Comparison of spoligotyping, mycobacterial interspersed repetitive units typing and IS*6110*-RFLP in a study of genotypic diversity of *Mycobacterium tuberculosis* in Delhi, North India

Mandira Varma-Basil[1], Sujeet Kumar[1], Jyoti Arora[2], Archana Angrup[1], Thierry Zozio[3], Jayant Nagesh Banavaliker[4], Urvashi Balbir Singh[2], Nalin Rastogi[3], Mridula Bose[1]/+

[1]Deptartment of Microbiology, Vallabhbhai Patel Chest Institute, University of Delhi, Delhi, India [2]Deptartment of Microbiology, All India Institute of Medical Sciences, Ansari Nagar, New Delhi, India [3]WHO Supranational TB Reference Laboratory, Unité de la Tuberculose et des Mycobactéries, Institut Pasteur de Guadeloupe, Guadeloupe, France [4]Rajan Babu Tuberculosis Hospital, Kingsway Camp, Delhi, India

The aim of the present study was to compare polymerase chain reaction (PCR)-based methods - spoligotyping and mycobacterial interspersed repetitive units (MIRU) typing - with the gold-standard IS6110 restriction fragment length polymorphism (RFLP) analysis in 101 isolates of Mycobacterium tuberculosis *to determine the genetic diversity of* M. tuberculosis *clinical isolates from Delhi, North India. Spoligotyping resulted in 49 patterns (14 clusters); the largest cluster was composed of Spoligotype International Types (SITs)26 [Central-Asian (CAS)1-Delhi lineage], followed by SIT11 [East-African-Indian (EAI) 3-Indian lineage]. A large number of isolates (75%) belonged to genotypic lineages, such as CAS, EAI and Manu, with a high specificity for the Indian subcontinent, emphasising the complex diversity of the phylogenetically coherent* M. tuberculosis *in North India. MIRU typing, using 11 discriminatory loci, was able to distinguish between all but two strains based on individual patterns. IS6110-RFLP analysis (n = 80 strains) resulted in 67 unique isolates and four clusters containing 13 strains. MIRUs discriminated all 13 strains, whereas spoligotyping discriminated 11 strains. Our results validate the use of PCR-based molecular typing of* M. tuberculosis *using repetitive elements in Indian isolates and demonstrate the usefulness of MIRUs for discriminating low-IS6110-copy isolates, which accounted for more than one-fifth of the strains in the present study.*

Key words: tuberculosis - *M. tuberculosis* - genotyping - spoligotyping - IS*6110*-RFLP

Mycobacterium tuberculosis infection is widespread and is one of the most successful human pathogens today. The propagation success of this disease remains directly linked to the social and hygienic conditions of human populations (Rastogi et al. 2007). One-fifth of the world's new tuberculosis (TB) cases and two-thirds of cases in the South East Asian region occur in India (RNTCP Status Report 2008). By working in a country with one of the highest TB burdens in the world, those attempting to control TB in India face new challenges such as the emergence of both multidrug-resistant and extensively drug resistant TB and a high incidence of human immunodeficiency virus (HIV)/TB coinfection.

The introduction of molecular techniques to the field of TB has provided new ways to study dissemination dynamics and evolutionary genetics of the pathogen and data collected using these methods could eventually impact TB control measures. IS*6110* restriction fragment length polymorphism (RFLP)-based fingerprinting has

been used to study the mycobacterial population structure in Southern India, Northern India and the Delhi region (Das et al. 1995, Radhakrishnan et al. 2001, Siddiqi et al. 2001, Bhanu et al. 2002). However, IS*6110* fingerprinting is of limited use because a significant proportion (40-44%) of *M. tuberculosis* isolates from several regions of India have been reported to either lack or have low copy numbers of IS*6110* (Das et al. 1995, Radhakrishnan et al. 2001, Narayanan et al. 2002, Chauhan et al. 2004, Mathuria et al. 2008). Furthermore, IS*6110* typing is labour intensive and requires several weeks for culturing the *M. tuberculosis* isolates. Fingerprinting methods targeting polymorphic spacer sequences in the direct repeat (DR) region, including spoligotyping, have been used in some regions in India (Mistry et al. 2002, Singh et al. 2004, 2007, Kulkarni et al. 2005, Mathuria et al. 2008, Narayanan et al. 2008, Sharma et al. 2008). However, when used alone, these methods may underestimate the clonal diversity of *M. tuberculosis* (Kremer et al. 1999). Molecular typing methods based on variable-number tandem repeats (VNTRs) of genetic elements called mycobacterial interspersed repetitive units (MIRUs) have also been developed by Supply et al. (2000, 2001). MIRU-VNTR typing shows a discriminatory power close to that of IS*6110* fingerprinting and can be used to differentiate between *M. tuberculosis* isolates that lack or have low copy numbers of IS*6110* elements (Supply et al. 2000). Very few studies in India have reported use of MIRU-VNTRs for molecular typing of *M. tuberculosis* isolates (Sharma et al. 2008, Stavrum et al. 2009, Narayanan et al. 2010).

