Mashhad University of Medical Sciences (MUMS)

Reviews in Clinical Medicine

Clinical Research Development Center Ghaem Hospital

Genetic Diversity of Clinical Isolates of Mycobacterium Tuberculosis in Northeast of Iran

Samane Saeidi[1], Kiarash Ghazvini[1], Masoud Youssefi[1], Hadi Safdar[1], Saman Soleimanpour[1], Parviz Afrough[1], Amir Azimian[2], Hamid Solgi[3], Masoud Keikha[4*]

[1.] *Depatment of Microbiology and Virology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran*
[2.] *Microbiology and Virology Research Center, Department of Bojnord University of Medical Sciences, Bojnord, Iran*
[3.] *Department of Mycobacteriology & Amp, Pulmonary Research, Pasteur Institue of Iran, Tehran, Iran*
[4.] *Department of Microbiology and Virology, Faculty of Medicine, Iranshahr University of Medicine Sciences, Iranshahr, Iran*

ARTICLE INFO

Article type
Original article

Article history
Received: 03 Jan 2024
Revised: 11 Jan 2024
Accepted: 12 Mar 2024

Keywords
KatG
Mycobacterium tuberculosis
Multi-locus sequence analysis
Molecular epidemiology
Rpsl
Tuberculosis

ABSTRACT

Introduction: One of the most hazardous infections for humans is still *Mycobacterium tuberculosis*. Finding the connections between various clinical strains has been a top focus in epidemiological research.
Methods: In this work, we created a very strong phylogeny of Mycobacterium TB using multilocus sequence analysis, or MLSA. To identify polymorphic nucleotide sites, five gene fragments from the Rpsl (302 bp), MprA (559 bp), LipR (322 bp), KatG (488 bp), and Fgd1 (266 bp) regions were subjected to MLSA, a single nucleotide polymorphism (SNP)-based technique. The Hunter-Gaston Index (HGI) was used to measure the discriminatory power of each locus for all genes.
Results: Each distinct allelic profile in this investigation was given a sequence type (ST) number; nine sequence types from a total of twenty strains were found, suggesting a significant degree of strain diversity in this region.
Conclusion: Our results showed that the presence of high genetic diversity among clinical isolates of *M. tuberculosis* in Northeast of Iran. There is no evidence for recent transmission.

Please cite this paper as:
Saeidi S, Ghazvini K, youssefi M, Safdar H, Soleimanpour S, Parviz Afrough P, Azimian A, Solgi H, Keikha M. Genetic Diversity of Clinical Isolates of Mycobacterium Tuberculosis in Northeast of Iran. Rev Clin Med. 2024;11(1): 8-12.

Introduction

Approximately one-third of the global population is infected with *Mycobacterium tuberculosis* (*M. tuberculosis*), the causative agent of tuberculosis (TB), which results in approximately 11 million TB infections and 1.6 million deaths annually [1–2]. *M. tuberculosis* is one of the most successful human pathogens. The public's health greatly benefits from the identification of the connections between various clinical strains of *Mtb* [3–4]. Although, IS6110 RFLP is known as gold standard of molecular epidemiology studies of tuberculosis; but this technique has serious limitation such as time-consuming, expensive and cannot differentiate the totally *Mtb* strains (particularly Beijing strains) [3-6]. Moreover; the spoligotyping and MIRU-VNTR are standard methodologies for *M. tuberculosis* genetic epidemiology, but these methods have low-discriminatory powers in various studies; recent studies suggest that SNPs are advantageous in phylogenetics [5-7]. It is dependable and easy to understand when the nucleotide sequences of many genes from several strains are determined. SNP-based multilocus sequence analysis (MLSA) proved

***Corresponding author:** Masoud Keikha,*
Department of Microbiology and Virology, Faculty of Medicine, Iranshahr University of Medicine Sciences, Iranshahr, Iran
***E-mail:** Masoud.keykha90@gmail.com*
***Tel:** 09386836425*

a potent technique for phylogenetic analysis of closely related strains [8]. Genetic diversity is low in *M. tuberculosis* compared to several other bacterial species [9–11].

