

Analysis of sequence diversity among IS6110 sequence of *Mycobacterium tuberculosis*: possible implications for PCR based detection

Sathish Sankar*, Suresh Kuppanan, Babu Balakrishnan, Balaji Nandagopal

Division of Biomedical Research, Sri Narayani Hospital and Research Centre, Thirumalaikodi, Sripuram, Vellore - 632 055, Tamil Nadu, India; Sathish Sankar - Email: sathish3107@gmail.com; Phone: 0416-2271202; *Corresponding author

Received June 01, 2011; Accepted June 06, 2011; Published June 23, 2011

Abstract:

The IS6110 belongs to the family of insertion sequences (IS) of the IS3 category. This insertion sequence was reported to be specific for *Mycobacterium tuberculosis* complex and hence is extensively exploited for laboratory detection of the agent of tuberculosis and for epidemiological investigations based on polymerase chain reaction. IS6110 is 1361-bp long and within this sequence different regions have been utilized as targets in the identification of *M. tuberculosis* by PCR. However, the results are not always consistent, specific and sensitive. In recent years, a few clinical investigations raised concerns over IS6110 specificity and sensitivity in the diagnosis of tuberculosis due to false-positive (homology with other target DNA besides *M. tuberculosis*) or false negative (due to absence of copies of IS6110) results with IS6110 specific primers. To unravel the variations in IS6110 sequences, an *in silico* analysis of IS6110 sequence of different strains of *M. tuberculosis* was carried out. Our results of comparative analysis of IS6110 insertion sequences of *M. tuberculosis* complex suggests that, IS6110 insertion sequences harbored variations in its sequence, which is evident from the phylogenetic analysis. Importantly, IS6110 sequence has divergence within the copies of same strain and formed different clusters. A list of IS6110 specific primers used in various clinical investigation of tuberculosis was obtained from the literature and their performance scrutinized. Our study emphasizes the need to develop PCR assays (multiplex format) targeting more than one region of the genome of *M. tuberculosis*.

Keywords: IS6110, *Mycobacterium tuberculosis*, Phylogenetic analysis, Tuberculosis diagnosis

Background and Description:

Tuberculosis (TB), an infectious disease affects millions of humans worldwide and is caused by the *Mycobacterium tuberculosis* complex. In 2009, WHO estimated that there were about 9.4 million new cases, with 1.3 million deaths globally due to tuberculosis [1]. This high incidence of TB worldwide necessitates research on developing precise diagnostic methods for specific treatment and management. Currently, the disease is diagnosed by sputum smear examination, a rapid and cheap method but lacks specificity. Traditional microbial culture utilizes solid (Lowenstein-Jensen) or liquid media which provides a definitive diagnosis of an active infection but is time consuming (6-8 weeks). With the advent of automated or semi-automated liquid culture system the time to detect the growth of mycobacterial species has been significantly shortened (roughly 14 days) [2]. However, the systems are expensive and not available even in many tertiary care centres in developing countries. In recent years, a number of molecular diagnostic methods for tuberculosis have been developed based on polymerase chain reaction (PCR) amplification targeting certain sequences of IS6110, *hsp65*, 16srRNA, 85B antigen, 38kDa antigen. The IS6110 belongs to a family of insertion sequences (IS) of the IS3 category and its amplification is most commonly used in the detection of *M. tuberculosis* because it is highly conserved. In addition to diagnosis, IS6110 insertion sequence has also been utilized for molecular epidemiological analysis of clinical isolates. However, the sensitivity and specificity of IS6110 sequence in the diagnosis of tuberculosis remains

uncertain and needs to be appraised by in-silico analysis. In recent years, IS6110 based diagnosis has been shown to be hampered by the presence of low copy number or absence of this IS6110 repetitive sequence.

A few clinical investigations reported the presence of low copy number of IS6110 in *M. tuberculosis* strains from regions such as Tunisia, where, 75% of the strains showed 6-10 copies [3]. In Denmark 50% of the strains analyzed showed 11-15 copies [4] and in different geographical regions of India 11 to 20% of the strains showed nil or 1-2 copies of IS6110 [5]. A small number (<1%) of IS6110 deficient strains have been reported in San Francisco [6], and 2% in Vietnam [7]. In the present study, our primary aim was to detect the IS6110 sequence differences within the *M. tuberculosis* complex by means of construction of phylogenetic trees based on available IS6110 sequence at GenBank database and the secondary aim was to analyze the discriminatory potential of a list of primer pairs reported in literature which were used in the PCR techniques to diagnose tuberculosis.

