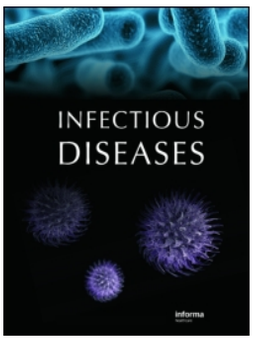

Infectious Diseases

ISSN: 2374-4235 (Print) 2374-4243 (Online) Journal homepage: https://www.tandfonline.com/loi/infd20

# Evaluation of the 15 and 24-*loci* MIRU-VNTR genotyping tools with spoligotyping in the identification of *Mycobacterium tuberculosis* strains and their genetic diversity in molecular epidemiology studies

**Dulanthi Weerasekera, Harshani Pathirane, Dushantha Madegedara, Neranjan Dissanayake, Vasanthi Thevanesam & Dhammika Nayoma Magana-Arachchi**

**To cite this article:** Dulanthi Weerasekera, Harshani Pathirane, Dushantha Madegedara, Neranjan Dissanayake, Vasanthi Thevanesam & Dhammika Nayoma Magana-Arachchi (2019): Evaluation of the 15 and 24-*loci* MIRU-VNTR genotyping tools with spoligotyping in the identification of *Mycobacterium tuberculosis* strains and their genetic diversity in molecular epidemiology studies, Infectious Diseases, DOI: 10.1080/23744235.2018.1551619

**To link to this article:** https://doi.org/10.1080/23744235.2018.1551619

Published online: 28 Jan 2019.

Taylor & Francis
Taylor & Francis Group

INFECTIOUS DISEASES,
2019; VOL. 0,
NO. 0, 1–10

https://doi.org/10.1080/23744235.2018.1551619

ORIGINAL ARTICLE



# Evaluation of the 15 and 24-*loci* MIRU-VNTR genotyping tools with spoligotyping in the identification of *Mycobacterium tuberculosis* strains and their genetic diversity in molecular epidemiology studies

Dulanthi Weerasekera[a], Harshani Pathirane[a], Dushantha Madegedara[b], Neranjan Dissanayake[c], Vasanthi Thevanesam[d] and Dhammika Nayoma Magana-Arachchi[a]

[a]National Institute of Fundamental Studies, Kandy, Sri Lanka; [b]Respiratory Disease Treatment Unit, Teaching Hospital, Kandy, Sri Lanka; [c]Consultant Respiratory Unit, District General Hospital, Nuwara-Eliya, Sri Lanka; [d]Department of Microbiology, Faculty of Medicine, University of Peradeniya, Peradeniya, Sri Lanka

## ABSTRACT

**Background:** The transmission dynamics of *Mycobacterium tuberculosis* (Mtb) using various genotyping tools has been studied globally and a particular tool for genotyping Mtb is the mycobacterial interspersed repetitive units-variable number tandem repeats (MIRU-VNTR). Tuberculosis (TB) remains an important public health problem worldwide and Sri Lanka being a country of tourist destination; because of major development projects undergoing, it has a high proportion of tourists and immigrants from Asia and Europe that are characterized with highest TB incidences and drug-resistant clinical isolates. Hence, in order to address the question of Mtb genetic diversity, we investigated the discriminatory power of both MIRU-VNTR typing of 15 and 24 *loci* with spoligotyping to differentiate Mtb isolates.

**Method:** Acid-fast bacilli positive sputum samples ($n = 150$) from first visit patients were collected. Decontamination of sputum and extraction of genomic DNA were carried out using standard techniques. The isolates were characterized by MIRU-VNTR for both the 15 and 24 *loci* and spoligotyping.

**Results:** In our study population, MIRU-VNTR 15 and 24 *loci* did not show a significant difference among the identified *M. tuberculosis* strains. However, MIRU 24 *loci* yielded an additional strain LAM, which is of T1 origin. 15 *loci* strain grouping had more clusters of strains grouped together while 24 *loci* differentiated the same cluster of strains into distinct strain types.

**Conclusion:** We conclude that the use of 15-*locus* MIRU-VNTR typing is sufficient for a first-line epidemiological study to genotype *M. tuberculosis*, but the additional discriminatory power of 24 *loci* MIRU-VNTR has been able to differentiate samples within highly homologous groups.

**KEYWORDS**
TB
; MTB
; MIRU-VNTR
; spoligotyping
; EAI
; Beijing

**ARTICLE HISTORY**
Received 30 January 2018
Revised 2 November 2018
Accepted 15 November 2018

**CONTACT**
Dhammika Nayoma Magana-Arachchi
cellbio@ifs.ac.lk
National Institute of Fundamental Studies,
Hantana Road, Kandy 20000, Sri Lanka.

Supplemental data for this article can be accessed on the publisher's website.

