

HOSTED BY



ELSEVIER

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/IJMYCO



Characterization of genomic variations in SNPs of PE_PGRS genes reveals deletions and insertions in extensively drug resistant (XDR) *M. tuberculosis* strains from Pakistan

Akbar Kanji ^a, Zahra Hasan ^{a,*}, Asho Ali ^a, Ruth McNerney ^b, Kim Mallard ^b, Francesc Coll ^b, Grant Hill-Cawthorne ^{c,d}, Mridul Nair ^c, Taane G. Clark ^b, Ambreen Zaver ^a, Sana Jafri ^a, Rumina Hasan ^a

^a Aga Khan University, Karachi, Pakistan

^b London School of Hygiene and Tropical Medicine (LSHTM), United Kingdom

^c Pathogen Genomics Laboratory, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

^d Sydney Emerging Infections and Biosecurity Institute and School of Public Health, Sydney Medical School, University of Sydney, NSW 2006, Australia

ARTICLE INFO

Article history:

Received 27 October 2014

Accepted 2 November 2014

Available online 21 January 2015

Keywords:

Whole genome sequencing

Mycobacterium tuberculosis

PE_PGRS genes

ABSTRACT

Background: *Mycobacterium tuberculosis* (MTB) PE_PGRS genes belong to the PE multigene family. Although the function of PE_PGRS genes is unknown, it is hypothesized that the PE_PGRS genes may be associated with antigenic variability in MTB.

Material and methods: Whole genome sequencing analysis was performed on ($n = 37$) extensively drug-resistant (XDR) MTB strains from Pakistan, which included Lineage 1 (East African Indian, $n = 2$); Other lineage 1 ($n = 3$); Lineage 3 (Central Asian, $n = 24$); Other lineage 3 ($n = 4$); Lineage 4 (X3, $n = 1$) and T group ($n = 3$) MTB strains.

Results: There were 107 SNPs identified from the analysis of 42 PE_PGRS genes; of these, 13 were non-synonymous SNPs (nsSNPs). The nsSNPs identified in PE_PGRS genes – 6, 9 and 10 – were common in all EAI, CAS, Other lineages (1 and 3), T1 and X3. Deletions (DELS) in PE_PGRS genes – 3 and 19 – were observed in 17 (80.9%) CAS1 and 6 (85.7%) in Other lineages (1 and 3) XDR MTB strains, while DELS in the PE_PGRS49 were observed in all CAS1, CAS, CAS2 and Other lineages (1 and 3) XDR MTB strains. All CAS, EAI and Other lineages (1 and 3) strains showed insertions (INS) in PE_PGRS6 gene, while INS in the PE_PGRS genes 19 and 33 were observed in 20 (95.2%) CAS1, all CAS, CAS2, EAI and Other lineages (1 and 3) XDR MTB strains.

Conclusion: Genetic diversity in PE_PGRS genes contributes to antigenic variability and may result in increased immunogenicity of strains. This is the first study identifying variations in nsSNPs and INDELS in the PE_PGRS genes of XDR-TB strains from Pakistan. It highlights common genetic variations which may contribute to persistence.

© 2015 Asian African Society for Mycobacteriology. Published by Elsevier Ltd. All rights reserved.

* Corresponding author at: Department of Pathology and Microbiology, The Aga Khan University, Karachi, Pakistan. Tel.: +92 21 3486 4527.

E-mail address: zahra.hasan@aku.edu (Z. Hasa).

Peer review under responsibility of Asian African Society for Mycobacteriology.

<http://dx.doi.org/10.1016/j.ijmyco.2014.11.049>

2212-5531/© 2015 Asian African Society for Mycobacteriology. Published by Elsevier Ltd. All rights reserved.

Introduction

Mycobacterium tuberculosis (MTB), the causative agents of tuberculosis (TB) kills approximately 2 million people worldwide each year [1]. The control of TB requires a better understanding of the mechanisms that allow MTB to evade the immune system and remain persistent with the host. The sequencing of the genomes of MTB strains has provided important insights into possible mechanisms of persistence within the host, along with the discovery of ~60 genes named PE_PGRS (Proline glutamic acid polymorphic GC rich repetitive sequence), which belong to the PE multigene family [2]. These multigene families represent approximately 10% of the coding capacity of the MTB genome and are characterized by their high GC content and repetitive sequences. This may be due to the high frequency of gene duplication, recombination and strand slippage mechanisms [2,3].

