Insight into population structure of *Mycobacterium tuberculosis* isolates in the multiethnic province of Alborz, Iran

Morassa Sadat Farnad[1], Hassan Momtaz[1], Nader Mosavari[2], Yahya Khosravi[3], Keyvan Tadayon[1,2*]

[1]Department of Microbiology, Shahrekord Branch, Islamic Azad University, Sharekord, Iran
[2]Department of Tuberculin & Mallein, Razi Vaccine & Serum Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran
[3]Department of Occupational Health and Safety Engineering, Non-communicable Diseases, Research Center, Alborz University of Medical Sciences, Karaj, Iran

ABSTRACT

Background and Objectives: Genetic diversity of *Mycobacterium tuberculosis* clinical isolates from tuberculosis patients in the multiethnic province of Alborz, Iran was assessed.

Materials and Methods: A total of 17 isolates in the period of 2012-2013 were collected and subjected to a Multiple-locus variable-number tandem repeat (VNTR) analysis (MLVA) consisted of 6 variable numbers of tandem repeats (VNTRs) including ETR-A, ETR-B, ETR-C, ElTR-D, ETR-E, ETR-F, 5 Mycobacterial Interspersed Repetitive Units including MIRU10, MIRU16, MIRU26, MIRU39, MIRU40, and 1 Queen University of Belfast locus, QUB11.

Results: This classified all isolates into 17 distinct MIRU-VNTR types, a reflection of a highly heterogenic population. Within the 12 used VNTR loci, ten proved highly or moderately discriminant according to the calculated HGDI scores. No cluster of isolates was identified in the study panel, giving a clustering rate of 0%, several events of SVL (N=5) and DVL (N=4) and TVL (N=3) were detected.

Conclusion: The greater heterogeneity observed here by MLVA-VNTR analysis is most likely due to limited background data in the study region rather than a genuine more heterogeneous population compared to other provinces of the country.

Keywords: *Mycobacterium tuberculosis*; Population; Genotype; Genetic heterogeneity; Polymorphism; Epidemiologic; Locus

INTRODUCTION

Tuberculosis (TB), known to mankind since antiquity, has yet remained a serious threat to the public health in low- and middle-income countries. Figures released by World Health Organization (WHO) showed an estimated 10 million new cases globally worsened by 1.6 million deaths only in 2017 (1). Moreover, in 2017 Asian and African continents were home to almost Two-thirds of the whole world TB cases where India (27%), China (9%) and Indonesia (8%) along with the Philippines (6%) jointly hosted over half of the world victims (1).

Substantial genomic and epidemiological knowledge of *Mycobacterium tuberculosis* complex (MTC) in the fields of transmission dynamics, genetic diversity and phylogenetic have been provided over the last three decades using various genotyping tools

*Corresponding author: Keyvan Tadayon, Ph.D, Department of Tuberculin & Mallein, Razi Vaccine & Serum Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran.
Tel: +98-2634502892
Fax: +98-2634552194
Email: k.tadayon@rvsri.ac.ir*

such as IS*6110*-RFLP, large sequence polymorphism (LSP), spoligotyping, Single Nucleotide Polymorphism (SNP), mycobacterial interspersed repetitive unit- variable number tandem repeat (MIRU-VNTR) typing and more recently Whole-genome sequencing. According to these findings, MTC strains are highly clonal bacteria representing relatively static genomic content where up to 99.9% nucleotide sequence identity is typically a normal observation (2). Until now seven globally important lineages of human-adapted MTC strains have been defined including lineages scattered globally (e.g. lineages 2 and 4) and other lineages (5 and 6) that tend to operate locally. This is believed to be an indication of the difference in transmission and drug resistance properties of strains (3).

The molecular epidemiology of MTC in human has recently been the subject of numerous studies by Iranian workers (4-7) though, need for further epidemiological analysis of actively circulating MTC genotypes in Iran seems not fulfilled as yet.

The Alborz province, located in the north-west of the country's capital, Tehran is the latest parliamentary-approved province of Iran formed by division of Tehran province in 2010. Surrounded by Tehran, Qazvin, Mazandaran and Markazi provinces, as many as 2.5 million people lived here in 2018. A typically multiethnic region with its population comprised of almost all the country's ethnicities and foreign nationals dominated by Afghan and Iraqi settlers. Alborz is represented by a relatively higher TB burden at the national level according to the official medical records. The genetic diversity of MTC population in this province has remained largely unknown so, to generate baseline data, in the present work, we used Multiple-locus variable- number tandem repeat (VNTR) analysis (MLVA).