## Background

Despite global efforts to strive against tuberculosis (TB), the etiologic agent *Mycobacterium tuberculosis* (Mtb) is responsible for a higher mortality rate by an infectious disease globally [1]. It is believed that about 15,000–20,000 years ago, Mtb, one of the oldest known pathogen has been evolved from a common progenitor [2]. TB has been documented 5000 years ago in Egypt, 3300 years ago in India and 2300 years ago in China [3]. The evolution of *M. tuberculosis* has been studied to define the diversity, phylogeny and the transmission dynamics using various markers [4–7]. However, because of the diversity of these markers in different populations, the utility of them has been limited [8]. One tool exploited for genotyping Mtb is the mycobacterial interspersed repetitive units (MIRU), which occur in variable number tandem repeats (VNTR) consisting of multiple *loci* scattered throughout the genome which are also hyper-variable and may have an evolutionary function [9]. Comparative sequence analysis of strains H37Rv, *M. bovis* strain AF212/97 and five exact tandem repeats (ETR-A to ETR-E) previously used as a genotyping tool, were included in the 41 MIRUs [10]. MIRU-VNTR is a Polymerase Chain Reaction (PCR) based method [10]. This method was found to be adequate for large-scale prospective studies but still lacked the discriminatory power of IS6110-RFLP which provides high discriminatory power for high copy number isolates [10,11]. In order to account for the large number of samples and for recent strain evolution, a higher level of discriminatory power was required for long-term population-based studies [12]. In 2006, Supply et al. proposed an expanded set of 24 MIRU *loci* with original 12 *loci* and includes a highly discriminatory subset of 15 *loci* [13]. These 15 *loci* include six of the previous 12 *loci* with nine additional *loci* and are recommended for front-line epidemiological studies such as outbreak investigation [13,14]. The additional VNTR *loci* of both the 15 and 24 *loci* methods have significantly increased discriminatory power over the 12 *loci* method. The complete set of 24 *loci* has been recommended for phylogenetic studies [15]. Spoligotyping, when used alone, tends to overestimate the proportion of clustered strains. Therefore, all spoligotype defined clusters were further discriminated using MIRU-VNTRs [16]. When the 24 *loci* MIRU-VNTR method is coupled with spoligotyping, discriminatory power can reach or exceed that of IS6110-RFLP [13,14,17,18]. Although it is considered that Sri Lanka is a low TB prevalent country in the South-East Asia region, around 9000 new tuberculosis cases are reported every year indicating that TB is an important public health problem in Sri Lanka. The prevalence rate of tuberculosis in Sri Lanka in 2012 was 109/100,000 population and the mortality rate was 5.9/100,000 population [19]. The fact that Sri Lanka being an immigration country for migrants from Asia to Europe and due to the current major development projects in the country, it has a high proportion of migrants originating from parts of China, India and Bangladesh that are characterized with highest TB incidences and drug-resistant clinical isolates [1]. However, only a few investigations have addressed the question of Mtb genetic diversity in Sri Lanka [20]. In this study, we investigated the discriminatory power of MIRU-VNTR method for differentiating Mtb isolates from the Central Province Sri Lanka during a period of two and half years using 15 and 24 *loci* VNTR sets with spoligotyping.

## Methods

### Study population

A total of 150 acid fast bacilli (AFB) positive sputum samples from first visit patients were collected over a period from January 2012 to April 2014 from three distinct population groups in the Central Province, Sri Lanka. For this study, we used the same samples from our previous study [21]. Group I were the general population who were culture positive and having TB attending the Central Chest Clinic, Kandy, Sri Lanka ($n = 78$), Group II were the prisoners who were culture positive and having TB in Bogambara prison, Kandy ($n = 22$). The prisoners in this study population were from the Central province, Sri Lanka. Group III were the estate workers in the Central province who were culture positive and having TB ($n = 50$). The estate workers were selected from the District General Hospital, Nuwara Eliya and Central Chest Clinic, Kandy, Sri Lanka. Ethical clearance was obtained from the Ethical Review Committee, Faculty of Medicine, University of Peradeniya, Sri Lanka.

### Collection of data and specimens

A validated questionnaire was administered to the patients of the study population to obtain information on gender, date and country of birth, nationality, immigration status (if relevant), number of years of residency in Kandy or elsewhere in Sri Lanka, present address (or whether the patient is homeless), whether the patient is living in a health care institution or any public

institution, the nature of the patient's employment, socio-economic status, any previous known exposure to other persons with TB (especially within the 6 months before development of any symptoms) and the identity of patient's household contacts and respective close contacts in occupational or crowded settings if relevant. The sputum samples from the patients who gave consent to the study were collected in autoclavable small wide mouth glass bottles.

### Specimen processing, culture and extraction of genomic DNA from mycobacteria

Decontamination of sputum was done using the modified Petroff's method using 4% NaOH [22]. Lowenstein–Jensen (LJ) medium were inoculated with the suspension and all slopes were observed for occurrence of bacterial growth for 8 weeks. Genomic DNA was extracted from culture positive isolates ($n = 150$) and the standard strains of $H_{37}Rv$ and *M. bovis*, using standard CTAB/NaCl method [23].

### Genotyping of the isolates

The isolates were characterized by two genotyping methods, spoligotyping and MIRU-VNTR. *Mycobacterium bovis* BCG P3 and $H_{37}Rv$ were used as positive controls and milliQ water as a negative control.

### Spoligotyping

Spoligotyping was performed as previously described [6] using the spoligotyping kit (Occimum Biosolutions, Hyderabad, India). The spoligo results were compared with the international spoligotype databases SPOTCLUST [24] and SITVIT2 [25] of the Pasteur Institute of Guadeloupe which provides information on the shared-type distribution of *M. tuberculosis* spoligotypes at the worldwide level [26]. At the time of the comparison, the updated SITVIT2 version contained more than 90,000 patterns from more than 160 countries of patient origin.

### MIRU-VNTR typing

MIRU-VNTR genotyping was performed using PCR-amplification of a panel of 15 and 24 [13] MIRU *loci* using primers described in the MIRU-VNTR standard protocol [26]. Hotstar Taq DNA Polymerase kit (QIAGEN, Hilden, Germany) was used in the preparation of PCR premixes for 15 and 24 *loci* MIRU-VNTR. Briefly, DNA amplification was carried out in Rotor Gene Q thermal cycler (Qiagen, Germany). The PCR products were analyzed using 2% agarose gels and the sizes of the amplicons were assessed manually using the Genetool software (Syngene, UK). The MIRU copy number per *locus* was calculated by using the conventions described [13]. $H_{37}Rv$ was used to verify the results for a particular *locus* by comparing with the allele number assigned for the *locus* used. MIRU-VNTR data were analyzed using the web application MIRU-VNTR*plus* [17].

### Phylogenetic tree and data analysis

Web tools available at http://www.miruvntrplus.org [10, 27] were used to plot the minimum spanning tree (MST) and the phylogenetic radial tree using unweighted pair group method with arithmetic mean (UPGMA). Genetic diversity of each *locus* was calculated as the Hunter Gaston diversity index (HGDI) using the formula described by Hunter and Gaston [28]. The discriminatory power and allelic diversity were calculated using h value and Hunter Gaston Discriminatory Index (HGDI) using the equations: $h = 1 - \Sigma xi^2$ where $xi$ is the frequency of the $i^{th}$ allele at each *locus* [28].