In order to achieve the epidemiological relationships of clinical strains in phylogenetic studies, we studied five gene fragments from the *Rpsl (*30S ribosomal subunit protein, 302 bp), *MprA* (Response regulator *MprA*, 559bp), *LipR* (Acetyl-hydrolase, 322bp), *KatG* (Catalase-peroxidase, 488bp) and *Fgd1* (Oxidoreductase, 266 bp), as common genes in all *M. tuberculosis* strains for the first time in the world.

Materials and Method

Bacterial isolates

In 2017, the Regional Reference Tuberculosis Laboratory in northeastern Iran identified twenty clinical isolates of *Mycobacterium tuberculosis* from patients suffering from pulmonary tuberculosis. This study also used H37RV (ATCC27294) as a reference strain. The sputum samples were inoculated on the Lowenstein-Jensen media after being decontaminated using the conventional Petroff's procedure [12]. For the purpose of confirming the species, the positive cultures underwent niacin testing and PCR for the IS6110 insertion sequence. Using the CTAB technique, DNA was isolated from *M. tuberculosis* isolates [13].

Gene selection

During the gene selection phase, all *M. tuberculosis* strains accessible in the NCBI's genome sequences were aligned using MEGA software [13–14] for each gene in order to identify SNPs. Polymorphic nucleotide sites were found using the genome sequence H37RV (NC 000962.3) as a reference. The Hunter-Gaston Index (HGI) was used to quantify the discriminatory strength of each locus for all isolates [15–16]. The genes must be a single copy throughout the whole genome and cannot be prone to horizontal gene transfer, according to the selection criterion [17]. Finally, the five genes (*rpsl, lipR, fgd1, mprA, katG*) were chosen for this study according to these criteria**.**

PCR amplification and sequencing of the genes

Primer3 [18] was used to generate primers for amplifying these genes, which are given in Table 1. With an initial denaturation at 95°C for 5 minutes, subsequent denaturation at 94°C for 45 seconds, annealing at 56°C–64°C for 45 seconds, depending on the primers, elongation at 72°C for 45 seconds, and a final extension at 72°C for 5 minutes, thermal cycling was done in 35 cycles. Next, the ABI 3730XL DNA Analyzer was used to sequence DNA.

Results & Discussion

BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and multiple alignments with the reference sequence H37Rv, which is accessible in the NCBI (GenBank: NC 000962.3), were used to compare the sequencing findings [18–19]. With the use of Mega software, the SNPs for every gene were identified; Table 2 displays the findings. Ultimately, two models—one character-based and the other distance-based—were compared in the analysis of the phylogeny and sequencing data. Following many alignments for each gene, separate alleles were found using the character-based technique, and the combination of the five MLSA allelic numbers was recognized as a distinct multilocus sequence type (ST). The distance matrix calculates the "genetic distance" for each gene between the sequences. Originally applied to phenetic data, this approach was employed in phylogeny as a non-parametric distance method that required an MSA (multiple sequence alignment) as an input. It used a matrix of pairwise distances. Using the sequence query set, distance techniques try to create an all-to-all matrix that describes the separation between each pair of sequences [19]. Finally, the Vector NTI program was used to examine SNP data, including sSNPs (synonymous) and nsSNPs

Table 1. Sequences of primers used in PCR and SNP analysis

Genes	Primers sequences	Annealing (C)	Size (bp)
Rpsl (30s ribosomal protien S12)	F:5'- TCGGGACAAGATCAGTAAGG-3' R:5'- GCCTGTTTGCGGTTCTTG-3'	47	390
MprA (Mycobacterial persistence regulator MprA)	F:5'- CGACATGATTGCCAGCG-3' R:5'- GTAGCACGTAACCCACTCCG-3'	48	188
LipR (Putative acetyl hydrolase LipR)	F:5'- GCTCATTCTTCCGTCGATGT-3' R:5'- CTGCGATTGGTGGAGGTG-3'	48	246
KatG (Catalase-peroxidase-peroxynitritase)	F:5'- CCAGCGGCCCAAGGTATC-3' R:5'-GCGGTCACACTTTCGGTAAG-3'	49	599
Fgd1 (F420-dependent glucose-6-phosphate dehydrogenase)	F:5'- AGCTGGCATTGAACAACACC-3' R:5'- CAGGTCCGACTGGAAGAGC-3'	49	549