MATERIALS AND METHODS

Ethics statement. All the study participants were briefed with the research objectives and its procedures with written informed consent obtained from all the subjects. Ethical approval required for this retrospective genotyping work was issued by the relevant authorities at the Ethics Committee of the Alborz University of Medical Sciences.

Collection of samples and isolation. Twenty sputum specimens were collected from randomly selected 20 pulmonary TB-suspected outpatients. These were visiting patients to the university hospitals of Alborz from December 2012 to October 2013.

Demographic data of patients were collected along with their consent through questionnaires filled out at the clinic when specimens were taken.

Treating of the samples by Petroff's method. The collected specimens were transferred to Razi institute in Hesarak, Karaj where they were subjected to Petroff's protocol for digestion/decontamination. For bacterial culture, processed (digested and decontaminated) specimens were used to inoculate slopes of plain (glycerinated) and Lowenstein–Jensen medium (LJ) supplemented with sodium pyruvate and glycerol. Incubation of cultured Falcon tubes at 37°C continued for 12 weeks with weekly close inspection in search for traces of bacterial growth. Slopes bearing visual properties of MTC bacterial mass were tested by Acid-fast staining microscopy examination (AFB).

DNA extraction and molecular amplification. Bacterial colonies from MTC-suspected cultures were scraped from the slopes and re-suspended in 200 µL of TB lysis® Solution (SinaClon BioScience, Iran) in microfuge tubes equipped with O-rings. Microtubes were heat-treated at 96°C for 30 min in a boiling water bath. The inactivated bacterial suspension was centrifuged (9,750 g; 10 min) and the DNA holding supernatant was retrieved into two new tubes including one that was stored at -20°C and the second that was used for PCR experiments. The primers were synthesized by Macrogen, (Seoul, South Korea). The PCR mixtures were prepared using 2×Taq Master Mix (Ampliquor, Denmark).

PCR-16S rRNA, PCR-IS*6110*. Two 543 and 245 bp long fragments of the 16S rRNA gene and IS*6110* marker from isolates were assessed according to the Huard (8) and McHugh (9) PCR methods, respectively.

RD-Typing. Four individual PCRs targeting specific portions of RD1, RD4, RD9, and RD12 were performed as previously described by Warren et al. to differentiate members of *M. tuberculosis* complex (10).

MLVA-VNTR typing and data analysis. A MLVA-VNTR typing method based on the 12 broadly-used loci namely, ETR-A, ETR-B, ETR-C, ETR-D,

ETR-E, ETR-F, MIRU10, MIRU16, MIRU26, MIRU39, MIRU40, and QUB11b, was conducted following the PCR protocol described elsewhere. The number of tandem repeats at the selected loci was determined through visual comparison of the size of the amplification products in Electrophoresis gels against the known size of bands from both the DNA step ladder and those from already-identified bands of *M. tuberculosis* H37Rv genome. Loci with failure PCR amplification or undefined results were re-tested for three times when sizing of the given locus was eventually possible or it was scored a zero.

The allele copy numbers represented in digital Binary-format where each digit signified the number of copies at a particular locus. These were added manually to a Microsoft Excel spreadsheet and used for drawing dendrogram and the minimum spanning tree (MST) to enable classification of the isolates and phylogenetic analysis. In order to determine the individual discriminatory power of the 12 tandem repeat loci, the Nei's diversity index (D) was calculated as D = 1 - Σ (allele frequency)2 and the loci were grouped into highly (HGDI>0.6), moderately (HGDI>0.3) or poorly discriminative (HGDI<0.3). Similarly, the Hunter–Gaston discriminatory index (HGDI) was used to determine the combinational discriminatory power of the 12 loci on the whole as a genotyping system (11). The HGDI represents the average prob- ability that the typing system will assign a different type to two unrelated strains that have been randomly sampled in the microbial population of a given taxon. The calculations were conducted online by using the calculator tool available at the website http://insilico.ehu.es/mini_tools/discriminatory_power/ (12).

