Published in final edited form as:

Infect Genet Evol. 2008 July ; 8(4): 474–483. doi:10.1016/j.meegid.2007.09.007.

Genomic Interrogation of Ancestral *Mycobacterium tuberculosis* from South India

Sujatha Narayanan^{1,2,*}, Sebastien Gagneux^{2,3}, Lalitha Hari¹, Anthony G. Tsolaki^{2,4}, Suganthi Rajasekhar¹, P.R. Narayanan¹, Peter M. Small^{2,3,5}, and Kathryn DeRiemer^{2,6}

¹ Tuberculosis Research Centre (TRC), Mayor VR Ramanathan Road, Chetput, Chennai, 600031, India.

² Division of Infectious Diseases and Geographic Medicine, Stanford University Medical Center, 300 Pasteur Drive, Stanford, California 94305-5107, USA

Abstract

Mycobacterium tuberculosis is a very important global pathogen. One quarter of the world's TB cases occur in India. The tuberculosis strains isolated from south Indian patients exhibit certain phenotypic characteristics like low virulence in guinea-pigs, resistance to isoniazid, thiophene-2-carboxylic acid hydrazide (TCH) and para-amino salicylic acid (PAS) enhanced susceptibility to H₂O₂. Besides this, a large percentage of the isolates harbor only a single copy of IS 6110 which makes these strains distinct. Hence we have studied the genotypic characteristics of these strains by using advanced techniques like Deletion Micro array, deletion PCR, allelic discrimination RT-PCR using several lineage specific markers and KatG G13887 (non synonymous) polymorphism along with spoligotyping. The analysis of 1,215 tuberculosis patient isolates from South India revealed that 85.2% belonged to the ancestral lineage of *Mycobacterium tuberculosis*.

Comparative whole-genome hybridization identified six new genomic regions within this lineage that were variably deleted.

INTRODUCTION

Mycobacterium tuberculosis is a very important global pathogen. Without HIV, the tuberculosis (TB) epidemic would now be in decline everywhere. India has the highest number of incident TB cases in the world. With effective DOTS implementation tuberculosis is expected to decline but as HIV continues to spread, the decline of TB is far from reality.

*Corresponding Author: Sujatha Narayanan, PhD, Tuberculosis Research Centre (TRC), Mayor VR Ramanathan Road, Chetput, Chennai, 600031, India. Phone: +91 44 28369627/9600, Fax: +91 44 2836 2528/2529. sujatha.sujatha36@gmail.com.

³Current address: Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103, USA

⁴Current address: Center for Molecular Biology and Infection (CMMI), Division of Investigative Sciences, Flowers Building, South Kensington Campus, Imperial College, London, SW7 2AZ, United Kingdom

⁵Current address: Bill and Melinda Gates Foundation, P.O. Box 23350, Seattle, WA, 98102, USA

⁶Current address: School of Medicine, University of California, One Shields Avenue, Davis, California, 95616, USA

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Although one fourth of the global tuberculosis burden stems from India there are a few studies which define the tuberculosis genogroup (Bhanu et al., 2002, Das, et al., 2005, Mistry et al., 2002, Narayanan et al., 1997, Kulkarni et al., 2005, Singh et al., 2004, Radhakrishnan et al., 2001). There has been a population based study from south India on transmission dynamics and risk factors associated with transmission (Narayanan, et al., 2002). The information obtained in this study suffers from a lack of portability which hinders comparison between global studies. In addition, the south Indian isolates included by others in the previous studies have been limited in number (Guitierrez et al., 2006, Singh et al., 2007). To avoid the problem in portability of data and limitation in number we have used spoligotyping to genotype the isolates from a population based study.

Spoligotyping is a PCR based genotyping method that allows to assess the *M. tuberculosis* genetic biodiversity and provides enough information about the epidemiologically important clones in various settings (Cowan et al., 2007, Streicher et al., 2004). The establishment of an international database is an advantage to this method because it is very informative regarding the endemicity or the ubiquitous strains (Filiol et al., 2002, 2003).

The tuberculosis strains isolated from south Indian patients exhibit certain phenotypic characteristics like low virulences in guineapigs, resistance to isoniazid (Mitchison et al., 1960, Middlebrook et al., 1953) thiophene-2-carboxylic acid hydrazide (TC4) and para-amino salicylic acid (PAS) (Joseph et al., 1964) enhanced susceptibility to H₂O₂ and a majority of them belonging to intermediate phage type. (Bhatia A.L. 1961, Grange et al., 1978) Besides this a large percentage of the isolates harbour only a single copy of IS6110 (Das et.al 1995) which make these strains distinct. Hence we wanted to study the genotypic characteristics of these *M.tuberculosis* strains by using advanced techniques like Deletion Microarray, deletion PCR, allelic discrimination RT-PCR using several lineage specific markers and KatG G13887 (non synonymous) polymorphism (Sreevatsan et al., 1997) along with spoligotyping and IS6110 RFLP which is considered the gold standard for genotyping techniques.

The deletion micro array is an approach complementary to comparative genomics. This involves the interrogation of unsequenced genomes by DNA-microarray to identify sequences present in a fully sequenced isolate but absent from interrogated isolate (Kato-Maeda et al., 2001, Tsolaki et al., 2004). Nucleotide sequences provide robust, portable and comparable data for studying population variation. The mutational processes that generate these variations are understood and sequence data have been successfully used in the study of evolution, bacterial epidemiology and population biology (Baker L. 2004).

MATERIALS AND METHODS

IS6110 RFLP & Spoligotyping

We used a standardized international protocol for IS6110 RFLP genotyping (van Embden J D et al., 1993). Spoligotyping was performed as previously described (Kamerbeek et al., 1997). All of the spoligotype patterns were coded using the octal code system (Dale et al., 2001) with *M.bovis* P₃ and H37Rv as positive controls and autoclaved milli Q water as negative control. The DR region was amplified using the DRa (5'biotinylated) and DR b

primers. The amplified product was hybridized to a set of 43 immobilised oligonucleotides, each corresponding to a unique spacer sequence within the DR locus. Hybridisation was detected by enhanced chemiluminescence (ECL Enhanced Chemo-Luminescence Detection Kit; Amersham Hongkong) followed by exposure to X-ray film according to the manufacturer's recommendations. The results were documented in the form of binary code (x and 0) according to the hybridization (respectively positive or negative) for each spacer nucleotide probe and entered into excel format.