Table 2. Allelic diversity and HGI indexes of each locus

Gene name	HGI	Variable allele	Number of strains
Rpsl	0.94	(G) 128	3
LipR	0.98	(C) 43 ,72	10
Fgd1	0.95	(C) 960	9
katG	0.95	(G) 1280 ,1212	4
MprA	0.99	(T) 399 ,621	5

(non-synonymous). SNP data analysis shows 31 variable SNP sites among these strains.

A sequence type (ST) number was assigned to each unique allelic profile and finally nine STs were identified from 20 strains (Table 3).

For each gene, distance matrix was determined as the distances of each strain in the total number of strains, and the final distance-based for 5 genes obtained, by taking the average of total distance matrix (Figure 1). These results provide the most accurate evolutionary distance.

Following character-based phylogeny, the ST dendrogram was computed and compared to the evolutionary distance matrix.

When two strains are placed in one ST type, they have a minimum difference, and their associated distance matrix is also zero, showing that two phylogenic methods are changing in the same direction.

There are 31 variable SNP sites among 20 isolates, these polymorphisms comprised of 6 nsSNPs, 25 sSNPs. In *Rpsl,* three nsSNP, in *KatG* three nsSNP and 1 sSNP, in *MprA* five sSNP, in *LipR* ten sSNP and in Fgd_1 nine sSNP were found. In our study, sSNPs were much more abundant than nsSNPs. The mutation in codon 128 of *rpsl* led to changes of lysine to arginine, and the mutation in codon 1280 of *katG* led to changes of alanine to glycine.

Table 3. Results of sequencing and allele counts of different STs in clinically isolates of *M. tuberculosis*

Code	*rpsl*	*katG*	*lipR*	*mprA*	*Fgd*	Sequence type
1	1	2	1	1	1	ST1
2	1	1	2	2	2	ST2
3	1	1	2	2	2	ST2
4	1	2	2	3	2	ST3
5	1	1	1	2	2	ST4
6	1	1	2	1	2	ST5
7	1	1	2	1	2	ST5
8	1	1	2	1	2	ST5
9	1	1	2	1	1	ST6
10	2	2	3	1	2	ST7
11	2	2	2	1	2	ST8
12	1	1	1	1	1	ST9
13	1	1	1	1	1	ST9
14	1	1	1	1	1	ST9
15	1	1	1	1	1	ST9
16	1	1	1	1	1	ST9
17	1	1	1	1	1	ST9
18	1	1	1	1	1	ST9
19	1	1	1	1	1	ST9
20	1	1	1	1	1	ST9

Thousands of SNPs have been found in clinical isolates of Mycobacterium TB as a result of recent improvements in DNA sequencing. Our knowledge of the distinctions and evolutionary links across strains has been altered by this genetic diversity. A comprehensive database including SNP data is still lacking, despite the significance of SNPs for our comprehension of the variety of *M. tuberculosis* populations. We identified 31 variable SNP locations in this investigation. This result is consistent with the Gagneux study's outcome, which demonstrated that SNPs are reliable markers for strain categorization and phylogeny inference [20–21]. The majority of SNPs in *M. tuberculosis* are