For clustering analysis, the assessment of genetic relationships and comparison of strain genotypes, the BioNumerics 7.6 software (Applied Maths, Sint-Martens-Latem, Belgium) was consulted where similarity matrices applying the unweighted pair-grouping method with arithmetic averages algorithm (UPGMA) was used. A minimum spanning tree (MST) was constructed to represent an undirected network of MIRU-VNTR genotypes that are linked together where nearest neighbors are displayed by the smallest possible linkages.

To calculate the recent transmission rate, the N-1 method was employed using the $CR=(N_c-C)/N$ equilibrium, where N is the total number of cases in the sample, c stands for the number of clusters, and N_c represents the total number of clustered cases (13).

Isolates representing identical MIRU-VNTR profiles were treated as members of the same cluster.

RESULTS

The study panel included 20 patients with Iranian nationality. The man to woman ratio was roughly 2:1 (12 males versus 5 females). In bacterial culture, preferential growth was observed for the LJ slopes in all the 20 isolates. All of the 20 isolates, produced the expected 543 bp long PCR target fragment of the 16S rRNA gene proving their identity as *Mycobacterium* species. Besides, 17 isolates produced the expected 245 bp fragment of IS*6110* locus and were therefore identified as *M. tuberculosis* complex. This was re-confirmed by RD-typing test.

The allele numbers at each MIRU-VNTR locus ranged between 2 (for loci ETR-B and ETR-D) and 5 (for locus MIRU16). The allelic diversity (h) was in general high with the exception of locus ETR-D that showed 3 repeat units in 99% of isolates, yielding h=0.10 indicated that locus ETR-D was highly conserved.

Successful amplification of 12 loci among the 17 isolates, gave 17 distinct non-clustered MIRU-VNTR profiles.

Comparative analysis of the identified MIRU-VNTR patterns helped to assess relative evolutionary rates of the individual loci. This was analyzed through calculating frequency of these loci in SLVs (Single-Locus Variables), DLVs (Double-locus Variable), or TLVs (Treble-Locus Variable) within the isolates. Out of the 17 isolates, 17 clusters were identified among which 9 were SLV and 13 were DLV and 17 were TLV (Table 1).

Using the minimum spanning tree algorithm, the MIRU-VNTR-based genetic relationships among the isolates in the study panel along with those from Khorasan Razavi (N=93) and Markazi (N=53) provinces were visualized. This was conducted in a way that clonal complexes were defined arbitrarily as groups of isolates gathering according to similarity level of their MIRU-VNTR profiles with a maximal tolerance of three locus difference (Fig. 1).

DISCUSSION

The current work seems to be is the first study

Table 1. Epidemiological and MLVA genotyping details of 17 isolates in the present study.

Isolate order	Isolate ID	Province	Sex	Nationality	ETR A	ETR B	ETR C	ETR D	ETR E	ETR F	MIRU 10	MIRU 16	Miru26	Miru39	Miru40	QUB11
1	ALZ 0043	Alborz	F	Iranian	2	1	4	2	3	2.2	2	3.2	6	2	2	2
2	ALZ 0038	Alborz	F	Iranian	2	1	4	3	3	2.2	2	3.2	6	2	2	2
3	ALZ 0045	Alborz	M	Iranian	2	1	4	3	3	2.2	2	4.2	6	2	2	2
4	ALZ 0040	Alborz	M	Iranian	2	2	4	3	3	2.3	3	2.2	6	1	1	3
5	ALZ 0033	Alborz	M	Iranian	3	1	3	3	3	2.2	2	3.2	6	1	2	2
6	ALZ 0042	Alborz	M	Iranian	3	1	4	3	3	2.1	2	3.2	5	2	2	2
7	ALZ 0036	Alborz	F	Iranian	3	1	4	3	3	2.2	2	3.2	5	2	2	2
8	ALZ 0034	Alborz	M	Iranian	3	1	4	3	3	2.2	2	3.2	6	2	2	2
9	ALZ 0046	Alborz	M	Iranian	3	1	4	3	3	2.2	2	4.2	6	2	2	2
10	ALZ 0041	Alborz	F	Iranian	3	1	4	3	4	2.2	2	2.2	5	2	2	2
11	ALZ 0031	Alborz	F	Iranian	3	1	4	3	4	2.2	2	3.2	5	1	3	2
12	ALZ 0039	Alborz	M	Iranian	3	1	5	3	3	1.2	2	3.2	5	2	2	2
13	ALZ 0050	Alborz	M	Iranian	3	1	5	3	4	2.2	2	2.2	5	2	2	2
14	ALZ 0049	Alborz	M	Iranian	3	2	3	3	5	2.3	5	4.2	7	3	2	2
15	ALZ 0035	Alborz	M	Iranian	3	2	4	3	2	2.3	3	1.2	5	2	3	3
16	ALZ 0032	Alborz	M	Iranian	4	2	2	3	5	2.3	6	5.2	7	3	2	1
17	ALZ 0047	Alborz	M	Iranian	4	2	3	3	5	2.3	5	4.2	7	2	2	2
	No of alleles				3	2	4	2	3	4	4	5	3	3	3	3
	Nei's diversity index				0.51	0.33	0.53	0.11	0.59	0.56	0.47	0.69	0.63	0.46	0.30	0.30
	Hunter-Guston (HG) diversity index				1											
	SLV (N=9)				1&2, 2&3, 2&8, 3&9, 5&7, 6&7, 7&8, 8&9, 10&13											
	DLV (N=13)				1&3, 1&8, 6&8, 6&10, 6&12, 2&7, 2&9, 3&8, 5&8, 7&9, 7&10, 7&12, 14&17											
	TLV (N=17)				1&7, 1&9, 2&5, 2&6, 2&10, 3&7, 5&9, 5&11, 6&9, 7&11, 7&13, 8&10, 8&12, 9&10, 10&11, 11&13, 12&13											