We then referred to a standardized international database of spoligotype patterns, SpolDB4 (Brudey et al., 2006) to determine whether each pattern had been previously reported. The updated SpolDB4 version contained 39,295 patterns distributed into 1939 shared types (patterns reported at least twice that grouped 36,925 clinical isolates) and 337 orphan patterns from more than 120 countries. The database provides detailed information about the country of origin and the geographic distribution within eight regions of the world.

Identification of Large Sequence Polymorphism

We selected 25 isolates for comparative whole-genome hybridization using an Affymetrix DNA chip (Santa Clara, California, USA) following procedures previously described. We identified putative deletions in the experimental strains relative to the sequenced reference strain H37Rv using DelScan software (AbaSci, San Pablo, California, USA), and confirmed the putative deletions by PCR and direct sequencing. We used phylogenetically informative large sequence polymorphisms (LSPs) or genomic deletions, to screen one isolate per spoligotype patterns (102 isolates) by PCR or multiplex real-time PCR. The screening results from the clustered isolates were extrapolated to the remaining isolates of their respective clusters.

We used previously published studies of genomic deletions (Tsolaki et al., 2004) to identify phylogenetically important genomic deletions and screen for them using PCR. For the detection of LSPs by multiplex real-time PCR, we designed a series of assays based on different TaqMan primer/probe combinations. The presence or absence of RD105 and RD239 was evaluated using following primers and probes: RD105-F, 5'-AACGAAGTGCCTGACTGAATC-3'; RD105-R, 5'-TCCCGCACCGGTTGAG-3'; RD105-probe, 5'-FAM-AGAGTGGACAGTTTCG-MGBNFQ-3'; RD239-F 5'-CGAGCTCAATCCGAACGAAA-3'; RD239-R, 5'-CCGGGCTTGGCTTTAACTG-3'; RD239-probe, 5'-VIC®-CCAGGTGCTTGCCATG-MGBNFQ-3'. Samples with no increase in fluorescence of either FAM or VIC® were considered deleted for the corresponding region of interest. Appropriate positive and negative controls were run on each plate. Isolates with an increase in both FAM and VIC® signal were considered not deleted for either RD and analyzed further. Similarly, the presence or absence of TbD1 and RD9 was evaluated with the following primers and probes: TbD1-F, 5'-CCGATTGACCACAGCTCGAT-3'; TbD1-R, 5'-CTGGCCGACGCTTTGC-3'; TbD1-probe, 5'-FAM-CCGTTTCAGATCAGC-3'; RD9-F, 5'-TGGTGGCGGTAGGTTTCAC 3'; RD9-R, 5'-ATGACCCGCGCGATGT-3'; RD9-probe, 5'-VIC®-TTCGACCCCAAGAC

Real Time PCR

The Euro-American lineage was defined based on a characteristic seven base pair deletion in pks15/1 (Constant et al., 2002) or the ctg to cgg substitution at codon 463 of *katG*. These two markers are known to be equivalent based on previous studies (Baker et al., 2004, Gutacker et al., 2002, Marmiesse et al., 2004). The *katG*463 SNP was analyzed by allelic discrimination using a TaqMan multiplex real-time PCR assay (Applied Biosystems, Foster City, CA, USA). The TaqMan-probes were labeled with 6-carboxyfluorescein (FAM) or VIC® at the 5'-end, and conjugated with a nonfluorescent quencher (NFQ) and a minor groove binding (MGB) group at the 3'-end. The following primers and probes were used: *katG*463-L: 5'-CCGAGATTGCCAGCCTTAAG-3', *katG*463-R: 5'-GAAACTAGCTGTGAGACAGTC AATCC-3', *katG*463-cgg probe: 5'-FAM-CAGATCCGGGCATC-MGBNFQ 3', *katG*463-ctg probe: 5'-VIC®-CCAGATCCTGGCATC-MGBNFQ 3'. Samples were run in 25µl reactions in 96-well plates on an ABI 7000 sequence detection system (Applied Biosystems, Foster City, California, USA). The reaction conditions were as recommended by the manufacturer.

RESULTS

Strain diversity based on spoligotyping

The study was approved by the ethical review committee of the Tuberculosis Research Centre (TRC, 1999) and Stanford University. There were 3,036 persons diagnosed with pulmonary tuberculosis during the study period, June 1999-June 2002, in the Tiruvallur District, Tamil Nadu state, South India. Of these, 1,456 (48.0%) were culture positive for *M. tuberculosis*. Spoligotyping was performed on 83.4% (1,215/1,456) of patient isolates. Based on the spoligotype patterns, there were 102 different genotypic clusters, each with at least two different patients (median, 3 patients; range, 2-338 patients) and a total of 1,038 tuberculosis patients (85.4%) in the clusters (Supplementary Table). The remaining 177 (14.6%) tuberculosis patients had an isolate with a unique spoligotype pattern.

The spoligopattern of all the isolates were compared to SpolDB4 (Brudey et al., 2006) to determine whether each pattern has been previously reported. We found 42 clusters with spoligotypes matching the SpolDB4 (Fig.1). There were 23 unique isolates with spoligopatterns which matched SpolDB4 database. 56 clusters had spoligotypes which did not match the SpolDB4 database (Fig. 2). These were considered as clusters of orphan isolates. There were 152 orphan strains which were not involved in clusters. The pattern of the 42 major clusters which matched the SpolDB4 data base has been shown (Fig. 1). Three clustered clinical isolates had spoligotype pattern lacking spacers 20, 21 and 33-36 a profile generally associated with *M. tuberculosis* H37Rv and 1 isolate had a pattern of *M. microti*. When we analyzed our data using spotclust (online software) the tuberculosis isolates were classified into 9 major clades with sub clades (Table 1). The Spotclust assigned the orphans to be variants of major clades and thus reduced the no of orphans to 30 (2.4%) (Table 1).

Table 1A and Supplementary table shows the different spoligotype clades and spoligotypes present in the Thiruvallur region respectively. EAI3 and EAI5 are represented predominantly upto 41% each whereas 5% of the isolates belong to EAI 1 & 4 and 1.6% belong to EAI 2.