The single nucleotide polymorphisms (SNPs) have been shown to be the most common form of genetic variation in MTB complex (MTBC). The SNPs have been shown to be of low frequency and limited horizontal gene transfer (HGT), which have resulted in low levels of independent occurrence of the same SNP in phylogenetically unrelated strains [4,5]. Depending on their position in the genome, SNPs can be either coding or non-coding. The coding density in the MTBC has been reported to be 90–96%, with most of the SNPs in the MTBC in the coding region of the genome [6]. SNPs which results in an amino acid change are non-synonymous (nsSNPs) while those which do not cause a change in amino acid coding are synonymous (sSNPs).

While the function of the members of the PE_PGRS multigene family is not known, it is hypothesized that the PE_PGRS genes may be associated with antigenic and genetic variability as well as virulence [7–10]. Previous studies have been shown that some members of the PE_PGRS family are expressed on the cell surface during MTB infection and recognized by the host immune system [7,11,12]. It has been suggested that MTB PE_PGRS genes are variably expressed during infection [13–15]. The roles of PE_PGRS16 and PE_PGRS26 in causing infection have previously been studied in mouse models; the expression of PE_PGRS16 was significantly up-regulated as compared with PE_PGRS26, suggesting that these two PE_PGRS genes may serve as a marker of latent TB infection [14]. In addition, PE_PGRS33 gene has been shown to encode for a surface expressed protein, while transposon mutagenesis-based studies have shown that the MTB PE_PGRS33 gene plays an important role in interactions with other mycobacteria as well as with macrophages [16].

The sequence variations such as SNPs, insertions and deletions (INDELs) in MTB clinical strains have also been investigated for PE_PGRS16, PE_PGRS26 and PE_PGRS33 [17,18]. Moreover, in a population-based study of 649 clinical MTB strains, patients infected with MTB strains with a significant change in the PE_PGRS33 protein were more likely to belong to a cluster of TB cases as compared with MTB strains with minimal changes to the PE_PGRS33 protein suggests that PE_PGRS33 may play an important role in MTB transmission [19].

The extensively drug-resistant tuberculosis (XDR-TB) has emerged worldwide as one of the biggest threats to public health and TB control programs. XDR-TB is defined as TB caused by the MTB strain that is resistant to at least rifampin (RIF) and isoniazid (INH) among the first-line anti-TB drugs and resistant to fluoroquinolones and to at least one of the three injectable second-line drugs [20–22].

Previous studies have reported SNPs in MTBC as drug resistance-conferring mutations, with 1447 mutations significant for most anti-TB drugs [23–26]. However, no data are available on sequence variations such as SNPs and INDELs in the PE_PGRS genes of XDR MTB isolates. This study analyzed 37 XDR MTB strains for the presence of SNPs, INDELs by whole genome sequencing (WGS) analysis method. The aim of this study is to specifically investigate SNPs and INDELs in XDR MTB strains to understand how genetic diversity contributes to antigenic variability.

Methodology

Strain selection

The MTB strains were obtained from the strain bank of Aga Khan University Clinical Microbiology Laboratory, Pakistan. The 37 XDR MTB strains were from the period of 2004–2009, which had previously been spoligotyped [27–29], and were randomly selected for the WGS analysis. The study samples selected for the WGS analysis belonged to all three Principal Genetic Groups: PGG1, PGG2 and PGG3, which were comprised of PGG1: 21 CAS1, 2 CAS, 1 CAS2, 2 EAI and 7 Other lineages (1 and 3) (89.2%); PGG2: 1 X3 (2.7%); and PGG3: 3 T (8.1%) groups, respectively.