on population genetics of MTC in Alborz province using MIRU-VNTR strategy. Detection of 17 distinct genotypes among the 17 isolates in the study panel displayed a high proportion of non-clustered and orphan isolates. A trivial explanation for this observation could be that there is a lack of previous data on genetic structure of MTC in the region. In fact, as shown by the published literature, in populations where extensive genetic studies have been already conducted, observing a lower number of orphan isolates is largely frequent (14, 15). When the MLVA-VNTR results from a further 93 isolates collected in Khorasan Razavi Province as well as 53 isolates collected in Markazi province combined with the results of present work, a number of clustered isolates including those from Alborz were observed. This is a supporting evidence on the fact that higher frequency of orphan genotypes in this study is a consequence of less-analyzed genetics of MTC in the region.

Using the "n-1" method on the study panel to calculate Recent Transmission Rate (RTR) was in support of no association between TB cases and recent transmission, a likely indication of association with latent tuberculosis reactivation. This observation seems to be not unusual in Iran as the overall clustering rate reported by Iranian workers ranged 0.0%-75% (6, 16-18). A variety of elements including spatial and temporal dimensions of study, size of examined isolates, ethnicity coverage and socioeconomic status of enrolled patients have been noted as contributors (6).

No sign of multiple alleles was found with any of the loci or isolates under examination in the present study though previous works have displayed the existence of this phenomenon among local populations of MTC in the Iranian environment with a frequency ranging from 1.6% to 26.6% (5, 6, 19, 20). In-patient independent microevolution of discrete MLVA-VN-
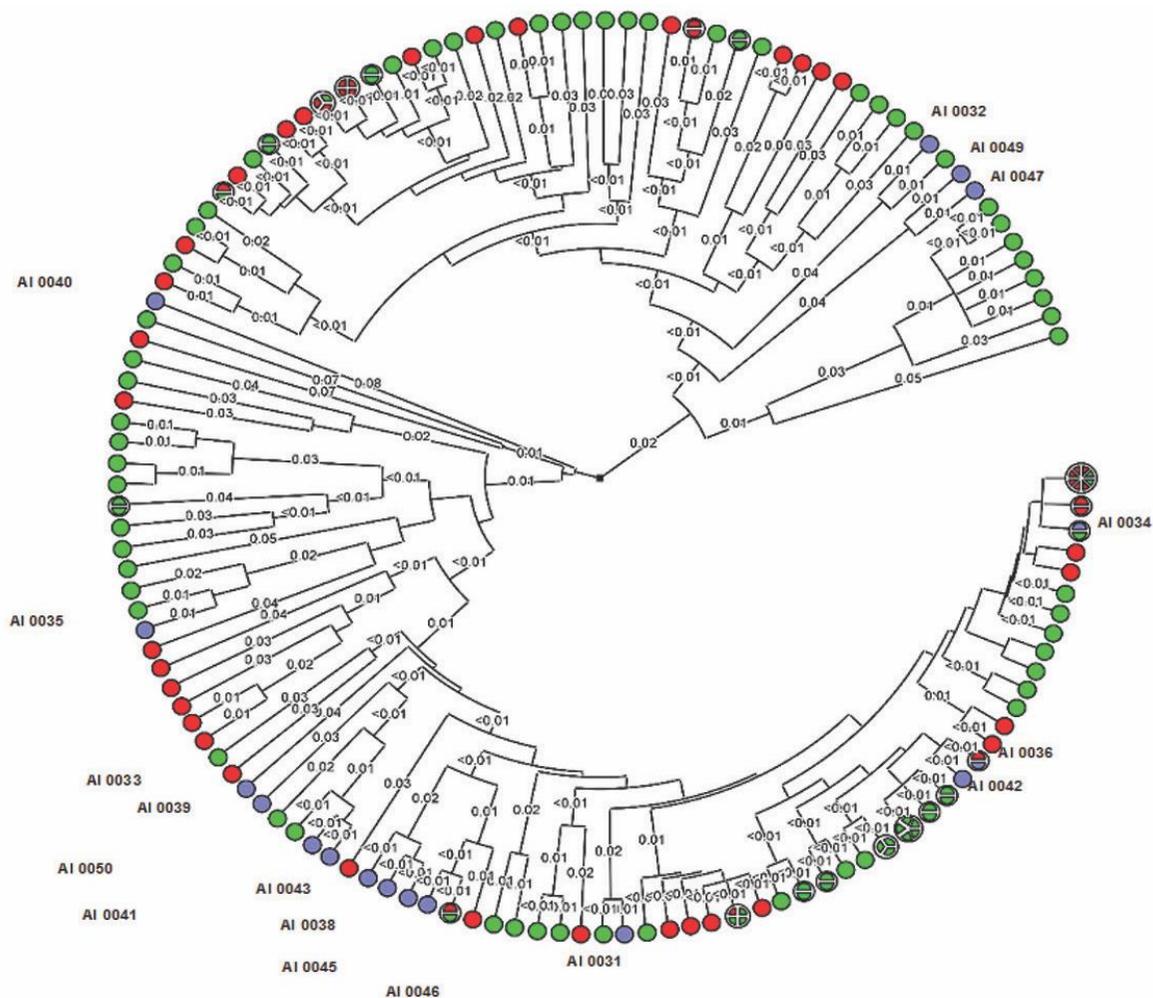
Fig. 1. Circular dendrogram displaying a cluster analysis of Iranian *M. tuberculosis* isolates from Ahmadi (N=53) and Ravansalar (N=94) studies plus the current work (N=17) based on the 12-locus MLVA. The province sources of the strains are indicated with colors, where green, red and turquoise represent Markazi, Khorasan Razavi and Alborz provinces, respectively.

TR loci and co-infection with multiple-strains are believed to be the two most relevant explanations for this observation which were not observed in the present work (6).

To visualize the evolutionary relationships of examined isolates along with those from previous studies (Khorasan Razavi province, N=93 and Markazi province, N=53) that used the same MLVA-VNTR genotyping system, a minimum spanning tree and a dendrogram were constructed (Figs. 1 and 2). A number of clonal structures were identified with Alborz isolates distributed across them showing a large genetic heterogeneity, forming several distinct complexes. Furthermore, as displayed by the circular dendrogram drawn using UPGMA method, most of

the Alborz isolates scattered through the genotypes which again reiterate on heterogeneity of studied MTC population in the region.