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

where $N$ is the total number of isolates, $s$ is the total number of different patterns and $n_j$ is the number of isolates belonging to the $j^{th}$ pattern [29]. Clustering rate was calculated using the formula $(n_c - c)/n$, where the $n_c$ is the total number of clustered cases, $c$ is the number of clusters, $n$ is the total number of cases in the study [30].

The HGDI varies between 0.00 and 1.00, with a value of 0.00 indicating that all strains are indistinguishable and a value of 1.00 indicating that all strains in the sample are differentiated. 'Clustering' is used to partition a set of data points into groups ('clusters'), so that the points in each group share some common characteristics, typically proximity according to some distance or similarity measure. The clustering rate is denoted as a percentage, with lower percentages indicating higher discriminatory power.

## Results

### Demographic characteristics of patients

Among the total 150 patients, 79.3% were males, and the majority (30.66%) was in the age group between 46

**Table 1.** Discriminatory index and clustering rate of MIRU-VNTR and spoligotyping.

| MIRU-VNTR locus set | Number of unique genotypes | HGDI | Clustering rate (%) |
|---|---|---|---|
| 15 loci + spoligotyping | 19 | 0.896 | 25.33 |
| 24 loci + spoligotyping | 20 | 0.897 | 12.66 |
| Spoligotyping | 16 | 0.814 | 52.66 |

and 60 years. EAI lineage and its sub lineages were identified in Kandy district, Sri Lanka. EAI lineage had the highest prevalence rate of 43.23%. Beijing (8.7%), Haarlem (20%) and MANU (12.7%) strains were also prevalent in Kandy district. Of the collected data, all the patients were born in Sri Lanka, majority were from the Central Province. However, there were a few patients born in other provinces outside Central Province in Sri Lanka (Group I and Group III; see Figure S1) and attended the Central Chest Clinic to receive treatments for TB. The majority of the patients in general population (group I) were from the Central Province (85.9%) including patients from Kandy (78.2%), Matale (6.4%) and Nuwara Eliya (1.3%) districts. All the prisoners (group II) in this study were from Kandy district (100%) and the majority of the estate workers (group III) were from Kandy (28%) and Nuwara Eliya (60%) districts (Figure S1).

### *MIRU-VNTR genotyping*

Molecular typing of Mtb isolates was performed using 15 and 24 MIRU-VNTR *loci* on 150 clinical isolates from 150 patients belonging to three population groups. As an initial approach to comparing the efficiency of MIRU-15 with MIRU-24, we studied the same sample population we had already used to test the efficiency of MIRU-15 [21]. Therefore, 15 MIRU-VNTR *locus* was performed on 150 samples and 19 unique lineages/sub-lineages were identified. The alternate subset of 24 highly discriminatory MIRU-VNTR *loci* were also used to genotype the 150 study samples which yielded 20 lineages/sub-lineages (Table 3). The 15 *locus* MIRU-VNTR set provided the highest clustering rate (25.33%) and lowest HGDI (0.896) whereas, the 24 *locus* MIRU-VNTR set provided the lowest clustering rate (12.66%) and highest HGDI (0.897) of the two sets of MIRU-VNTR *loci* (Table 1).

In our study population, MIRU-VNTR 15 and 24 *loci* did not show a significant difference among the identified *M. tuberculosis* strains. However, MIRU 24 *loci* yielded an additional strain LAM, which is of T1 origin. The 24 *loci* grouping of strains differed when compared to the 15 *loci* strain grouping. 15 *loci* MIRU-VNTR yielded a total of 57 clustered cases in 19 distinct clusters whereas the 24 *loci* MIRU-VNTR yielded 33 clustered cases in 14 clusters (Figure 1(a,b)). Therefore, in comparison to the 15 *loci* strain grouping, the 24 *loci* strain grouping differentiated the cluster of strains into distinct strain types due to its higher discriminatory power. MIRU Mtub 21 and QUB-11b were found to be the most discriminatory of all MIRU *loci* with an HGDI of 0.72. The least discriminatory *locus* was *locus* MIRU 20 with the lowest HGDI (Table 2).

The HGDI of the *M. tuberculosis* isolates at each MIRU-VNTR *locus* varied significantly. Among the 24 *loci*, the HGDI for 7 *loci* (Mtub 21, QUB-2163b, ETRA, MIRU 31, Mtub 39, QUB 26, MIRU 39) exceeded 0.6, suggesting that they were highly discriminating. 12 *loci* (Mtub 04, ETRC, MIRU 04, MIRU 40, MIRU 10, Mtub 30, MIRU 26, QUB-4156, MIRU 23, MIRU 24, ETRB, Mtub 35) showed moderate discrimination ($0.3 \leq h \leq 0.6$), but MIRU 16 ($h = 0.15489$), MIRU 02 ($h = 0.015625$), MIRU 27 ($h = 0.161319$) and Mtub 29 ($h = 0.177042$) were less polymorphic. No diversity was observed for the MIRU 20 *locus* ($h = 0$). In the study presented here, the HGDI increased to 0.082 and 0.083 with the addition of spoligotyping data to 15 *loci* and 24 *loci* MIRU-VNTR data respectively (Table 1).

### *Analysis of spoligotyping data*

The spoligotyping analysis was done by assigning SIT numbers and genotypic clade designations by comparing with the SPOTCLUST and SITVIT2 databases, and the results obtained are summarized in Table 3. This table shows data both for all the 150 isolates versus data by taking only one isolate per patient ($n = 150$ isolates). The overall partition of strains according to major phylogenetic *M. tuberculosis* clades was performed according to signatures provided in SITVITWEB [25]. The six major clades observed ranked in the following order: East-African-Indian (EAI-59/150; 39.33%), Haarlem (H-30/150; 20%), Beijing (13/150; 8.6%), Central European family T (10/150; 6.5%), European Family X (8/150; 5.2%) and Central and Middle Eastern Asian (Delhi/CAS-1/150; 0.6%). Additionally, six other sub-lineages were identified including; MANU (19/150; 12.6%), EAI3-IND (4/150; 2.6%), EAI6-BGD1 (2/150; 1.3%) and *M. tuberculosis* S/Québec (1/150; 0.6%), Family36 (2/150; 1.3%) and Family35 (1/150; 0.6%). Of the 13 isolates belonging to Ural (U) lineages, six 'U likely H' (SIT124), five 'U likely H3' (SIT124) and two 'U likely EAI' (SIT523) were identified (Table 3 and Table S1). These results indicate that EAI is the most endemic lineage in Sri Lanka with reference to the