Culture and antibiotic susceptibility testing

The XDR MTB strains used had been isolated from the specimens using Lowenstein–Jensen media and MGIT (Becton Dickinson, Franklin Lakes, NJ, and USA). MTB was identified using BACTEC NAP TB differentiation test (Becton Dickinson), growth on para-nitrobenzoic acid containing media, nitrate reduction, and niacin accumulation [30,31]. The drug susceptibility testing (DST) of these isolates was previously performed using an agar proportion method on enriched Middlebrook 7H10 medium (BBL Microbiology Systems, Cockeysville, MD, USA) at the following concentrations: rifampicin 1 µg/mL, isoniazid 0.2 µg/mL, streptomycin 2 µg/mL and 10 µg/mL, and ethambutol 5 µg/mL. Pyrazinamide sensitivity was determined by using BACTEC 7H12 medium, pH 6.0, at 100 µg/mL (BACTEC PZA test medium, Becton Dickinson). MDR TB strains were further tested with capreomycin 10 µg/mL, ciprofloxacin 2 µg/mL, ethionamide 5 µg/mL, amikacin 5 µg/mL, and kanamycin 6 µg/mL. Reference strain MTB H37Rv was used as a control with each susceptibility testing batch [32].

DNA extraction and spoligotyping

Spoligotyping was performed as described previously [33] and inferred *in silico*; the spoligotypes were also confirmed from the sequencing reads using SpolPred software [34].

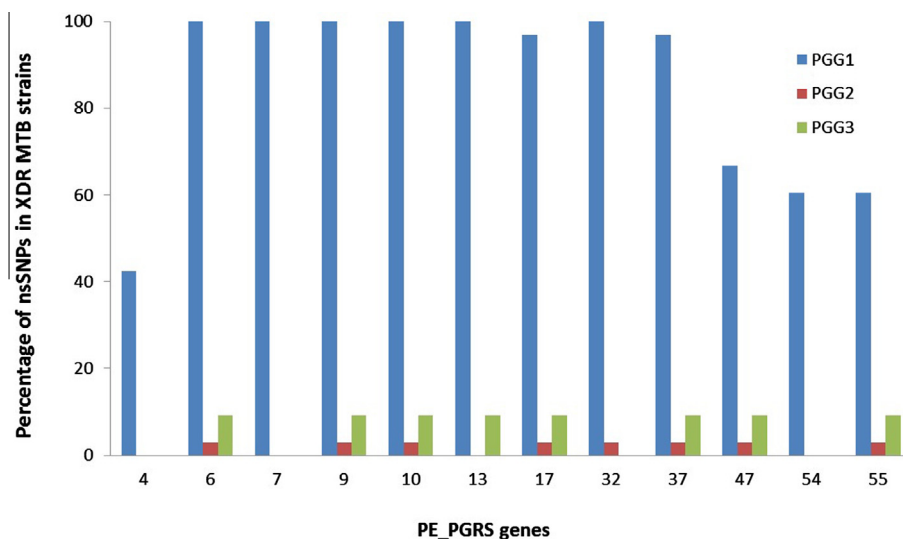


Fig. 1 – Occurrence of SNPs in PE_PGRS genes according to PGG lineage of strains. The figure indicates the occurrence of non-synonymous (ns) SNPs in different PE_PGRS genes according to their PGG lineages.

was present in 35 (94.6%) with a 9 bp INS, while PE_PGRS53 INS was present in 36 (97.3%) of XDR MTB strains with INS size range of 9–18 bp. The INS in the PE_PGRS genes 6, 19, 28, 30, 33, 50, 57 and 61 were variably present in the XDR MTB strains with INS size ranging from 1 to 21 bp.

The INS were also observed to be significantly more in PE_PGRS6 ($p = 0.05$) and PE_PGRS30 ($p = 0.0447$) in XDR CAS1 as compared with XDR Other lineages strains (T1 and X3) groups analyzed by Pearson's Chi-square test (Table 2).

Genetic diversity among the XDR MTB Principal Genetic Groups (PGG)

Of the 42 PE_PGRS genes investigated for detection of 37 XDR MTB strains by WGS, 34 (91.90%), 1 (2.7%) and 3 (8.1%) of the 37 MTB strains belonged to PGG1, PGG2 and PGG3, respectively. CAS1, CAS CAS2, EAI and Orphan belong to PGG1, while X3 and T1 belong to the PGG2 and PGG3 groups, respectively. There were significantly more nsSNPs present in PGG1 strains ($n = 34$) as compared with PGG2 ($n = 1$) and PGG3 ($n = 2$) strains (Fig. 1). However, given the larger proportion of PGG1 strains in this sample size of XDR strains as compared with others, it is difficult to get a complete picture for this comparison.