While MIRU-VNTR still remains a broadly used protocol for genotyping at global scale, with the advent of the enabling technology of next-generation sequencing, the whole-genome sequencing (WGS) is now available to TB reference laboratories though mostly in the developed world. The currently high costs and complexity of data analysis in connection with WGS has limited its broad acceptance. Despite delivering greater specificity, the quantitative prospect comparisons in favor of WGS are yet to be completed (21). In reality, as the running costs gradually decline, WGS will very likely replace oth-
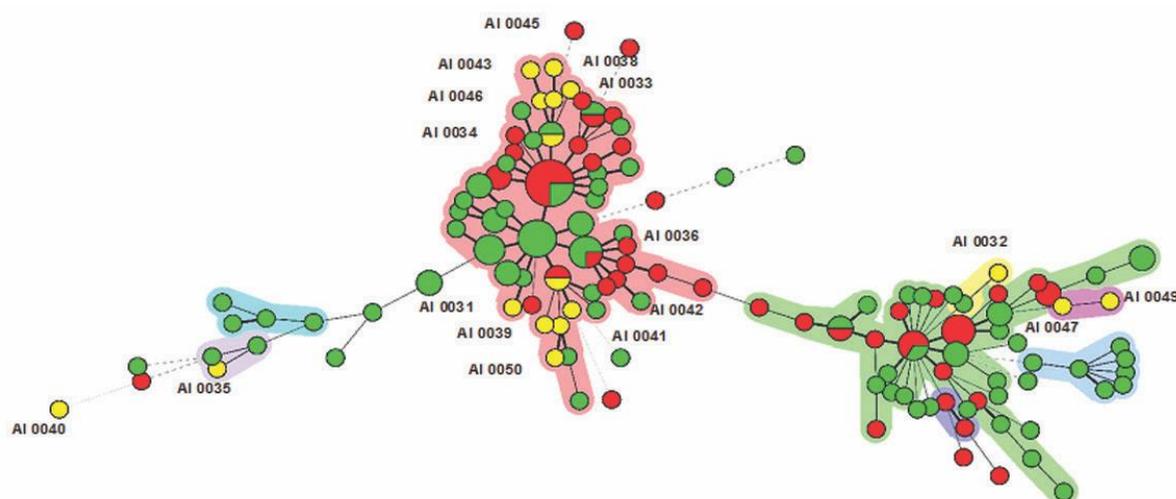
Fig. 2. Minimum spanning tree of 164 Iranian *M. tuberculosis* isolates from Ahmadi (N=53) and Ravansalar (N=94) studies plus the current work (N=17) based on the 12-locus MLVA. Single circles denote MLVA types with the Alborz types indicated beside the corresponding circle. Frequency of the MLVA types is represented by the diameter of circles. The different colors of the circles indicate the geographical origin of isolates. The allelic differences between the MLVA types are shown by the thickness of the lines connecting the circles where thick lines indicate that a given MLVA type (circle) is an SLV to its immediate neighboring circle. For colored partitioning, nodes closer than 2 have been put into the same partition and every partition contains at least 2 entries.

er genotyping methods in the near future. If open-source software with capability to connect WGS and MIRU-VNTR typing (so-called MIRU-profiler programs) become available then the prediction of MIRU-VNTR profiles from WGS of *M. tuberculosis* isolates will be very convenient (22). In conclusion, the present work provides baseline data on population genetic of *M. tuberculosis* strains that are operating in Alborz, a step forward for further molecular epidemiology works at local as well as national level.

REFERENCES

1. Mendis C, Thevanesam V, Kumara A, Wickramasinghe S, Madagedara D, Gamage C, et al. Insight into genetic diversity of *Mycobacterium tuberculosis* in Kandy, Sri Lanka reveals predominance of the Euro-American lineage. *Int J Infect Dis* 2019;87:84-91.

2. Dos Vultos T, Mestre O, Rauzier J, Golec M, Rastogi N, Rasolofo V, et al. Evolution and diversity of clonal bacteria: the paradigm of *Mycobacterium tuberculosis. PLoS One* 2008;3(2):e1538.

3. Gagneux S. Ecology and evolution of *Mycobacterium tuberculosis. Nat Rev Microbiol* 2018;16:202-213.

4. Kargarpour Kamakoli M, Hadifar S, Khanipour S, Farmanfarmaei G, Fateh A, Siadat SD, et al. Comparison of MIRU-VNTR genotyping between old and fresh clinical samples in tuberculosis. *Infect Dis (Lond)* 2019;51:659-667.

5. Hadifar S, Shamkhali L, Kargarpour Kamakoli M, Mostafaei S, Khanipour S, Mansoori N, et al. Genetic diversity of *Mycobacterium tuberculosis* isolates causing pulmonary and extrapulmonary tuberculosis in the capital of Iran. *Mol Phylogenet Evol* 2019;132:46-52.

6. Mansoori N, Yaseri M, Vaziri F, Douraghi M. Genetic diversity of *Mycobacterium tuberculosis* complex isolates circulating in an area with high tuberculosis incidence: Using 24-locus MIRU-VNTR method. *Tuberculosis (Edinb)* 2018;112:89-97.