Compared to northern regions like Delhi and Lucknow, the CAS isolates are represented by low numbers in Thiruvallur (4.0%) (Singh et al., 2007). The Beijing, LAM, T1 & Family 33 genotypes accounted for 2.0 %, 1.0 %, 3.0 % and 1.0 % respectively.

1B shows the phylogenetic tree. Phylogenetic analysis was carried out with spoligotyping data of all the isolates included in the study. Nei's original measures of genetic identity and genetic distance analysis was carried out using POPgene software version 1.32 (<http://www.ualberta.ca/fyeh/download>). The phylogenetic tree has grouped the subclades EAI 2, 3&5 subclades together which is farther from EAI 1&4 subgroups, represented poorly in Thiruvallur. EAI 3 and 5 are present as predominant sub clades in Thiruvallur. More number of EAI 1& 4 sub clades are present in Denmark, Netherlands, USA, Bangladesh and Vietnam (SpolDB4) EAI 1 & 4 are grouped closer to Harlem, CAS and LAM 8&9.

Spoligotyping and IS6110 RFLP

Among the 1192 isolates which had been typed by IS6110, 20 isolates had no copy. Four hundred and ninety three isolates had single copy, 249 isolates had 2-5 copies. Four hundred and thirty isolates were IS6110 high copy isolates. Among the 20 no copy IS6110 strains, 14 had unique spoligotype patterns and remaining 6 isolates were in 3 clusters of 2 each. Thus among the no copy IS6110 strains, spoligotyping has been highly discriminatory (17 types among 20 strains). We tried to find out whether the major spoligotype clusters could be differentiated by IS6110 RFLP. Table 2 & 3 show the typed clusters (>2 strains in a cluster) and orphan clusters (< than 3 isolates in a cluster) and the degree of differentiation by IS6110 RFLP. Among the total 733 isolates involved in clusters with types matching the database 332 strains (45%) had a single copy, 195 (26.6%) showed low copy strains and 8 (1.0%) showed no copy isolates. Totally 63% of strains could not be discriminated by both spoligotype and IS6110. Similarly among the isolates involved in orphan clusters, all the spoligotype clusters (except 2) had isolates with low copy IS6110 strains. The East African variants or the ancient TB lineage forms the predominant group among the south Indian isolates and 63% of the EAI isolates are low copy isolates. The TbD1 ancestral isolates belonging to EAI family were less discriminatory by IS6110 RFLP typing. 76% of the isolates belonging to the typed and orphan clusters by spoligotyping are low copy IS6110 isolates (Tables 2, 3).

Genomic deletions detected by comparative whole-genome hybridization

To identify genomic differences putatively associated with the altered virulence, we performed whole-genome hybridization on a subset of 25 strains belonging to the dominant ancestral lineage using the Affymetrix *M. tuberculosis* GeneChip™ (Santa Clara, CA, USA) and published methods (Gingeras et al., 1998, Kato-Maeda et al., 2001, Tsolaki et al., 2004, 2005). We confirmed putative deletions by PCR and sequencing as previously reported (Chen 2006, Gagneux et al., 2006, Glynn et al., 2002, Gutacker et al., 2006).

In the 25 strains, we identified a total of 150 putative deletions, including 38 that involved highly repetitive PE-/PPE-PGRS sequences and which were excluded from further analysis because of the difficulties in precisely determining deletion boundaries (Tsolaki et al., 2004). Analysis of the remaining 112 putative deletions determined that 31 (27.6%) were false

positives and 81 (72.3%) were true deleted sequences. The locations of the true deletions were mapped to the base pair and their sequences were named regions of difference (RD) with respect to H37Rv (Brosch et al., 2002, Cole et al., 1998).

All 25 strains had a genomic deletion is RD239 (as expected), RD147C and RD198a. These three large sequence polymorphisms (LSPs) have been previously reported (Gagneux et al., 2006, Tsolaki et al., 2005). In addition, we found six new LSPs (Table 4). To confirm that the RD239-deleted strains from south India belonged to the ancestral lineage of *M. tuberculosis*, we used multiplex real-time PCR to screen for TbD1 (Brosch et al., 2002) in the 25 strains following published procedures (Gagneux et al., 2006). As expected, we found that the TbD1 region was intact in all 25 strains.

To estimate the frequency of the six newly identified LSPs (RD768-RD773), 100 isolates belonging to the ancestral lineage were chosen from the largest spoligotype clusters and screened by PCR. We found that RD769 and RD771 were each detected in only 2 of the 100 isolates, and RD768, RD770, RD772 and RD773 were not deleted in any of the 100 isolates. Earlier studies have reported that 40% of the strains in south India are low virulent in guinea pigs (Prabhakar et al., 1987). Because of the low frequency of these LSPs, these genomic alterations are unlikely to account for the low virulence of south Indian strains in animal models.

Ancestral strain lineage predominates in south India

To define the population structure of *M. tuberculosis* in the study area, we selected one isolate per spoligotype pattern (102 isolates) and screened by multiplex real-time PCR for the main strain lineages using the lineage-specific DNA polymorphisms (Gagneux 2006). We first tested for the presence or absence of the genomic region of difference (RD) 105 and RD239 using the primers and probes published in (Gagneux 2006). RD105 and RD239 are markers for the East Asian (“W-Beijing”) and ancestral strain lineages, respectively (Gagneux 2006, Tsolaki 2005). The screening results from the clustered isolates were extrapolated to the remaining isolates of their respective clusters. Based on this first round of screening, we found that 85.2% (1,035/1,215) of the isolates had a deletion in RD239, 2.3% (28/1,215) of isolates had a deletion in RD105, 11.8% (143/1,215) of isolates had no deletion, and the genomic deletions in 0.7% (9/1,215) of isolates remained undetermined (Table 5). We performed a second round of screening on the isolates without a confirmed deletion in RD105 or RD239 using a TaqMan allelic discrimination assay for the *katG*463 ctg to cgg single nucleotide polymorphism, which defines the Euro-American lineage (Principal Genetic Groups 2 and 3) (Gagneux 2006). We found 6.8% (83/1,215) of isolates belonged to this lineage. Finally, close inspection of the spoligotypes of the remaining isolates (Supplementary Table) revealed that 4.4% (54/1,215) belonged to the CAS lineage or the Delhi genogroup (Baker et al., 2004, Brudey et al., 2006).