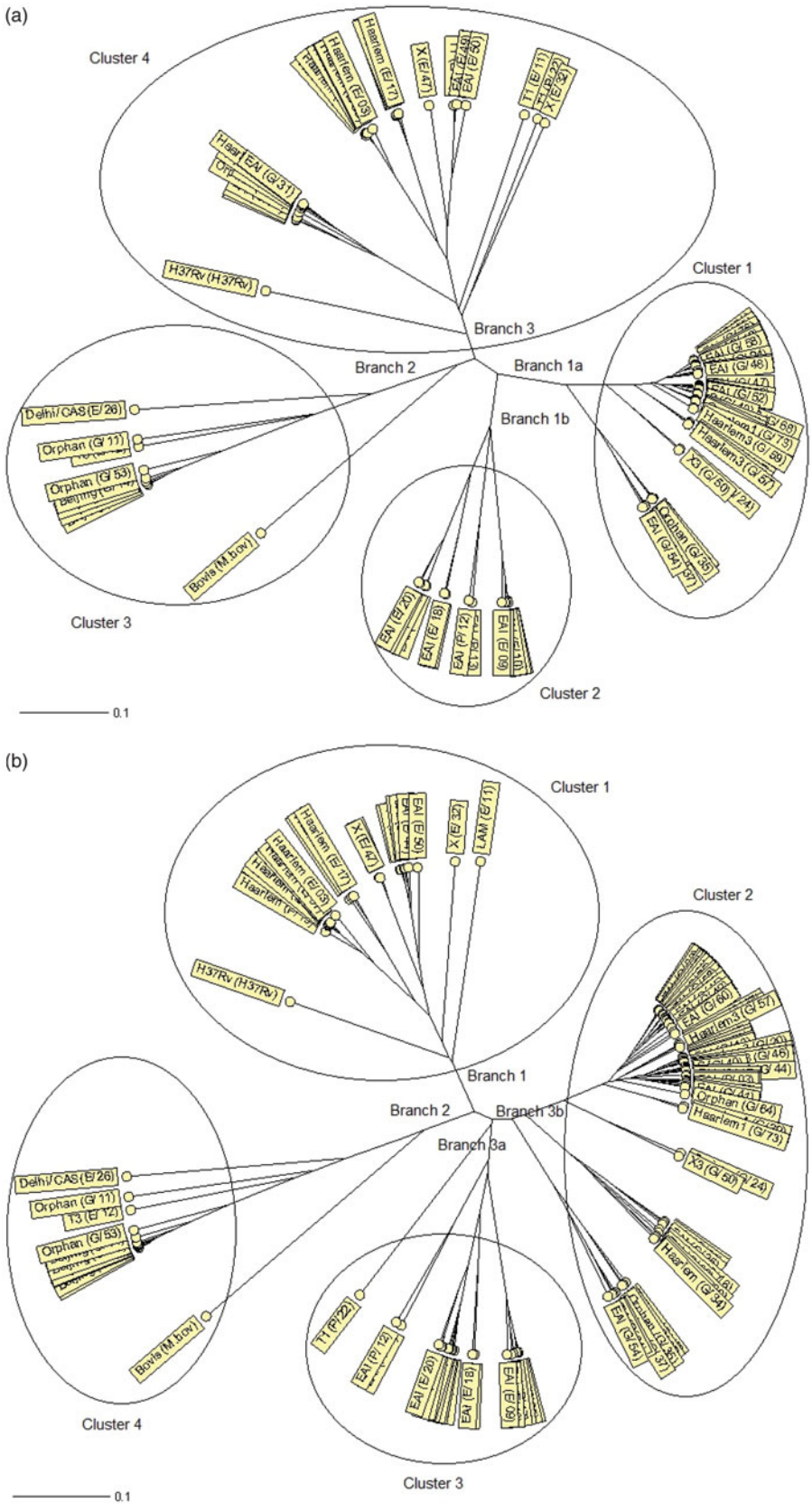


**Figure 1.** (a) UPGMA-radial tree displaying the genetic relationships of 150 isolates (Group I, Group II and Group III) of *M. tuberculosis* based on 15 MIRU-VNTR *loci* and spoligotyping. The linkage distance scale is indicated at the bottom. (b) UPGMA-radial trees displaying the genetic relationships of 150 isolates (Group I, Group II and Group III) of *M. tuberculosis* based on 24 MIRU-VNTR *loci* and spoligotyping. The linkage distance scale is indicated at the bottom.