7. Riyahi Zaniani F, Moghim S, Mirhendi H, Ghasemian

Safaei H, Fazeli H, Salehi M, et al. Genetic lineages of *Mycobacterium tuberculosis* i in Isfahan, Iran. *Curr Microbiol* 2017;74:14-21.

8. Huard RC, de Oliveira Lazzarini LC, Butler WR, van Soolingen D, Ho JL. PCR-based method to differentiate the subspecies of the *Mycobacterium tuberculosis* complex on the basis of genomic deletions. *J Clin Microbiol* 2003;41:1637-1650.

9. McHugh TD, Gillespie SH. Nonrandom association of IS*6110* and *Mycobacterium tuberculosis*: implications for molecular epidemiological studies. *J Clin Microbiol* 1998;36:1410-1413.

10. Warren R, Gey van Pittius N, Barnard M, Hesseling A, Engelke E, De Kock M, et al. Differentiation of *Mycobacterium tuberculosis* complex by PCR amplification of genomic regions of difference. *Int J Tuberc Lung Dis* 2006;10:818-822.

11. 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.

12. Salvato RS, Schiefelbein S, Barcellos RB. Molecular characterisation of multidrug-resistant *Mycobacterium tuberculosis* isolates from a high-burden tuberculosis state in Brazil. *Epidemiol Infect* 2019;147: e216.

13. 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;330:1703-1709.

14. Choisy M, Nguyen DH, Tran THT, Pham KLT, Dinh PTT, Philippe J, et al. High prevalence of Beijing and EAI4-VNM genotypes among *M. tuberculosis* isolates in northern Vietnam: sampling effect, rural and urban disparities. *PLoS One* 2012;7(9):e45553.

15. Somphavong S, Berland JL, Gauthier M, Vu TT, Nguyen QH, Iem V, et al. First insights into the genetic characteristics and drug resistance of *Mycobacterium tuberculosis* population collected during the first national tuberculosis prevalence survey of Lao PDR (2010-2011). *BMC Infect Dis* 2019;19:851.

16. Asgharzadeh M, Kafil HS, Roudsary AA, Hanifi GR. Tuberculosis transmission in Northwest of Iran: using MIRU-VNTR, ETR-VNTR and IS6110-RFLP methods. *Infect Genet Evol* 2011;11:124-131.

17. Asgharzadeh M, Samadi Kafil H, Pourostadi M, Asadi Faezi N, Rashedi J, Mahdavipour B. Strain differentiation of *Mycobacterium tuberculosis* for epidemiology in northwest of Iran. *Cell Mol Biol (Noisy-le-grand)* 2016;62:15-20.

18. Zamani S, Aflaki M, Fooladi AA, Darban-Sarokhalil D, Bameri Z, Khazaee S, et al. MIRU-VNTR analysis of the *Mycobacterium tuberculosis* isolates from three provinces of Iran. *Scand J Infect Dis* 2013;45:124-130.

19. Farmanfarmaei G, Kargarpour Kamakoli M, Sadegh HR, Masoumi M, Abdolrahimi F, Fateh A, et al. Bias in detection of *Mycobacterium tuberculosis* polyclonal infection: Use clinical samples or cultures? *Mol Cell Probes* 2017;33:1-3.

20. Jafarian M, Aghali-Merza M, Farnia P, Ahmadi M, Masjedi MR, Velayati AA. Synchronous comparison of *Mycobacterium tuberculosis* epidemiology strains by "MIRU-VNTR" and "MIRU-VNTR and Spoligotyping" technique. *Avicenna J Med Biotechnol* 2010; 2: 145-152.

21. Wyllie DH, Davidson JA, Grace Smith E, Rathod P, Crook DW, Peto TEA, et al. A quantitative evaluation of MIRU-VNTR typing against whole-genome sequencing for identifying *Mycobacterium tuberculosis* transmission: A prospective observational cohort study. *EBioMedicine* 2018;34:122-130.

22. Rajwani R, Shehzad S, Siu GKH. MIRU-profiler: a rapid tool for determination of 24-loci MIRU-VNTR profiles from assembled genomes of *Mycobacterium tuberculosis*. *PeerJ* 2018;6:e5090.