DISCUSSION

For decades there has been a belief that the outcome of a person's exposure to and infection with *M. tuberculosis* depends only on individual host characteristics such as age, gender and genetic background. However recent studies increasingly suggest that bacterial factors also

contribute to the differences in outcomes seen in the human host population. Hence it becomes imperative to characterize the clinical isolates of *M. tuberculosis* from a large population based study in order to design vaccine and new drugs for the population. One quarter of the world's TB cases occur in India and the major lineage which has been identified in this study is likely to be responsible for a high percentage of the TB cases.

This study describes the spoligotyping of 1215 isolates of *M. tuberculosis* from Tiruvallur area in South India where the largest BCG trial was conducted. The total population of the study area has been 580,000. The incidence of smear positive TB is 76 per 100,000 in the population. We have compared the spoligotyping with the widely accepted IS6110 RFLP typing from a 3 year population based study. In addition to these two methods Deletion Micro array based large sequence polymorphism (LSP), deletion PCR and RT PCR were used as large scale screening methods to delineate the phylogeny and evolutionary characteristics of the south Indian isolates of *M. tuberculosis*. To our knowledge this is the first study from India which combines deletion micro array along with the widely used genotyping tools in a population based study.

The genetic diversity observed in our study has been 22.5% which is corroborating with the fact that endemic area might have relatively few circulating strains (Hermans et al., 1995). Ours is a population based study and restricted to southern most region of the country. Hence this data contradicts the study of (Singh et al., 2007) which included 7 different regions of India and showed a large number of circulating strains. This highlights the rich diversity among different regions of India.

Spoligotyping has been a very useful technique to differentiate the *M. tuberculosis* isolates into different geographical clades. It is not a highly discriminatory marker. Studies from parts of the globe have highlighted the predominant spoligotypes responsible for most of the pulmonary TB cases (Brudey et al., 2003, Dale et al., 1999, Puustinen et al., 2003). These studies reveal that different spoligotypes occur at variable frequencies in different countries and continents.

The CAS or Delhi type corresponding to ST26 which predominates in North India (Bhanu et al., 2002, Singh et al., 2004, Kulkarni et al., 2005, Guitierrez *et al.*, 2006) represents 1.02 % of isolates SpolDB4 database. This type has been reported from 34 countries. None of the orphan strains identified by Kulkarni et al., 2005 or Singh U.B. et al., 2007) were found in our study. Spoligotyping has categorically identified the major clade present in south India to East African lineage with 28,32 and 34 spacers missing. This clearly shows the difference in origin and evolution of *M. tuberculosis* in North and South of India. In our study the EAI family corresponding to ubiquitous spoligotype ST11, is present as a large group among the clusters matching the SpolDB4. 336 isolates among 1215 isolates belong to this group. This family is present in other countries like Bangladesh, Malaysia, Indonesia, Pakistan, UK, Denmark, Netherlands, France, New Zealand and USA.

The family W-Beijing which has been reported from several countries has a distinct genotype has been associated with rapid transmission across large parts of different

continents, accounted to only 1.9% in Thiruvallur. The other clades have been represented by a fewer percentage in Thiruvallur (Table 1).

Sixteen strains of type 53 were seen in the present study from Tiruvallur. This type is linked to be a major clade of European descent in a multicentric study (Kremer et al., 1999). The presence or absence of spacer31 discriminates between the two highly prevalent spoligotypes worldwide, ST 50 & ST 53 with its absence being linked to IS6110 insertional events. While Mumbai study (Kulkarni et al., 2005) showed the presence of one strain of ST 50, we have seen ST 53 in the south Indian region.

In spite of the endemicity of the disease a large proportion of the patients were infected with the tuberculosis isolates having spoligotype that had not been described in the global database, of more than 39000 tuberculosis isolates. But there has been a major cluster of isolates constituting 85% of the total isolates from this endemic area. Twelve major clusters ST 1, 8, 11, 126, 340, 43, 48, 591, 236, 26, 355, 53 comprised 86%. Except 1 & 53 all the remaining clusters are variants of EAI with the characteristic spoligotype pattern with absence of spacers 28-32 and 34 which may be considered as a major type in the Tiruvallur district.

Overall the spoligotyping analysis showed diversity in clinical isolates especially among the strains with no copy of IS6110. The 20 no copy isolates were differentiated to 17 different spoligotypes. 30% of the IS6110 low copy strains were differentiated by spoligotyping. Spoligotyping alone is not sufficient as a secondary marker for IS6110 low copy strains, instead this genotyping method has been an excellent tool to define the geno clade prevalent in the Tiruvallur region of South India in comparison with the other regions of the globe. This region has been predominated by IS6110 low copy strains. 63% the EAI variants are comprised of single and low copy IS6110 isolates. As tables 5 & 6 show that major spoligo clusters could not be differentiated by IS6110 typing because of the predominance of low copy isolates. This warrants the use of other secondary markers like VNTR, MIRU besides spoligotyping, and other PCR based methods to study the transmission dynamics of *M. tuberculosis* in the region. Even though the discriminating capacity of MIRU is not greater than IS6110, it has been shown to be efficient to discriminate the low IS6110 strains (Sun et al., 2004).

The excellent congruence observed between all the independent sets of genetic markers used here lends strong support to the assignment of different prevalent lineages. Our LSP followed by large scale screening by RT PCR and PCR has confirmed the prevalent genotypic clade in this region. The presence of TbD1 in a large percentage of isolates indicates that S. India could have been the initial focus of the ancient lineage. Another characteristic feature of these strains is the deletion of RD239 region which is also a characteristic feature of the Manila strains (Tsolaki, et al., 2004). Aggregation of 236–239 deletion is confined to the so-called “Manila” clade (Tsolaki, et al., 2004), and is likely to result from a genetic event that occurred in the ancestor of these isolates. The spoligotyping of 41 of the 48 *M. tuberculosis* isolates from metropolitan Manila yielded characteristic identical patterns lacking hybridization to eight spacers (Bhanu et al., 2002) the remaining 35 spacers showed a positive hybridization signal.