**Table 2.** Hunter–Gaston discriminatory index (HGDI) for the loci in MIRU-15 and MIRU-24 sets.

| | | Copy number | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Allele | $n = 150$ | 0s | 1s | 2s | 3s | 4s | 5s | 6s | 7s | 8s | 9s | HGDI |
| **Mtub 04** | 149 | 0 | 80 | 52 | 1 | 16 | 0 | 0 | 0 | 0 | 0 | **0.58** |
| **ETRC** | 149 | 0 | 0 | 5 | 36 | 107 | 1 | 0 | 0 | 0 | 0 | **0.43** |
| **MIRU 04** | 147 | 0 | 0 | 60 | 0 | 14 | 73 | 0 | 0 | 0 | 0 | **0.58** |
| **MIRU 40** | 147 | 0 | 0 | 71 | 59 | 0 | 17 | 0 | 0 | 0 | 0 | **0.59** |
| **MIRU 10** | 146 | 0 | 0 | 17 | 14 | 102 | 13 | 0 | 0 | 0 | 0 | **0.48** |
| **MIRU 16** | 145 | 0 | 1 | 10 | 133 | 1 | 0 | 0 | 0 | 0 | 0 | **0.15** |
| **Mtub 21** | 148 | 0 | 1 | 2 | 22 | 38 | 10 | 12 | 0 | 0 | 63 | **0.72** |
| **QUB-2163b** | 148 | 0 | 0 | 15 | 37 | 10 | 0 | 18 | 0 | 3 | 65 | **0.72** |
| **ETRA** | 149 | 0 | 0 | 2 | 38 | 84 | 0 | 12 | 10 | 3 | 0 | **0.61** |
| **Mtub 30** | 148 | 0 | 10 | 78 | 0 | 60 | 0 | 0 | 0 | 0 | 0 | **0.56** |
| **MIRU 26** | 147 | 0 | 0 | 88 | 1 | 13 | 42 | 0 | 0 | 0 | 3 | **0.55** |
| **MIRU 31** | 146 | 0 | 0 | 0 | 37 | 70 | 29 | 10 | 0 | 0 | 0 | **0.66** |
| **Mtub 39** | 148 | 0 | 0 | 57 | 54 | 19 | 4 | 14 | 0 | 0 | 0 | **0.69** |
| **QUB-26** | 148 | 0 | 0 | 0 | 0 | 3 | 25 | 51 | 65 | 4 | 0 | **0.66** |
| **QUB-4156** | 148 | 0 | 40 | 82 | 25 | 1 | 0 | 0 | 0 | 0 | 0 | **0.59** |
| **MIRU 02** | 128 | 0 | 0 | 127 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | **0.015** |
| **MIRU 23** | 128 | 0 | 0 | 0 | 0 | 18 | 78 | 31 | 1 | 0 | 0 | **0.55** |
| **MIRU 39** | 120 | 0 | 8 | 49 | 53 | 10 | 0 | 0 | 0 | 0 | 0 | **0.63** |
| **MIRU 20** | 132 | 0 | 0 | 132 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **0** |
| **MIRU 24** | 146 | 3 | 86 | 52 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | **0.53** |
| **MIRU 27** | 116 | 0 | 2 | 0 | 106 | 8 | 0 | 0 | 0 | 0 | 0 | **0.16** |
| **Mtub 29** | 128 | 0 | 6 | 1 | 3 | 116 | 2 | 0 | 0 | 0 | 0 | **0.18** |
| **ETRB** | 128 | 0 | 10 | 20 | 98 | 0 | 0 | 0 | 0 | 0 | 0 | **0.38** |
| **Mtub 35** | 121 | 0 | 12 | 10 | 92 | 2 | 5 | 0 | 0 | 0 | 0 | **0.41** |

**Table 3.** Distribution of study isolates in various lineages as determined using spoligotyping and MIRU-VNTR*plus* databases.

| Lineage/ Sub-lineage | Genotype/SpolDB4 SIT | Strain assignments using 15 MIRU-VNTR loci | Strain assignments using 24 MIRU-VNTR loci |
|---|---|---|---|
| EAI | Orphan | 34 | 34 |
| EAI5 | 126, 138, 236, 934, 1090, 2674, Orphan | 19 | 19 |
| MANU1 | 100, 937 | 15 | 15 |
| Ural | 124, 523 | 13 | 13 |
| H3 | 50, 124, 533, Orphan | 7 | 7 |
| T1 | 53, 159, 573 | 8 | 7 |
| MANU2 | 54, 1634 | 4 | 4 |
| EAI6_BGD1 | 591 | 1 | 1 |
| EAI3_IND | 11, Orphan | 2 | 2 |
| S | 34 | 1 | 1 |
| X3 | 1157 | 2 | 2 |
| Beijing | 1, Orphan | 13 | 13 |
| Delhi/CAS | Orphan | 1 | 1 |
| H1 | 47, 1952 | 2 | 2 |
| T4 | Orphan | 1 | 1 |
| T3 | Orphan | 1 | 1 |
| H | Orphan | 10 | 10 |
| X | 2180, Orphan | 6 | 6 |
| Orphan | – | 10 | 10 |
| LAM | 159 | 0 | 1 |

isolates in the study presented here. This result also supports the findings that the EAI (CAS) lineage was the most common and widely prevalent across India, whereas EAI lineage was more common in south and eastern India. In our study, 126/150 isolates belonged to 25 shared types which have already been classified in SITVIT2 database. The EAI lineage with SIT126 pre-dominated the strains among the general population. Beijing strain was only identified among the general population where shared type (ST) 1 of Beijing lineage being the most prevalent ST among the Beijing isolates in this study (12; 8%) and Haarlem (20%) being the pre-dominant strain among the estate workers. Regarding the world-wide distribution of major shared types encountered in the present study, it is important to underline that EAI pre-dominated in our setting with 59/150 (39.33%) of the isolates. It was followed by SIT124 (Haarlem lineage) with 17/150 (11.33%) of strains and SIT100 (MANU lineage) 14/150 (9.3%) of the isolates. Spoligotyping of the 150 isolates led to a total of 26 different spoligo patterns, whereas 10 patterns were not yet reported and corresponded to orphans (Table S1). Used alone, spoligotyping grouped 100 strains in 21 clusters (2–14 strains per cluster) with an overall clustering rate of 52.66% (Table 1).

### *Phylogenetical analysis using UPGMA radial tree and minimum spanning tree (MST)*

We have used UPGMA radial trees and MSTs to determine the phylogeny of the isolates based on spoligotyping and 15 and 24 *loci* MIRU-VNTR. In both the 15 and 24 radial trees, we have observed that there were three major branches in which one of the three branches gave rise to two sub-branches, EAI and the European lineages (Figure 1(a,b)). Branch three had the largest number of nodes as compared to branches 1 and 2. Additionally, in both the trees, EAI strains were more homogeneous than the other strains. Beijing isolates were the most homogeneous with the least number of nodes and longest branch lengths. According to previous studies by Chatterjee et al., it was noted that the analysis based on UPGMA may mislead due to the assumption of the nearest ancestor, but it still yielded the differential clonal nature of the strains. The Beijing isolates were found to be the most clonal with least genetic distances in the branches while the Haarlem isolates with longer branches reflected the greater divergence among the isolates. When the 15 and 24 *loci* radial trees were compared, the 24 *loci* tree had replaced T1 strain of the 15 *loci* tree in the branch 1 by LAM strain and have added T3 strain to the branch 2 and T1 strain to the branch 3 in 24 *loci* radial tree (Figure 1(b)). Additionally, when the two trees were compared, the arrangements of strains in clusters were observed to be different (Figure 1(a,b)).