Retrieval of IS6110 sequences:

The IS6110 sequences were retrieved from the GenBank repository for the *M. tuberculosis* complex, which includes *M. tuberculosis* H37Rv (Accession No: NC_000662.2), *M. tuberculosis* H37Ra (Accession No: NC_009525), *M. tuberculosis* CDC1551 (Accession No: NC_002755), *M. bovis* strain Pasteur 1173P2 (Accession No: NC_008769), *M. bovis* AF2122/97 (Accession No:

NC_002495.3), *M. avium* ssp. *paratuberculosis* K-10, *M. fortuitum*, *M. tuberculosis* clone TB3A, *M. tuberculosis* clone TB4C, *M. tuberculosis* clone TB2A, *M. tuberculosis* SAW1690, *M. tuberculosis* H2255, *M. tuberculosis* D7030, *M. tuberculosis* B9, *M. tuberculosis* isolate patient 18 and *M. tuberculosis* Norway strain. These 35 strains were subjected to in-silico analysis.

Sequence alignment and phylogeny construction:

The ClustalW program of the MEGA 5.03 [8] was used to align the IS6110 nucleotide sequences. Each alignment was visually examined to detect short misalignments and poorly aligned regions. From these aligned IS6110 nucleotide sequences, phylogenetic tree was constructed using Maximum Likelihood method using Tamura-Nei model. The discriminatory potential of a few primer sets which have been used successfully to amplify different regions of IS6110 sequences in PCR assays for the detection of *M. tuberculosis* in various clinical investigations was also analyzed.

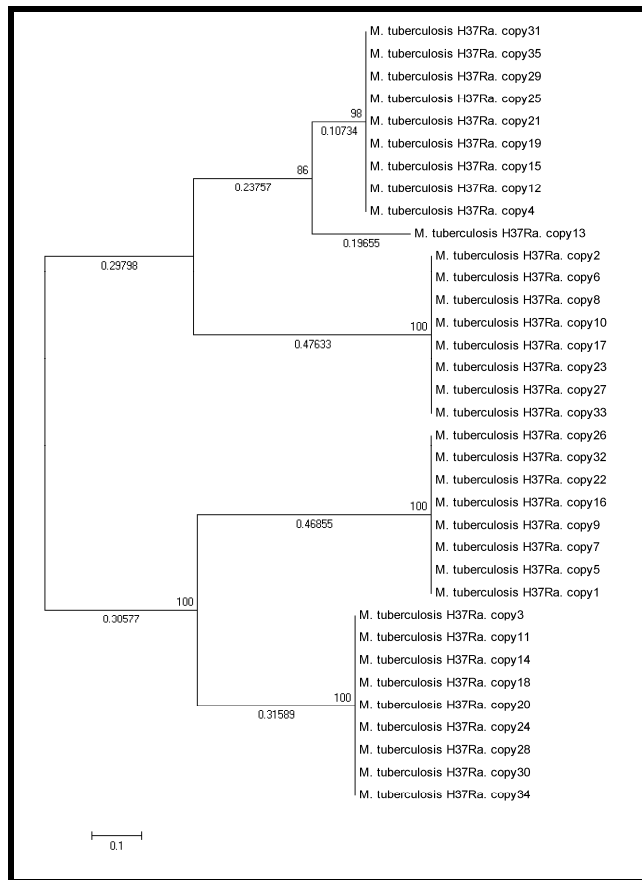


Figure 1: A dendrogram showing IS6110 sequence divergence of *M. tuberculosis* H37Ra. Bootstrap values were calculated for 500 replications. The values above the branch indicate bootstrap value; the values below the tree indicate branch length.