A similar systematic association has been observed in strains from Singapore (Sun et al., 2004) and from Bangladesh (Sola et al., 2001). It remains to be tested whether the deletion of TbD1 from the genome of ancient strains conferred or not some selective advantage to the descendent strains enhancing their epidemic potential. Due to trade links the probable route of migration to and fro from south India to East Africa could have been through East Asian countries like Philipines.

This study is population based and has been spread over 3 years. There have also been studies from this region describing the prevalence of low copy IS6110 strains in the past (Das et al 1995) In spite of the presence of all the 9 clades in this region the only lineage which has been predominating in this region is the EAI clade. Recently there have been reports on the notion that *M. tuberculosis* has adapted to human population. In this region the ancient genotype or the EAI strain has adapted to the host during several centuries and thus has a capacity for higher transmission. The prevalence of such a high number in this part of the continent and also its prevalence for a long time raise the question whether these have a unique transmission dynamics and are they suppressing the spread of other genotypes. In the largest BCG trial conducted in the same region earlier, BCG did not afford any protection. It is known that genetic variation of the strains is a method of vaccine escape and has been demonstrated in several bacterial species (Van Loo et al., 2002, Mbelle et al., 1999). Further investigation in this line would be helpful to ascertain the role of these highly prevalent strains on the host immunity.

Starting 50 years ago, studies using the guinea pig and other animal models have repeatedly reported that strains of *M. tuberculosis* from south India were less virulent than other strains (Mitchison et al., 1960, Singh et al., 1964, Prabhakar et al 1987). Furthermore, BCG vaccine trials in south India showed no significant efficacy (TRC, 1999). Considering the global phylogeography of *M. tuberculosis* and the dominance of the ancestral lineage in south India, BCG failure in this part of the world could partially be due to strain-specific effects (Brewer et al., 1995). Based on our approach, which was designed to detect genomic regions that are absent relative to the laboratory reference strain, H37Rv, deletion array was performed.

Deletion micro array followed by large scale screening by PCR had identified six new deletions in the genome of only 1-2% of the south Indian isolates. The phenotypic characteristics lost or gained by these deletions have to be explored. However, considering that this strain lineage is ancestral to all *M. tuberculosis* strains that have been sequenced thus far (Gagneux et al., 2006), this lineage could harbour unique sets of genes that are absent (i.e. deleted) in other *M. tuberculosis* and thus not detectable using H37Rv as a reference.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Baker L, Brown T, Maiden MC, Drobniewski F. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. *Emerg. Infect. Dis.* 2004; 10:1568–1577. [PubMed: 15498158]
- Bhanu NV, van Soolingen D, van Embden JD, Dar L, Pandey RM, Seth P. Predominance of a novel *Mycobacterium tuberculosis* genotype in the Delhi region of India. *Tuberculosis*. 2002; 82:105–112. [PubMed: 12356462]
- Bhatia AL. Characteristics of Indian tubercle bacilli. *Ind. J. Chest Dis.* 1961; 3:147–153.
- Brewer TF, Colditz GA. Relationship between bacille Calmette-Guerin (BCG) strains and the efficacy of BCG vaccine in the prevention of tuberculosis. *Clin. Infect. Dis.* 1995; 20:126–135. [PubMed: 7727638]
- Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci.* 2002; 99:3684–3689. [PubMed: 11891304]
- Brudey K, Filliol I, Sola C, Bebear C, Elia-Pasquet S, Texier-Maugein J, Rastogi N. Molecular characterization and biodiversity of *Mycobacterium tuberculosis* in the Antilles-Guiana region and comparative analysis in a metropolitan region, Aquitaine. *Pathol. Biol.* 2003; 51:282–289. [PubMed: 14567196]
- Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajj SA, Allix C, Aristimuno L, Arora J, Baumanis V, Binder L, Cafrune P, Cataldi A, Cheong S, Diel R, Ellermeier C, Evans JT, Fauville-Dufaux M, Ferdinand S, Garcia de Viedma D, Garzelli C, Gazzola L, Gomes HM, Gutierrez MC, Hawkey PM, van Helden PD, Kadival GV, Kreiswirth BN, Kremer K, Kubin M, Kulkarni SP, Liens B, Lillebaek T, Ho ML, Martin C, Martin C, Mokrousov I, Narvskaja O, Ngeow YF, Naumann L, Niemann S, Parwati I, Rahim Z, Rasolofo-Razanamparany V, Rasolonavalona T, Rossetti ML, Rusch-Gerdes S, Sajduda A, Samper S, Shemyakin I, Singh U, Somoskovi A, Skuce R, van Soolingen D, Streicher E, Suffys PN, Tortoli E, Tracevska T, Vincent V, Victor TC, Warren R, Yap SF, Zaman K, Portaels F, Rastogi N, Sola C. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol.* 2006; 6:6–23. [PubMed: 16448556]
- Chen I. Thinking big about global health. *Cell.* 2006; 124:661–663. [PubMed: 16497574]
- Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE III, Tekaiia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream M-A, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead SC, Barrell BG. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature.* 1998; 393:537–544. [PubMed: 9634230]
- Collins FM, Smith MM. A comparative study of the virulence of *Mycobacterium tuberculosis* measured in mice and guinea pigs. *Am. Rev. Respir. Dis.* 1969; 100:631–639. [PubMed: 4987090]
- Constant P, Perez E, Malaga W, Laneelle MA, Saurel O, Daffe M, Guilhot C. Role of the pks15/1 gene in the biosynthesis of phenolglycolipids in the *Mycobacterium tuberculosis* complex. Evidence that all strains synthesize glycosylated p-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the pks15/1 gene. *J. Biol. Chem.* 2002; 277:38148–38158. [PubMed: 12138124]
- Cowan LS, Diem L, Brake MC, Crawford JT. Transfer of a *Mycobacterium tuberculosis* genotyping method, Spoligotyping, from a reverse line-blot hybridization, membrane-based assay to the Luminex multianalyte profiling system. *J. Clin. Microbiol.* 2004; 42:474–477. [PubMed: 14715809]
- Dale JW, Nor RM, Ramayah S, Tang TH, Zainuddin ZF. Molecular epidemiology of tuberculosis in Malaysia. *J. Clin. Microbiol.* 1999; 37:1265–1268. [PubMed: 10203468]
- Dale JW, Brittain D, Cataldi AA, Cousins D, Crawford JT, Driscoll J, Heersma H, Lillebaek T, Quitugua T, Rastogi N, Skuce RA, Sola C, van Soolingen D, Vincent V. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardised nomenclature. *Int. J. Tuberc. Lung Dis.* 2001; 5:216–219. [PubMed: 11326819]