	ID1	ID2	ID3	ID5	ID6	ID7	ID8	ID9	ID10	ID11	ID12	ID13	ID14	ID15	ID16	ID17	ID18	ID19	ID21	ID23
ID1	0																			
ID2	0.0006	0																		
ID3	0	0.0006	0																	
ID5	0	0.0006	0	0																
ID6	0.0022	0.0028	0.0022	0.0022	0															
ID7	0	0.0006	0	0	0.0022	0														
ID8	0.0032	0.0036	0.0032	0.0022	0.0018	0.0032	0													
ID9	0.004	0.0044	0.004	0.004	0.0018	0.004	0.0008	0												
ID10	0.0014	0.002	0.0014	0.0014	0.0008	0.0014	0.0026	0.0026	0											
ID11	0.0008	0.0014	0.0008	0.0008	0.0014	0.0008	0.0024	0.0032	0.0024	0										
ID12	0.0018	0.0024	0.0018	0.0018	0.0004	0.0018	0.0014	0.0022	0.0012	0.001	0									
ID13	0	0.0006	0	0	0.0024	0	0.0032	0.004	0.0014	0.0008	0.0018	0								
ID14	0	0.0006	0	0	0.0024	0	0.0032	0.004	0.0014	0.0008	0.0018	0	0							
ID15	0.0032	0.0036	0.0032	0.0032	0.0014	0.0036	0	0.0008	0.003	0.0028	0.0014	0.0032	0.0032	0						
ID16	0.0024	0.0028	0.0024	0.0024	0	0.0024	0.0018	0.0018	0.0008	0.0014	0.0004	0.0024	0.0024	0.0018	0					
ID17	0.0018	0.0024	0.0018	0.0018	0.0004	0.0018	0.0014	0.0022	0.0012	0.001	0	0.0018	0.0018	0.0014	0.0004	0				
ID18	0	0.0006	0	0	0.0024	0	0.0032	0.004	0.0014	0.0008	0.0018	0	0	0.0032	0.0022	0.0018	0			
ID19	0	0.0006	0	0	0.0024	0	0.0032	0.004	0.0014	0.0008	0.0018	0	0	0.0032	0.0022	0.0018	0	0		
ID21	0.0018	0.0024	0.0018	0.0018	0.0004	0.0018	0.0014	0.0022	0.0012	0.001	0	0.0018	0.0018	0.0014	0.0004	0	0.0018	0.0018	0	
ID23	0	0.0006	0	0	0.0024	0	0.0032	0.004	0.0014	0.0008	0.0018	0	0	0.0032	0.0024	0.0018	0	0	0.0018	0

Figure 1. Pairwise genetic distances based on the five genes nucleotide polymorphism among M. tuberculosis strains

found in genomic regions that code for proteins. sSNP and nsSNP are two further classifications for coding SNPs based on whether or not they alter the sequence of amino acids [22–23]. There were six nsSNPs and twenty-five sSNPs in the polymorphisms in this study. Our strains were categorized into nine sequence types (ST) using a distance matrix and a character-based phylogenetic dendrogram. This suggests that there are a large variety of strains in this region of the nation. Between strain 9 (containing Mtb1,2,3,5,7,11,1,19) and strain 8 (containing Mtb2,3,7,14), the greatest evolutionary gap was found. Additionally, if a mutation rate is known, SNPs may be used to calculate the evolutionary distances between strains and, if genetic distance is known, to estimate the period of strain divergence [24]. Non-sequence-based technologies like spoligotyping and VNTR-based approaches have established the gold standard for regular genotyping of these species because typical sequence-based genotyping like MLST is not suitable for these bacteria [25]. But given the intrinsic phylogenetic constraints of these techniques and the results shown here, it is not recommended to utilize them for evolutionary research. Instead, the final phylogeny should be based on more reliable markers, such as SNPs for DNA sequence data. However, several models of molecular evolution have been created from empirical data, and a strong statistical framework has been established to assess these models' suitability for phylogenetic connection inference [26]. *M. tuberculosis* strains were shown to have a reduced rpsl phylogenetic capacity than other strains based on partial gene sequence analysis of rpsl, lipR, katG, fgd1, and mprA. Based on the SNP data shown in table 3, 10 strains belonged to ST9, 3 strains belonged to ST5, 2 strains belonged to ST2 and only one isolate belonged to other sequence types. In Bing *et al*. study, 11 sequence types (STs) among 44 *M. tuberculosis* strains were discovered [25]. In their 108 strains of 89 genes, Hershberg et al. found 488 SNPs in total. These genes included those encoding a novel bacterial secretion mechanism, DNA repair genes, novel antigens, and housekeeping genes and antigens [26]. They were also genes of interest as potential new therapeutic targets.