Discussion

The PE_PGRS genes have been shown to be limited to mycobacterial species; they have specially expanded within the genomes of pathogenic mycobacteria, probably through widespread gene duplication events and genetic divergence during their adaptation to the intra-macrophage environment [9].

It has been hypothesized that gene conversion may have contributed to the evolution of members of the PE_PGRS subfamily and may have participated in the generation of antigenic variation in their members [2,7,17]. A previous

study has shown nucleotide substitution of A to G in the PE_PGRS17 genes of CAS, EAI and LAM strains. In addition, SNPs of G to A, C to A and G to C nucleotide substitutions were observed in PE_PGRS18 genes of CAS, EAI, LAM and Haarlem MTB strains [36].

The PGG1 is considered to be the most ancient of the MTB lineages, which has also been reported for other regions of Asia [37]. In this study, it was observed that all XDR MTB strains belonging to the PGG1 group had nsSNPs in the PE_PGRS6, 7, 9, 10, 13 and 32 genes, whereas it was variable in the PGG2 and PGG3 XDR MTB strains. However, the large proportion of PGG1 strains in this study as compared with PGG2 and PGG3 strains makes it difficult to get a complete picture of the proportion of SNPs. In this study, the nucleotide substitution of A to G and T to C in the PE_PGRS17 genes of CAS1, CAS, CAS2, EAI, T, X and orphan (lineages 1 and 3) XDR MTB strains was also observed. Moreover, nucleotide substitution of A to G, C to T and A to C in the PE_PGRS18 genes of CAS1, CAS, CAS2, EAI3, T, X and Other lineages (1 and 3) XDR MTB strains was also observed. However, the function of these PE_PGRS genes are yet unknown [36].

In a population-based study of 649 MTB clinical isolates it was shown that in PE_PGRS33 gene, Thr47Ile was found in 1 isolate, but Thr1172Ile was found in 11 (1.7%) isolates, respectively [38]. However, non-synonymous SNPs (nsSNPs) in the PE_PGRS33 genes were not observed in the XDR MTB strains in the present study. Also, previously, it has been observed that a Gly686Asp in PE_PGRS26 genes was observed in 20 (10%) MTB clinical strains belonging to the Principal Genetic Group (PGG2), while Gly1122Gly in the PE_PGRS26 genes was observed in 1 (0.5%) of the MTB clinical strains also belonging to PGG2 [18]. However, this study did not observe nsSNPs in the PE_PGRS26 genes of XDR MTB strains.

In addition to SNPs, deletions have also been shown in the PE_PGRS genes 16, 26 and 33. In the XDR strains studied here, a 9 bp DEL was observed in the PE_PGRS16 gene sequence of 2 (25%) XDR Other lineage 1 MTB strains

belonging to PGG1 strains. Previously, in the PE_PGRS16 gene sequence, a 3 bp deletion in PGG3 group strains and a 252 bp deletion in three PGG2 group strains have been reported [18]. Also, in the PE_PGRS26 gene, both 3 bp and 150 bp DELs have been reported in PGG2 group strains and PGG1 group strains [18]. In this study, 2 (9.5%) CAS1 MTB strains showed 9 bp DELs in the PE_PGRS26 gene belonging to the PGG1 group. It was observed that 1–9 bp DELs were present in the PE_PGRS33 gene in EAI (100%) and Other lineage 1 (28.57%). This correlates with previous work which has 3 bp, 9 bp and 273 bp, respectively, in the PE_PGRS33 gene of MTB strains [38].

In this study, the variable 18 bp INS in PE_PGRS26 and 9 bp INS PE_PGRS33 was observed. INS of 9 bp and 18 bp have been reported in the PE_PGRS26 gene [18].

Overall, this work demonstrates the variability in the genome of MTB clinical strains. As these PE_PGRS genes contribute to immunogenicity, the variations in XDR strains may be associated with their ability to persist within the host.

Conclusion

Multiple DELs and INS were observed in the PE_PGRS genes in CAS1 XDR–MTB strains. As these strains are predominant in this endemic region, it indicates a need to further understand the functions of the PE_PGRS genes and its persistence.