Figure 2(a,b) illustrate the MST summarizing potential evolutionary relationships of *M. tuberculosis* 15 and 24 *loci* MIRUs [Figure 2(a,b)]. Using this approach, one considers that all intermediate stages are present within the

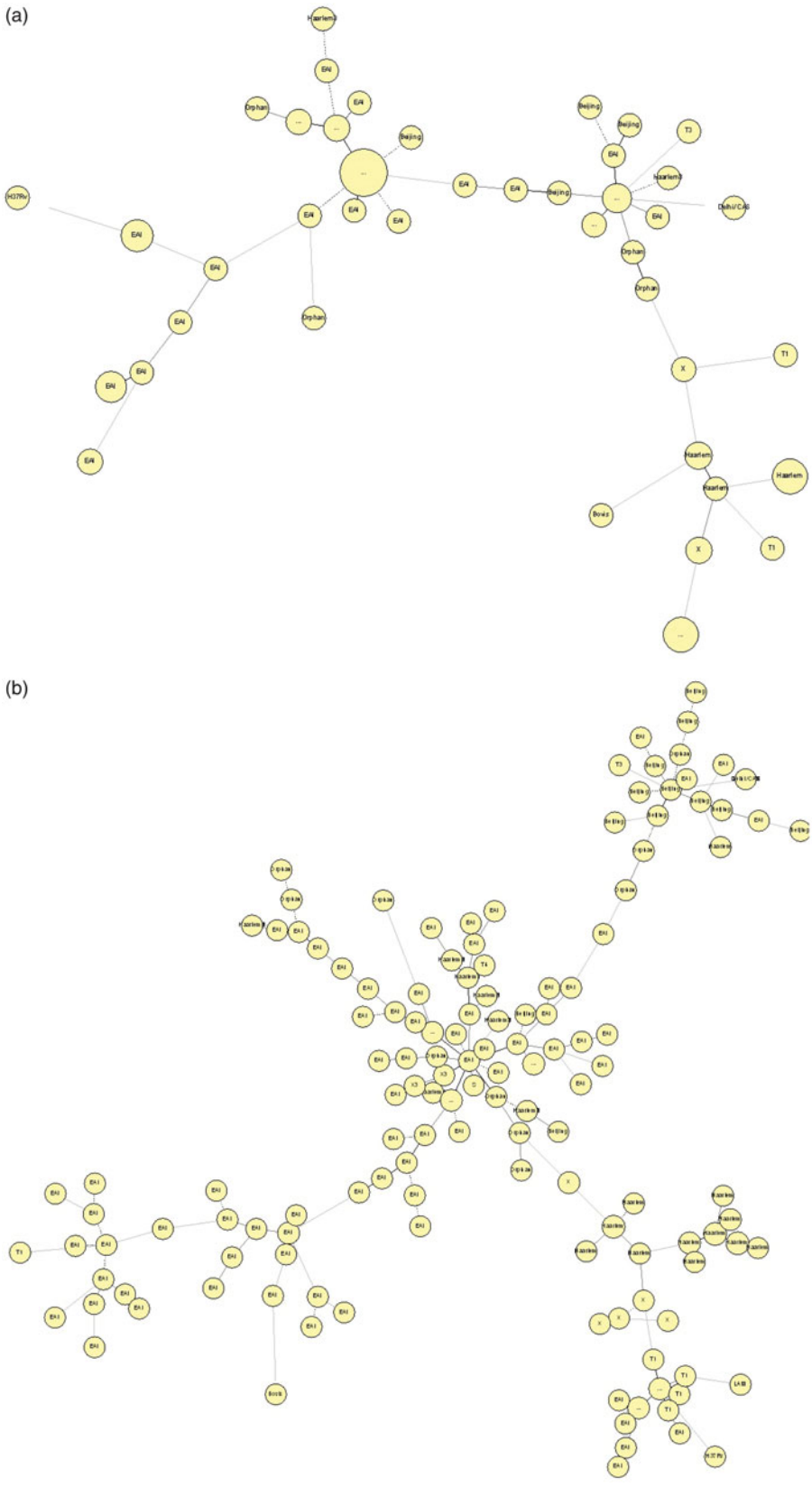


**Figure 2.** (a) Minimum spanning trees (MSTs) belonging to 15 *loci* MIRU-VNTR. The length of the branches represents the distance between patterns while the complexity of the lines (continuous, black dotted and grey dotted) denotes the number of allele/spacer changes between two patterns. (b) Minimum spanning trees (MSTs) belonging to 24 *loci* MIRU-VNTR. The length of the branches represents the distance between patterns while the complexity of the lines (continuous, black dotted and grey dotted) denotes the number of allele/spacer changes between two patterns.

sample analyzed by first including the individual that shows the greatest number of possible linkages to other individuals in the population studied. We used this method to highlight the links between the 15 *loci* MIRU and 24 *loci* MIRU differing by changes observed in their number of copies for 15 and 24 MIRU patterns (lineages/sub-lineages are marked on this tree). The structure of the tree is represented by branches (continuous vs. dotted lines) and circles representing each individual pattern (Figure 2(a,b)).

15 MIRU led to well organized and highly linked strains belonging to the evolutionary recent EAI lineage that in our study comprised MANU, Family33, EAI3-IND, with two major sub branches, as well as the EAI6-BGD1 sub-lineage. In addition, the strains Haarlem, T and X were identified. Finally, MIRU typing resulted in an efficient splitting of the spoligotyping clusters (compatible with results summarized in Table 3), with sub-cluster strains being well organized around their closest parental strain.