- Das S, Paramasivan CN, Lowrie DB, Prabhakar R, Narayanan PR. IS6110 restriction fragment length polymorphism typing of clinical isolates of *Mycobacterium tuberculosis* from patients with pulmonary tuberculosis in Madras, South India. *Tuberc. Lung Dis.* 1995; 76:550–554.
- Das SD, Narayanan S, Hari L, Hoti SL, Thangathurai RK, Charles N, Jaggarajamma K, Narayanan PR. Differentiation of highly prevalent IS6110 single-copy strains of *Mycobacterium tuberculosis* from a rural community in South India with an ongoing DOTS programme. *Infect. Genet. Evol.* 2005; 5:67–77. [PubMed: 15567140]
- Filliol I, Driscoll JR, Van Soolingen D, Kreiswirth BN, Kremer K, Valetudie G, Anh DD, Barlow R, Banerjee D, Bifani PJ, Brudey K, Cataldi A, Cooksey RC, Cousins DV, Dale JW, Dellagostin OA, Drobniewski F, Engelmann G, Ferdinand S, Gascoyne-Binzi D, Gordon M, Gutierrez MC, Haas WH, Heersma H, Kallenius G, Kassa-Kelembho E, Koivula T, Ly HM, Makristathis A, Mammina C, Martin G, Mostrom P, Mokrousov I, Narbonne V, Narvskaya O, Nastasi A, Niobe-Eyangoh SN, Pape JW, Rasolofo-Razanamparany V, Ridell M, Rossetti ML, Stauffer F, Suffys PN, Takiff H, Texier-Maugein J, Vincent V, De Waard JH, Sola C, Rastogi N. Global distribution of *Mycobacterium tuberculosis* spoligotypes. *Emerg. Infect. Dis.* 2002; 8:1347–1349. [PubMed: 12453368]
- Filliol I, Driscoll JR, van Soolingen D, Kreiswirth BN, Kremer K, Valetudie G, Dang DA, Barlow R, Banerjee D, Bifani PJ, Brudey K, Cataldi A, Cooksey RC, Cousins DV, Dale JW, Dellagostin OA, Drobniewski F, Engelmann G, Ferdinand S, Gascoyne-Binzi D, Gordon M, Gutierrez MC, Haas WH, Heersma H, Kassa-Kelembho E, Ho ML, Makristathis A, Mammina C, Martin G, Mostrom P, Mokrousov I, Narbonne V, Narvskaya O, Nastasi A, Niobe-Eyangoh SN, Pape JW, Rasolofo-Razanamparany V, Ridell M, Rossetti ML, Stauffer F, Suffys PN, Takiff H, Texier-Maugein J, Vincent V, de Waard JH, Sola C, Rastogi N. Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. *J. Clin. Microbiol.* 2003; 41:1963–1970. [PubMed: 12734235]
- Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, Hilty M, Hopewell PC, Small PM. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci.* 2006; 103:2869–2873. [PubMed: 16477032]
- Gingeras TR, Ghandour G, Wang E, Berno A, Small PM, Drobniewski F, Alland D, Desmond E, Holodniy M, Drenkow J. Simultaneous genotyping and species identification using hybridization pattern recognition analysis of generic *Mycobacterium* DNA arrays. *Genome Res.* 1998; 8:435–448. [PubMed: 9582189]
- Glynn JR, Whiteley J, Bifani PJ, Kremer K, van Soolingen D. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. *Emerg. Infect. Dis.* 2002; 8:843–849. [PubMed: 12141971]
- Grange JM, Aber VR, Allen BW, Mitchison DA, Goren MB. The correlation of bacteriophage types of *Mycobacterium tuberculosis* with guinea pig virulence and *in vitro* indicators of virulence. *Tubercle.* 1978; 45:354–359.
- Gutierrez MC, Ahmed N, Willery E, Narayanan S, Hasnain SE, Chauhan DS, Katoch VM, Vincent V, Loch C, Supply P. Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India. *Emerg. Infect. Dis.* 2006; 12:1367–1374. [PubMed: 17073085]
- Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN, Graviss EA, Musser JM. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. *J Infect Dis.* 2006; 193:121–128. [PubMed: 16323140]
- Gutacker MM, Smoot JC, Migliaccio CA, Ricklefs SM, Hua S, Cousins DV, Graviss EA, Shashkina E, Kreiswirth BN, Musser JM. Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics.* 2002; 162:1533–1543. [PubMed: 12524330]
- Hermans PW, Messadi F, Guebrexabher H, van Soolingen D, de Haas PE, Heersma H, de Neeling H, Ayoub A, Portaels F, Frommel D. Analysis of the population structure of *Mycobacterium tuberculosis* in Ethiopia, Tunisia, and The Netherlands: usefulness of DNA typing for global tuberculosis epidemiology. *J. Infect. Dis.* 1995; 171:1504–1513. [PubMed: 7769285]