Competing interest

Authors declare that they have no competing interests.

Ethical approval

This work received approval from the Ethical Review Committee of the Aga Khan University.

Acknowledgements

We thank the Department of Pathology and Microbiology, Clinical Microbiology, Aga Khan University Hospital for providing support for collection of samples used in the study. Our thanks to Pathogen Genomics Laboratory, King Abdullah University of Science and Technology (KAUST) for whole genome sequencing of MTB strains.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ijmyco.2014.11.049>.

REFERENCES

- [1] WHO, The global plan to stop TB, 2006–2015.
- [2] S.T. Cole et al, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, *Nature* 393 (6685) (1998) 537–544.
- [3] G. Lamichhane et al, A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to *Mycobacterium tuberculosis*, *Proc. Natl. Acad. Sci. U. S. A.* 100 (12) (2003) 7213–7218.
- [4] I. Comas et al, Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved, *Nat. Genet.* 42 (6) (2010) 498–503.
- [5] R. Hershberg et al, High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography, *PLoS Biol.* 6 (12) (2008) e311.
- [6] A. Namouchi et al, After the bottleneck: genome-wide diversification of the *Mycobacterium tuberculosis* complex by mutation, recombination, and natural selection, *Genome Res.* 22 (4) (2012) 721–734.
- [7] S. Banu et al, Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens?, *Mol. Microbiol.* 44 (1) (2002) 9–19.
- [8] G. Delogu et al, Rv1818c-encoded PE_PGRS protein of *Mycobacterium tuberculosis* is surface exposed and influences bacterial cell structure, *Mol. Microbiol.* 52 (3) (2004) 725–733.
- [9] M.J. Brennan, G. Delogu, The PE multigene family: a ‘molecular mantra’ for mycobacteria, *Trends Microbiol.* 10 (5) (2002) 246–249.
- [10] L. Ramakrishnan, N.A. Federspiel, S. Falkow, Granuloma-specific expression of *Mycobacterium* virulence proteins from the glycine-rich PE-PGRS family, *Science* 288 (5470) (2000) 1436–1439.
- [11] K.K. Singh et al, Antigens of *Mycobacterium tuberculosis* expressed during preclinical tuberculosis: serological immunodominance of proteins with repetitive amino acid sequences, *Infect. Immun.* 69 (6) (2001) 4185–4191.
- [12] G. Delogu, M.J. Brennan, Comparative immune response to PE and PE_PGRS antigens of *Mycobacterium tuberculosis*, *Infect. Immun.* 69 (9) (2001) 5606–5611.
- [13] M.I. Voskuil et al, Regulation of the *Mycobacterium tuberculosis* PE/PPE genes, *Tuberculosis (Edinb.)* 84 (3–4) (2004) 256–262.
- [14] V. Dheenadhayalan et al, Variable expression patterns of *Mycobacterium tuberculosis* PE_PGRS genes: evidence that PE_PGRS16 and PE_PGRS26 are inversely regulated in vivo, *J. Bacteriol.* 188 (10) (2006) 3721–3725.
- [15] G. Delogu et al, PE_PGRS proteins are differentially expressed by *Mycobacterium tuberculosis* in host tissues, *Microbes Infect.* 8 (8) (2006) 2061–2067.
- [16] M.J. Brennan et al, Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells, *Infect. Immun.* 69 (12) (2001) 7326–7333.
- [17] S. Talarico et al, Variation of the *Mycobacterium tuberculosis* PE_PGRS 33 gene among clinical isolates, *J. Clin. Microbiol.* 43 (10) (2005) 4954–4960.
- [18] S. Talarico et al, *Mycobacterium tuberculosis* PE_PGRS16 and PE_PGRS26 genetic polymorphism among clinical isolates, *Tuberculosis (Edinb.)* 88 (4) (2008) 283–294.
- [19] F. Chaves et al, Usefulness of the secondary probe pTBN12 in DNA fingerprinting of *Mycobacterium tuberculosis*, *J. Clin. Microbiol.* 34 (5) (1996) 1118–1123.
- [20] World Health Organization (WHO), Case definition for extensively drug-resistant tuberculosis, in *Wkly. Epidemiol. Rec.* 2006, World Health Organization.
- [21] R. Hasan, Drug resistant tuberculosis: challenges of urbanization, *Int. J. Mycobacteriol.* 3 (2) (2014) 79–81.
- [22] R. Macedo, A.F. Antunes, M. Villar, I. Portugal, Multidrug and extensively drug-resistant tuberculosis in Lisbon and Vale do Tejo, Portugal, from 2008 to 2010, *Int. J. Mycobacteriol.* 1 (3) (2014) 131–136.
- [23] P.F. Riska, W.R. Jacobs Jr., D. Alland, Molecular determinants of drug resistance in tuberculosis, *Int. J. Tuberc. Lung Dis.* 4 (2 Suppl. 1) (2000) S4–S10.