Conclusion

There is serious challenging with molecular epidemiology of *M. tuberculosis* strains; However, SNPs analysis can reliable technique for molecular epidemiology studies of circulating strains within the geographic regions and can be applied in the regional control disease plans.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, Pulcini C, Kahlmeter G, Kluytmans J, Carmeli Y, Ouellette M, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect Dis*. 2018;18(3):318-27.
2. Bagcchi S. WHO's global tuberculosis report 2022. *The Lancet Microbe*. 2023;4(1):e20.
3. Keikha M, Esfahani BN. The relationship between tuberculosis and lung cancer. *Adv Biomed Res*. 2018;7:58.
4. Lindstedt BA. Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis*. 2005;26(13):2567-82.
5. 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(11):e7815.
6. Keikha M. The necessity of conducting studies for mycobacterial interspersed repetitive-unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *Egypt J Chest Dis Tuberc*. 2018;67(4):479.
7. Lopes JS, Marques I, Soares P, Nebenzahl-Guimaraes H, Costa J, Miranda A, Duarte R, Alves A, Macedo R, Duarte TA, Barbosa T, et al. SNP typing reveals similarity in *Mycobacterium tuberculosis* genetic diversity between Portugal and Northeast Brazil. Infection, Genetics and Evolution. 2013;18:238-46.
8. Vinuesa P, 2010. Multilocus sequence analysis and bacterial species phylogeny estimation. In: Oren A, Papke RT, eds. Molecular Phylogeny of Microorganisms. Norfolk, UK: Caister Academic Press, 41-64.
9. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, DeBoy R, Dodson R, Gwinn M, Haft D, Hickey E. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol*. 2002;184(19):5479-90.
10. Hughes AL, Friedman R, Murray M. Genomewide pattern of synonymous nucleotide substitution in two complete genomes of *Mycobacterium tuberculosis*. *Emerg Infect Dis*. 2002;8(11):1342.
11. Kapur V, Whittam TS, Musser JM. Is *Mycobacterium tuberculosis* 15,000 years old?. J Infect Dis. 1994;170(5):1348-9.
12. Petroff SA. A new and rapid method for the isolation and cultivation of tubercle bacilli directly from the sputum and feces. *J Exp Med*. 1915;21(1):38-42.
13. van Soolingen DI, Hermans PW, De Haas PE, Soll DR, Van Embden JD. 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(11):2578-86.
14. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28(10):2731-9.
15. 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(11):2465-6.
16. Hunter PR. Reproducibility and indices of discriminatory power of microbial typing methods. *J Clin Microbiol*. 1990;28(9):1903-5.
17. Adékambi T, Butler RW, Hanrahan F, Delcher AL, Drancourt M, Shinnick TM, et al. Core gene set as the basis of multilocus sequence analysis of the subclass *Actinobacteridae*. *PLoS One*. 2011;6(3):e14792.
18. S Rozen, H Skaletsky. Primer3 on the WWW for general

users and for biologist programmers S Krawetz, S Misener (Eds.), Bioinformatics methods and protocols: methods in molecular biology, Humana Press, Totowa, NJ (2000), pp. 365-386.
19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-10.
20. Mount, D. W. Bioinformatics: Sequence and Genome Analysis (Cold Spring Harbor Laboratory Press, New York, 2001).
21. Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. Lancet Infect Dis. 2007;7(5):328-37.
22. Hershberg R, Petrov DA. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet*. 2010;6(9):e1001115.
23. Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, Mohaideen N, Ioerger TR, Sacchettini JC, Lipsitch M, Flynn JL,et al. Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat Genet*. 2011;43(5):482.
24. Mathema B, Kurepina NE, Bifani PJ, Kreiswirth BN. Molecular epidemiology of tuberculosis: current insights. *Clin Microbiol Rev*. 2006;19(4):658-85.
25. Bing LU, Dong HY, Zhao XQ, Liu ZG, Liu HC, Zhang YY, Jiang Y, Wan KL, et al. A new multilocus sequence analysis scheme for *Mycobacterium tuberculosis*. *Biomedical Environ Sci*. 2012;25(6):620-9.
26. Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW, Gagneux S, et al. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol*. 2008;6(12):e311.