## Discussion

Genotyping of Mtb isolates is a useful tool, not only for local contact investigations but for its ability to allow comparisons of TB worldwide. Molecular typing by MIRU-VNTR has been used in epidemiology studies and its stability is found to be adequate for tracking recent transmission and distinguishing relapses and re-infections [31]. The system based on 12 *loci* is most widely used among the different sets of MIRU-VNTR *loci* [32]. However, it is not effective for the analysis of clustered isolates [33]. Other sets of MIRU-VNTR *loci*, such as the 14 *loci* set and the 15 *loci* set have improved the discrimination of unrelated isolates [34]. An optimized set of 24 *loci* has also been defined. However, studies have suggested that all 24 *loci* are not required for genotyping *M. tuberculosis* strains in any given situation [13] as the number of *loci* required depends on the lineage known to be prevalent in the investigated area [35].

In the present study, the analysis of the two sets of MIRU-VNTR *loci* showed the expected results. Though the HGDI is only increased by 0.1% when comparing 15 and 24 MIRU-VNTR *loci*, the clustering rate showed a noticeable decrease of 12.67%. According to previous studies, it was noted that the basic idea behind the identification of a cluster is the notion that clustering indicates an ongoing or recent transmission, whereas unique patterns indicate reactivation events [36]. The 24 *loci* subset, as is predicted, provided the increased resolution compared to the 15 *loci* method, with the 24 *loci* method providing slight improvements over the 15 *loci* subset. However, the 15 *loci* methodology is known as a highly discriminatory method for first-line genotyping of Mtb isolates and is recommended to consider as a replacement for the 12 *loci* method [12,34]. Approximately, one-third of the isolates were clustered into groups of 19 indistinguishable 15 *loci* MIRU patterns. These groups generally contained samples from a single district, but in some cases there are isolates from multiple districts involved in the clusters. The application of the 15 and 24 *loci* MIRU-VNTR methods allow us to further differentiate these groups. Instances where indistinguishable patterns cross district boundaries, highlight the need for a national surveillance program that can facilitate real-time tracking and contact investigations associated with a potentially highly mobile individual or group. Studies have shown that transmission across regions is not rare and tracking can be accomplished efficiently when adequate data is available [12]. Data of our study also illustrates that the higher discriminatory power provided by the 24 MIRU *loci* could help to eliminate the investigations of spurious relationships.

In our study, *loci* MIRU 10 and Mtub 30 showed moderate diversity which was similar to studies done previously [35,37] by genotyping the isolates epidemic, particularly in Asia. The 7 *loci* (Mtub 21, QUB-2163b, ETRA, MIRU 31, Mtub 39, QUB 26, MIRU 39) in our study which showed a higher diversity has shown a similar result in a previous study as well [37]. When compared to previous studies, the Sri Lankan samples represented here has a relatively low HGDI and high clustering rates [17,18]. This is due to the highly homologous nature of the samples. Moreover, the application of the 15 or 24 *loci* MIRU-VNTR sets were able to discriminate indistinguishable spoligo patterns of the isolates into distinct clusters based on the MIRU pattern. According to previous studies, in areas with more diverse populations, this level of discrimination was not required for immediate investigations, but has proved fruitful over the long-term tracing in regional outbreaks or the evolution of a strain over time [38].

For a successful TB control program, studying the molecular epidemiology of *M. tuberculosis* clinical isolates is important, as this unravels the disease transmission dynamics and highlights the most successful strains circulating within a population in a specific geographical region. The addition of spoligotyping data to 24 *loci* MIRU-VNTR data provides a very high level of discrimination. To address regional concerns and to quickly

differentiate strains of interest in an outbreak setting, alternate sets of MIRU *loci* have been proposed [13,39]. However, these methods are found to be inadequate for standardized, long-term, nation-wide surveillance. According to a study by Christianson et al., it is noted that in order to provide long-term surveillance of high volumes of samples, higher level of discriminatory power is required and according to their study, this high level of discrimination is provided using a 24 *loci* MIRU-VNTR typing [12]. Furthermore, the addition of spoligotyping can help to further differentiate clustered samples to the level of IS6110-RFLP.

## Conclusion

The genotyping using 24 MIRU-VNTR *loci* had a greater utility in the classification of Mtb strains into lineages and sub-lineages than 15 *loci* typing in the setting of the Central Province, Sri Lanka. The use of 15 *loci* MIRU-VNTR typing was sufficient for a first-line epidemiological study to genotype *M. tuberculosis* isolates. However, the additional discriminatory power of 24 *loci* MIRU-VNTR was able to differentiate samples within highly homologous groups and to identify occurrences of indistinguishable patterns among different geographical localities at a high level of discrimination when compared to 15 *loci* MIRU-VNTR method. Therefore, the use of the 24 MIRU *loci* for long-term surveillance will be advantageous for providing rapid and highly discriminatory genotyping data.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

The authors thank the grants received from National Research Council (NRC/07/47 and NRC/11/059) and National Science Foundation (RG/2006/HS/07) of Sri Lanka for providing instrumental support for this work.

## References

[1] World Health Organization. Global tuberculosis report 2015. 20th edition. World Health Organization; 2015;204. Available from: http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf?ua=1

[2] Gutierrez MC, Brisse S, Brosch R, et al. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. PLoS Pathog. 2005;1:55–61.

[3] Daniel TM. The history of tuberculosis. Respir Med. 2006; 100:1862–1870.

[4] Groenen PMA, Bunschoten AE, Van Soolingen D, et al. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. Mol Microbiol. 1993; 10:1057–1065.

[5] Hershberg R, Lipatov M, Small PM, et al. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. PLoS Biol. 2008;6:2658–2671.

[6] Kamerbeek J, Schouls L, Kolk A, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol. 1997;35: 907–914.

[7] Van Soolingen D, Hermans PWM, De Haas PEW, et al. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: Evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. J Clin Microbiol. 1991;29:2578–2586.

[8] Cronin WA, Golub JE, Magder LS, et al. Epidemiologic usefulness of spoligotyping for secondary typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS6110. J Clin Microbiol. 2001;39:3709–3711.

[9] Chatterjee A, Mistry N. MIRU-VNTR profiles of three major *Mycobacterium tuberculosis* spoligotypes found in western India. Tuberculosis (Edinb). 2013;93:250–256.