- [24] J.M. Musser, Antimicrobial agent resistance in mycobacteria: molecular genetic insights, *Clin. Microbiol. Rev.* 8 (4) (1995) 496–514.
- [25] A. Telenti, Genetics of drug resistance in tuberculosis, *Clin. Chest Med.* 18 (1) (1997) 55–64.
- [26] S. Ramaswamy, J.M. Musser, Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update, *Tuber. Lung Dis.* 79 (1) (1998) 3–29.
- [27] A. Kanji et al, Occurrence of RD149 and RD152 deletions in *Mycobacterium tuberculosis* strains from Pakistan, *J. Infect. Dev. Ctries.* 5 (2) (2011) 106–113.
- [28] A. Ali et al, Characterization of mutations conferring extensive drug resistance to *Mycobacterium tuberculosis* isolates in Pakistan, *Antimicrob. Agents Chemother.* 55 (12) (2011) 5654–5659.
- [29] Z.H. Asho Ali, S. Jafri, R. Inayat, R. Hasan, *Mycobacterium tuberculosis* Central Asian Strain (CAS) lineage strains in Pakistan reveal lower diversity of MIRU loci than other strains, *Int. J. Mycobacteriol.* 3 (2) (2014) 108–116.
- [30] E.W. Koneman, S.D. Allen, V.R. Dowell, H.M. Sommers, *Color Atlas Textbook of Diagnostic Microbiology*, JB Lippincott Company, Philadelphia, 1983.
- [31] M. Hasan, S.K. Munshi, M.S. Banu Momi, F. Rahman, R. Noor, Evaluation of the effectiveness of BACTEC MGIT 960 for the detection of mycobacteria in Bangladesh, *Int. J. Mycobacteriol.* 2 (4) (2014) 214–219.
- [32] G.L. Woods, B.A. Brown-Elliott, P.S. Conville, E.P. Desmond, G.S. Hall, G. Lin, et al., *Susceptibility Testing of Mycobacteria, Nocardiae, and Other Aerobic Actinomycetes*; Approved Standard, NCCLS, Wayne, PA, 2003, Document No. M24-A.
- [33] J. Kamerbeek et al, Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology, *J. Clin. Microbiol.* 35 (4) (1997) 907–914.
- [34] F. Coll et al, SpolPred: rapid and accurate prediction of *Mycobacterium tuberculosis* spoligotypes from short genomic sequences, *Bioinformatics* 28 (22) (2012) 2991–2993.
- [35] S. Honore-Bouakline et al, Rapid diagnosis of extrapulmonary tuberculosis by PCR: impact of sample preparation and DNA extraction, *J. Clin. Microbiol.* 41 (6) (2003) 2323–2329.
- [36] A. Karboul et al, Insights into the evolutionary history of tubercle bacilli as disclosed by genetic rearrangements within a PE_PGRS duplicated gene pair, *BMC Evol. Biol.* 6 (2006) 107.
- [37] M. Arjomandzadegan et al, Determination of principal genotypic groups among susceptible, MDR and XDR clinical isolates of *Mycobacterium tuberculosis* in Belarus and Iran, *Tuberk Toraks* 60 (2) (2012) 153–159.
- [38] S. Talarico et al, Association of *Mycobacterium tuberculosis* PE PGRS33 polymorphism with clinical and epidemiological characteristics, *Tuberculosis (Edinb.)* 87 (4) (2007) 338–346.