- Joseph S, Mitchison DA, Ramachandran K, Selkon JB, Subbaiah TV. Virulence in the guinea-pig and sensitivity to pas and thiacetazone of tubercle bacilli from south Indian patients with pulmonary tuberculosis. *Tubercle*. 1964; 45:354–359. [PubMed: 14239317]
- Kato-Maeda M, Rhee JT, Gingeras TR, Salamon H, Drenkow J, Smittipat N, Small P. Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res*. 2001; 11:1796–1799.
- Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol*. 1997; 35:907–914.
- Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PW, Martin C, Palittapongarnpim P, Plikaytis BB, Riley LW, Yakus MA, Musser JM, van Embden JD. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. *J. Clin. Microbiol*. 1999; 37:2607–2618. [PubMed: 10405410]
- Kulkarni S, Sola C, Filliol I, Rastogi N, Kadival G. Spoligotyping of *Mycobacterium tuberculosis* isolates from patients with pulmonary tuberculosis in Mumbai, India. *Res. Microbiol*. 2005; 156:588–596. [PubMed: 15862459]
- Marmiesse M, Brodin P, Buchrieser C, Gutierrez C, Simoes N, Vincent V, Glaser P, Cole ST, Brosch R. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiol*. 2004; 150:483–496.
- Mbelle N, Huebner RE, Wasas AD, Kimura A, Chang I, Klugman KP. Immunogenicity and impact on nasopharyngeal carriage of a nonavalent pneumococcal conjugate vaccine. *J. Infect. Dis*. 1999; 180:1171–1176. [PubMed: 10479145]
- Middlebrook G, Cohn ML. Some observations on the pathogenicity of isoniazid-resistant variants of tubercle bacilli. *Science*. 1953; 118:297–299. [PubMed: 13089681]
- Mistry NF, Iyer AM, D'souza DT, Taylor GM, Young DB, Antia NH. Spoligotyping of *Mycobacterium tuberculosis* isolates from multiple-drug-resistant tuberculosis patients from Bombay, India. *J. Clin. Microbiol*. 2002; 40:2677–2680. [PubMed: 12089307]
- Mitchison DA, Wallace JG, Bhatia AL, Selkon JB, Subbaiah TV, Lancaster MC. A comparison of the virulence in guinea-pigs of South Indian and British tubercle bacilli. *Tubercle*. 1960; 41:1–22. [PubMed: 14423002]
- Narayanan S, Sahadevan R, Narayanan PR, Krishnamurthy PV, Paramasivan CN, Prabhakar R. Restriction fragment length polymorphism of *Mycobacterium tuberculosis* strains from various regions of India, using direct repeat probe. *Indian J. Med. Res*. 1997; 106:447–454. [PubMed: 9415738]
- Narayanan S, Das S, Garg R, Hari L, Rao VB, Frieden TR, Narayanan PR. Molecular epidemiology of tuberculosis in a rural area of high prevalence in South India: implications for disease control and prevention. *J. Clin. Microbiol*. 2002; 40:4785–4788. [PubMed: 12454197]
- Prabhakar R, Venkataraman P, Vallishayee RS, Reeser P, Musa S, Hashim R, Kim Y, Dimmer C, Wiegshauss E, Edwards ML. Virulence for guinea pigs of tubercle bacilli isolated from the sputum of participants in the BCG trial, Chingleput District, South India. *Tubercle*. 1987; 68:3–17. [PubMed: 3116731]
- Puustinen K, Marjamäki M, Rastogi N, Sola C, Filliol I, Ruutu P, Holmstrom P, Viljanen MK, Soini J. Characterization of Finnish *Mycobacterium tuberculosis* isolates by spoligotyping. *J. Clin. Microbiol*. 2003; 41:1525–1528. [PubMed: 12682140]
- Radhakrishnan I, Manju YK, Kumar RA, Mundayoor S, Harris. Implications of low frequency of IS6110 in fingerprinting field isolates of *Mycobacterium tuberculosis* from Kerala, India. *J. Clin. Microbiol*. 2001; 39:1683. [PubMed: 11354035]
- Singh B, Sharma MD. Niacin production and virulence in the guinea-pig of Indian tubercle bacilli. *Tubercle*. 1964; 45:239–240. [PubMed: 14227812]
- Singh UB, Suresh N, Bhanu NV, Arora J, Pant H, Sinha S, Aggarwal RC, Singh S, Pande JN, Sola C, Rastogi N, Seth P. Predominant tuberculosis spoligotypes, Delhi, India. *Emerg. Infect. Dis*. 2004; 10:1138–1142. [PubMed: 15207071]

- Singh UB, Arora J, Suresh N, Pant H, Rana T, Sola C, Rastogi N, Pande JN. Genetic biodiversity of *Mycobacterium tuberculosis* isolates from patients with pulmonary tuberculosis in India. *Infect. Genet. Evol.* 2007; 7:441–448. [PubMed: 17317334]
- Sola C, Filliol E, Legrand E, Mokrousov I, Rastogi N. *Mycobacterium tuberculosis* phylogeny reconstruction based on combined numerical analysis with IS1081, IS6110, VNTR and DR-based spoligotyping suggests the existence of two new phylogeographical clades. *J. Mol. Evol.* 2001; 53:680–689. [PubMed: 11677628]
- Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl. Acad. Sci.* 1997; 94:9869–9874. [PubMed: 9275218]
- Sun YJ, Lee AS, Ng ST, Ravindran S, Kremer K, Bellamy R, Wong SY, van Soolingen D, Supply P, Paton NI. Characterization of ancestral *Mycobacterium tuberculosis* by multiple genetic markers and proposal of genotyping strategy. *J. Clin. Microbiol.* 2004; 42:5058–5064. [PubMed: 15528696]
- Streicher EM, Warren RM, Kewley C, Simpson J, Rastogi N, Sola C, van der Spuy GD, van Helden PD, Victor TC. Genotypic and phenotypic characterization of drug-resistant *Mycobacterium tuberculosis* isolates from rural districts of the Western Cape Province of South Africa. *J. Clin. Microbiol.* 2004; 42:891–894. [PubMed: 14766882]
- Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, Goguet de la Salmoniere, Y.-O.L. Aman K, Kato-Maeda M, Small PM. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proc. Natl. Acad. Sci.* 2004; 101:4865–4870. [PubMed: 15024109]
- Tsolaki AG, Gagneux S, Pym AS, Goguet de la Salmoniere, Y.-O.L. Kreiswirth BN, van Soolingen D, Small PM. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 2005; 43:3185–3191. [PubMed: 16000433]
- Indian J. Med. Res. Vol. 110. Tuberculosis Research Centre (ICMR); Chennai: 1999. Fifteen year follow up of trial of BCG vaccines in south India for tuberculosis prevention.; p. 56-69.
- van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J. Clin. Microbiol.* 1993; 31:406–409. [PubMed: 8381814]
- Van Loo IH, Mooi FR. Changes in the Dutch *Bordetella pertussis* population in the first 20 years after the introduction of whole-cell vaccines. *Microbiol.* 2002; 148:2011–2018.