[10] Weniger T, Krawczyk J, Supply P, et al. MIRU-VNTRplus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. Nucleic Acids Res. 2010;38: 326–331.

[11] Han H, Wang F, Xiao Y, et al. Utility of mycobacterial interspersed repetitive unit typing for differentiating *Mycobacterium tuberculosis* isolates in Wuhan, China. J Med Microbiol. 2007;56:1219–1223.

[12] Christianson S, Wolfe J, Orr P, et al. Evaluation of 24 *locus* MIRU-VNTR genotyping of *Mycobacterium tuberculosis* isolates in Canada. Tuberculosis. 2010;90:31–38.

[13] Supply P, Allix C, Lesjean S, et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol. 2006;44: 4498–4510.

[14] Oelemann MC, Diel R, Vatin V, et al. Assessment of an optimized mycobacterial interspersed repetitive- unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. J Clin Microbiol. 2007;45:691–697.

[15] Kato-Maeda M, Metcalfe JZ, Flores L. Genotyping of *Mycobacterium tuberculosis*: application in epidemiologic studies. Future Microbiol. 2011;6:203–216.

[16] Sola C, Filliol I, Gutierrez MC, et al. Spoligotype database of *Mycobacterium tuberculosis*: biogeographic distribution of shared types and epidemiologic and phylogenetic perspectives. Emerg Infect Dis. 2001;7:390–396.

[17] Allix-Béguec C, Fauville-Dufaux M, Supply P. Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol. 2008; 46:1398–1406.

[18] Maes M, Kremer K, van Soolingen D, et al. 24-*Locus* MIRU-VNTR genotyping is a useful tool to study the molecular epidemiology of tuberculosis among Warao Amerindians in Venezuela. Tuberculosis. 2008;88:490–494.

[19] World Health Organization. Global tuberculosis report 2013. WHO 2013;306. Available from: http://apps.who.int/iris/bitstream/10665/91355/1/9789241564656_eng.pdf

[20] Magana-arachchi D. Pattern of circulating *Mycobacterium tuberculosis* strains in Sri Lanka. In: Cardona PJ, editor. Understanding tuberculosis-global experiences and innovative approaches to the diagnosis. London: InTech; 2012. p. 511–26. ISBN: 978-953-307-938-7. Available from: http://www.intechopen.com/books/understanding-tuberculosis-global-experiences-and-innovative-approaches-to-the-diagnosis/pattern-of-circulating-mycobacterium-tuberculosis-strains-in-sri-lanka-1999.

[21] Weerasekera D, Magana-Arachchi D, Madegedara D, et al. Genetic diversity of *Mycobacterium tuberculosis* isolates obtained from three distinct population groups in the Central Province, Sri Lanka. Asian Pacific J Trop Dis. 2015;5:385–392.

[22] Morcillo N, Imperiale B, Palomino JC. New simple decontamination method improves microscopic detection and culture of mycobacteria in clinical practice. Infect Drug Resist. 2008;1:21–26.

[23] Somerville W, Thibert L, Schwartzman K, et al. Extraction of *Mycobacterium tuberculosis* DNA: a question of containment. J Clin Microbiol. 2005;43:2996–2997.

[24] Vitol I, Driscoll J, Kreiswirth B, et al. Identifying *Mycobacterium tuberculosis* complex strain families using spoligotypes. Infect Genet Evol. 2006;6:491–504.

[25] Demay C, Liens B, Burguière T, et al. SITVITWEB – A publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. Infect Genet Evol. 2012;12:755–766.

[26] Alonso-Rodríguez N, Martínez-Lirola M, Herránz M, et al. Evaluation of the new advanced 15-*loci* MIRU-VNTR genotyping tool in *Mycobacterium tuberculosis* molecular epidemiology studies. BMC Microbiol. 2008;8:34.

[27] Ali A, Hasan Z, Tanveer M, et al. Characterization of *Mycobacterium tuberculosis* Central Asian strain 1 using mycobacterial interspersed repetitive unit genotyping. BMC Microbiol. 2007;7:76.

[28] 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;26:2465–2466.

[29] Affolabi D, Anyo G, Faïhun F, et al. First molecular epidemiological study of tuberculosis in Benin. Int J Tuberc Lung Dis. 2009;13:317–322.

[30] Torkaman MRA, Nasiri MJ, Farnia P, et al. Estimation of recent transmission of *Mycobacterium tuberculosis* strains among Iranian and Afghan immigrants: a cluster-based study. J Clin Diagn Res. 2014;8:5–8.

[31] Supply P, Lesjean S, Savine E, et al. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J Clin Microbiol. 2001;39:3563–3571.

[32] Supply P, Mazars E, Lesjean S, et al. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. Mol Microbiol. 2002;36:762–771.

[33] Shabbeer A, Cowan LS, Ozcaglar C, et al. TB-Lineage: an online tool for classification and analysis of strains of *Mycobacterium tuberculosis* complex. Infect Genet Evol. 2012;12:789–797.

[34] Wang J, Liu Y, Zhang CL, et al. Genotypes and characteristics of clustering and drug susceptibility of *Mycobacterium tuberculosis* isolates collected in Heilongjiang province, China. J Clin Microbiol. 2011;49:1354–1362.

[35] Zozio T, Allix C, Gunal S, 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.

[36] Ahmed MM, Mohammed SH, Nasurallah HAA, et al. Snapshot of the genetic diversity of *Mycobacterium tuberculosis* isolates in Iraq. Int J Mycobacteriol. 2014;3:184–196.

[37] Crawford JT. Genotyping in contact investigations: a CDC perspective. Int J Tuberc Lung Dis. 2003;7:453–457.

[38] Ashworth M, Horan KL, Freeman R, et al. Use of PCR-based *Mycobacterium tuberculosis* genotyping to prioritize tuberculosis outbreak control activities. J Clin Microbiol. 2008;46:856–862.

[39] Millet J, Miyagi-Shiohira C, Yamane N, et al. High-resolution MIRU-VNTRs typing reveals the unique nature of *Mycobacterium tuberculosis* Beijing genotype in Okinawa, Japan. Infect Genet Evol. 2012;12:637–641.