Phylogenetic analysis:

The phylogenetic analysis of the IS6110 sequence of *M. tuberculosis* H37Ra strain revealed the existence of sequence divergence among the individual copies of IS6110 in a given strain (Figure 1) and showed four different clusters. A similar kind of situation was also found for *M. tuberculosis* H37Rv. It showed that two different main clusters and a few copies of IS6110 of H37Rv formed individual clusters (data not shown). Phylogenetic analysis of IS6110 sequences for *M. tuberculosis* CDC1551 revealed sequence divergence by forming two different clusters (data not shown). In our analysis, *M. tuberculosis* clone TB3A, *M. tuberculosis* TB4C, *M. tuberculosis* clone TB2 contributes to cluster I, whereas, *M. tuberculosis* SAW1690 formed sub-cluster of cluster I. *M. tuberculosis* D7030, *M. avium* ssp. *paratuberculosis* K-10, *M. tuberculosis* B9, *M. tuberculosis* isolate patient 18 and *M. tuberculosis* Norway strains formed independent clusters (data not shown). The IS6110 sequences of both the *M. bovis* AF2122/97 and *M. bovis* strain Pasteur were also included

for phylogenetic analysis. Of them, *M. bovis* AF2122/97 seg7/14 copy 1 and *M. bovis* Pasteur copy 1 contributes to cluster I. *M. bovis* AF2122/97 seg9/14 and *M. bovis* Pasteur copy 4 formed cluster 2, whereas, the other 7 copies of IS6110 formed independent clusters (data not shown). The above phylogenetic results of the IS6110 sequences revealed the divergence in IS6110 sequences within the strains of *M. tuberculosis* and other species. However, the IS6110 sequence is still a favorite target sequence in the diagnosis and epidemiological investigations of *M. tuberculosis* to properly categorize mycobacteria and for public health officials to understand the disease epidemiology.

Detection of IS6110 sequence by RFLP is the most widely used genotyping method besides others (Spoligotyping, MIRU-VNTR, 16SrRNA), for epidemiologic studies. At the same time, detection of IS6110 sequence was the commonest approach unlike *hsp65* and *mip40* target sequences used for diagnosis of tuberculosis. Reports suggested that, the sensitivity and specificity of the IS6110 for *M. tuberculosis* varies between laboratories [9]. Further, false positive results are being reported frequently. In this context, we have collected a list of primers which were successfully used in the conventional PCR for the diagnosis of *M. tuberculosis* from the 13 articles [10-22]. Of the 13 investigations, 7 investigations used primers which amplified nucleotides from nucleotide position 762 to 883, a 123-bp final product has been produced as a result of PCR and the remaining 6 investigations used different regions of IS6110 (Figure 2). Different regions of IS6110 sequence are being used as target for PCR amplification. Much of these do not describe the sensitivity and specificity of the assay. The reported data shows varied sensitivity and specificity for different regions of IS110 sequence. This further baffles the laboratorian in selecting an optimum primer set. A majority of the assays target nt 762-883 region for PCR amplification. A primer set used for PCR identification from buffy coat and sputum as well has been reported and found to be having 100% sensitivity and 95.1% specificity in nested format compared to blood culture (BacT/ALERT 3D, Biomeriux) [22]. These primers were found to be more efficient than other primers used to identify IS6110 lacking *M. tuberculosis* strains [23]. Thus, information on each primer sets along with its clinical- and assay- sensitivity and specificity, and its detection limit would be useful to identify an optimal primer set to facilitate rapid diagnosis.

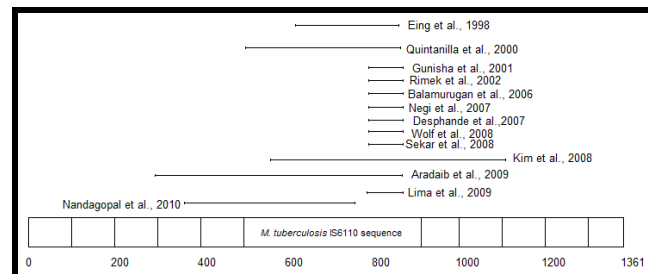


Figure 2: A graphical representation showing different target regions used for PCR detection of *M. tuberculosis*