Fig. 1.
Clustered Spoligo types matching Spol DB 4.0

Fig. 2.
Orphan strains in cluste

Table - 1 ASpoligotyping of *M. tuberculosis* strains from Tiruvallur, Tamilnadu

Spoligotype	Isolates	Clusters	Unique
EAI 1	5	4	1
EAI 2	20 (1.6%)	8	12
EAI 3	498 (40.9%)	448	50
EAI 4	5	4	1
EAI 5	498 (40.9%)	406	92
Beijing	24 (1.9%)	23	1
CAS	48 (3.9%)	34	14
Harlem 1	12 (0.9%)	7	5
LAM 7	1	0	1
LAM 8	7	2	5
LAM 9	6	2	4
T1	36 (2.9%)	22	14
T2	2	2	0
X2	1	0	1
X3	1	0	1
Family 33	14 (1.1%)	10	4
Family 36	3	2	1
H37RV	3	2	1
Microti	1	0	1
Orphan	30 (2.4%)	16	14
Total	1215	992 (81.6%)	223 (14.4%)

I B

Phylogenetic analysis of *M. tuberculosis* - Spoligotype

Table - 2

Orphan clusters and IS6110 patterns

Sl.No.	Orphan clusters	No. of Isolates	IS single copy	IS low copy	IS high copy	No copy
1.	467 777 777 413 071	15	14	1	0	0
2.	400 000 377 413 071	14	11	1	1	1
3.	477 777 777 413 171	14	3	7	4	0
4.	476 377 777 413 771	13	7	5	1	0
5.	477 777 377 413 071	9	7	2	0	0
6.	077 777 777 400 071	8	0	2	6	0
7.	777 737 757 413 771	8	0	1	7	0
8.	777 777 774 000 771	7	0	0	7	0
9.	477 777 777 413 011	8	7	0	0	1
10.	417 777 777 413 771	5	4	1	0	0
11.	477 773 177 413 771	5	5	0	0	0
12.	401 177 777 413 071	4	4	0	0	0
13.	473 777 077 410 771	4	1	3	0	0
14.	737 777 747 413 771	4	1	3	0	0
15.	777 777 777 733 771	4	0	1	3	0

Table - 3

Typed Spoligo clusters and IS6110 RFLP pattern

Sl.No.	Spoligotype	No. of Isolates	IS single copy	IS low (2-5 copy)	IS high copy	No copy
1.	11	333	144	129	58	3
2.	126	80	59	69	3	2
3.	340	75	48	19	7	1
4.	43	40	5	1	34	0
5.	48	37	26	5	6	0
6.	591	31	2	2	27	0
7.	355	19	14	3	2	0
8.	236	24	9	2	11	2
9.	26	20	0	2	18	0
10.	53	16	2	1	13	0
11.	8	18	13	4	1	0
12.	473	5	0	5	0	0
13.	474	8	6	2	0	0
14.	288	4	0	0	4	0
15.	25	4	0	0	4	0
16.	138	5	2	1	2	0
17.	100	6	2	3	0	1
18.	372	4	0	0	4	0
19.	124	4	0	0	4	0

Table - 4

Description of six new large sequence polymorphisms (LSPs) or genomic deletions observed in the clinical isolates from South India

Isolate	Name *	Coordinates		Primer sequence	Size [†] (bp)	Putative function of deleted genes and base pairs
		start	stop			
M394	RD768	70289	72023	GGG GCG GCT GTT GGA CCC GCA TAT CCT ATC AAG ACC GGT AAC GAG CGG GCC AAA CTC	1735	Rv0064 Trans membrane protein Rv0065 Conserved hypothetical protein
M357	RD769	1957030	1959686	GAC AGC AAC CGC GAC GCC CGG AAT C CCC GCC CTC GTC GTC ACC TTC ATC TGT AA	2657	Rv1730c Pencillin binding protein Rv1731 (<i>gabD2</i>) Succinate-semialdehyde dehydrogenase (NADP+) dependent (SSDH) <i>gabD2</i> Rv1732c Conserved hypothetical protein
M395	RD770	2196902	2200337	CCG GTG ACC GTC GTG GTG AGC ACC A CCA GGA CGG AGG TCA CAG TTG CGG GGT	3436	Rv1946 (<i>IppG</i>) Possible lipoprotein Rv1947 Hypothetical protein Rv1948 Hypothetical protein
M118	RD771	89500	90450	CCG GGC GCG CGA ACA TGG ACT GC GGC TCG GCG CCT CCG GGT GG	951	Rv0081 Probable transcriptional regulatory protein Rv0082 Probable oxidoreductase Rv0083 Probable oxidoreductase
M165	RD772	30669	34074	GCC ATC GCG GAG GCG GAA GCA GCT CT TTT GCC CGG CCT AGC GGT TGC CCA TC	3406	Rv0027 Conserved hypothetical protein Rv0028 Conserved hypothetical protein Rv0029 Conserved hypothetical protein Rv0030 Conserved hypothetical protein Rv0031 Possible remnant of a transposase
M461	RD773	3434523	3441337	CGG CCC TGA CGG TGG CAA TCT GGA TC GAG CAG GGT CGC CAG CCA GTT GCC	6815	Rv3071-Rv3076 Conserved hypothetical Rv3077 Possible hydrolase Rv3078 (<i>hab</i>) Probable hydroxylaminobenzene mutase <i>Hab</i> Rv3079c Conserved hypothetical protein Rv3080(<i>pknK</i>) Probable serine/threonine-protein kinase transcriptional regulatory protein <i>pknK</i> (protein kinase k)

bp = base pairs

* Name assigned to the specific genomic deletion

[†] Size of genomic deletion (bp) = (stop coordinates – start coordinates)

Table - 5

Results of real-time pCR screening of one isolate per strain, as defined by spoligotyping

Genomic deletion RD105	Genomic deletion RD239	Number of isolates	Percentage of isolates
Deleted	Present	28	(13.0)
Present	Present	143	(11.8)
Deleted	deleted	4	(0.3)
Present	Deleted	1035	(85.2)
Indeterminate	Indeterminate	5	(0.4)
Total	Total	1215	(100.0)