A number of clinical investigations suggested that the reason behind the false positivity is due to carryover contamination of *M. tuberculosis* DNA which is probably not the story. Another reason was suggested by Sandhu et al. [24] i.e., IS6110 likely to have originated from (or been passed on to) other organisms and certain regions of DNA that may have remained conserved among these organisms during evolution. This raised the possibility that some laboratories may have amplified stretches of DNA related to IS6110 from organisms that were not *M. tuberculosis*. This would explain some of the false positive tests reported in the literature. In a search of GenBank with 181-bp sequence, 63.3% homology in a 177-bp overlap was found with the insertion sequence IS629 of *Shigella sonnei* and 63.8% homology in a 177-bp overlap was found with the insertion sequence IS3411 for *Escherichia coli*. In addition, with the IS6110 specific primers this central region was amplified and Southern hybridized with IS6110 derived probe from the majority of non-*M. tuberculosis* species, thereby we confirmed that this central region is homologous to mycobacteria other than tuberculosis. Together, these data suggests to us that care must be taken while designing specific primers for IS6110 sequence of *M. tuberculosis*. The presence of multiple copies of IS6110 in the majority of *M. tuberculosis* strains enhances the sensitivity of PCR tests by providing a source of "preamplified" sequence and in addition, deletion or mutation of one of the sequences should still allow for others to be intact for amplification. It is well documented that, *M. tuberculosis* isolates containing five or fewer IS6110 copies cannot be reliably differentiated by the RFLP method [25]. In conclusion, we suggest that great care must be given in designing primer pairs for IS6110 to prevent false positive or false negative results. In such cases where IS6110 RFLP typing is

not possible due to low or zero copy of the element, one should consider an alternate genotyping method for accurate epidemiologic or diagnostic purposes. For diagnostic PCRs, multiplexing by targeting two regions like IS6110 and hsp65 could be a good strategy.

Acknowledgement:

Financial assistance of the institution is gratefully acknowledged and the authors thank Prof. G. Sridharan for his useful comments in the preparation of the manuscript.

References:

- [1] http://www.who.int/tb/publications/global_report/2010/en/index.html
- [2] Drobniewski FA *et al. Lancet Infect Dis* 2003 **3**: 141 [PMID: 12614730]
- [3] Sola C *et al. J Med Evol*. 2001 **53**: 680 [PMID: 11677628]
- [4] Chevrel-Dellagi D *et al. J Clin Microbiol*. 1993 **31**: 2446 [PMID: 8104956]
- [5] Chauhan DS *et al. Indian J Med Res*. 2007 **125**: 577 [PMID: 17598945]
- [6] Agasino CB *et al. Int J Tuberc Lung Dis*. 1998 **2**: 518 [PMID: 9626611]
- [7] Park YK *et al. J Clin Microbiol*. 2000 **38**: 191 [PMID: 10618086]
- [8] Tamura K *et al. Mol Biol Evol*. 2011 (In Press) [PMID: 21546353]
- [9] Noordhoek GT *et al. J Clin Microbiol*. 1994 **32**: 277 [PMID: 8150935]
- [10] Eing BR *et al. J Clin Microbiol*. 1998 **36**: 2023 [PMID: 9650955]
- [11] Garcia-Quintanilla A *et al. J Clin Microbiol*. 2000 **38**: 1166 [PMID: 10699014]
- [12] Gunisha P *et al. Indian J Pathol Microbiol*. 2001 **44**: 97 [PMID: 11883143]
- [13] Rimek D *et al. J Clin Microbiol*. 2002 **40**: 3089 [PMID: 12149389]
- [14] Balamurugan R *et al. J Clin Microbiol*. 2006 **44**: 1884 [PMID: 16672431]
- [15] Negi SS *et al. Indian J Med Microbiol*. 2007 **25**: 43 [PMID: 17377352]
- [16] Deshpande PS *et al. Cerebrospinal Fluid Res*. 2007 **4**: 10 [PMID: 17976247]
- [17] Wolf H *et al. Am J Trop Med Hyg*. 2008 **79**: 893 [PMID: 19052299]
- [18] Sekar B *et al. Indian J Med Microbiol*. 2008 **26**: 352 [PMID: 18974489]
- [19] Kim MH *et al. Yonsei Med J*. 2008 **49**: 537 [PMID: 18729295]
- [20] Lima JF *et al. J Bras Pneumol*. 2009 **35**: 690 [PMID: 19669008]
- [21] Aradaib I *et al. International Journal of Tropical Medicine* 2009 **4**: 123
- [22] Nandagopal B *et al. Indian J Med Microbiol*. 2010 **28**: 227 [PMID: 20644311]
- [23] Sankar S *et al. Indian J Med Microbiol*. 2010 **28**: 303 [PMID: 20966559]
- [24] <http://www.patentstorm.us/patents/5731150/description.html>
- [25] Burman WJ *et al. Am J Respir Crit Care Med*. 1997 **155**: 1140 [PMID: 9001331]

Edited by P Kanguane

Citation: Sankar *et al.* Bioinformation 6(7): 283-285 (2011)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.