The aim of this investigation was to evaluate the usefulness of the aforementioned typing methods in genotyping *M. tuberculosis* isolates in North Delhi and to obtain an initial assessment of the circulation of any dominant genotypes in our region. For this reason, we used a panel of epidemiologically unrelated *M. tuberculosis* isolates from the North Delhi region.

SUBJECTS, MATERIALS AND METHODS

Clinical isolates - The study included a total of 101 positive *M. tuberculosis* cultures obtained from as many new smear-positive pulmonary TB patients over a two-year period from January 2005-December 2006 in the North Delhi region. Patients suffering from TB were recruited from the outpatient facility of the Department of Respiratory Medicine at Vallabhbhai Patel Chest Institute, which serves as a referral centre for patients with respiratory diseases in North India, and the Directly Observed Treatment, Short Course Centre at Rajan Babu TB Hospital in Delhi. Only adult patients ≥ 18 years of age were enrolled for the study (81% of the patients were between 20-40 years old and the remaining were over 40 years old; the sex ratio was 3.37). All of the patients lived within a 10 km radius. Culture and identification were performed at the Vallabhbhai Patel Chest Institute; all isolates were confirmed to be *M. tuberculosis* by niacin, nitrate and catalase tests. Following clearance from the Institutional Ethical Committee, informed consent and a detailed history of contact were taken from each patient prior to sample collection.

Genotyping and database comparison - The bacterial DNAs were prepared as reported previously (van Soolingen et al. 1991) and were submitted to spoligotyping (Kamerbeek et al. 1997), 12-loci MIRU typing (Supply et al. 2000) and IS*6110*-RFLP (van Embden et al. 1993), using published protocols. MIRU typing was performed manually using polymerase chain reaction amplification of 11 loci because the 12th locus, locus 20, was found to be poorly discriminative in a preliminary study in the North Delhi region (results not shown) and was therefore not included in the present study. For database comparison, spoligotypes in binary format and MIRU patterns were entered in the SITVIT2 proprietary database of the Pasteur Institute of Guadeloupe, which is an updated version of the previously released SpolDB4 database (Brudey et al. 2006). At the time of the comparison, it contained genotyping data on 65,000 strains of *M. tuberculosis* complex from 160 countries of origin. In this database, "SIT" (Spoligotype International Type) designates spoligotyping shared by two or more patient isolates, "MIT" (MIRU International Type) designates 12-loci MIRU patterns shared by two or more patient isolates and "orphan" designates patterns reported for a single isolate. Major spoligotyping-based phylogenetic clades were assigned according to signatures provided in SpolDB4 (Brudey et al. 2006). These clades include specific signatures for *M. tuberculosis* complex members and rules defining major lineages/sub-lineages for *M. tuberculosis sensu stricto*, i.e., the Central-Asian (CAS) clade, the East-African-Indian (EAI) clade, the Haarlem

(H) clade, the Latin-American-Mediterranean (LAM) clade, the "Manu" family, the IS*6110*-low-banding X clade and an ill-defined T clade.

For IS*6110*-RFLP, the membranes were scanned and the banding patterns were analysed using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) and Jaccard coefficient with GelCompar II, version 5.1 software (Applied Maths NV, Sint-Martens-Latem, Belgium). A final combined analysis of the spoligotyping, MIRU, and IS*6110*-RFLP data was performed using BioNumerics software, version 3.5 (Applied Maths), to establish the genetic relationships among strains using UPGMA following the Dice similarity index.

Phylogenetic analysis - The evolutionary relationships among all of the observed spoligotypes were studied by drawing a minimum spanning tree (MST) with BioNumerics software as reported previously (Helal et al. 2009). MST is an undirected network in which all of the samples are linked together with the fewest possible linkages between nearest neighbours. Using this approach, one considers that all intermediate stages are present within the sample analysed by including first the individual that shows the greatest number of possible linkages to other individuals in the population studied. We used this method to highlight the phylogenetic links between two spoligotypes differing by changes observed in direct variable repeats (Helal et al. 2009). Consequently, the lengths of the branches in the tree indicate the level of change induced by the loss or gain of spoligotype spacers in the 43-oligonucleotide format that induce a shift from one allele to another; the solid lines represent a single spacer change, dotted lines represent two spacer changes (black) or more than two spacer changes (gray). The colours of the circles in the tree indicate the number of clinical isolates represented by each circle, i.e., unique isolates (sky blue) vs. clustered isolates (deep blue: 2-5 strains; red: 6 strains or more). Principal genetic groups in the present study were described based on spoligo-defined lineages (Brudey et al. 2006), which have previously been correlated with the katG-gyrA-based classification of Sreevatsan et al. (1997).

Statistical analysis - Allelic diversity of each of the 11 VNTR loci studied was calculated by the equation $h = 1 - \sum x_i^2 [n/(n-1)]$, where x_i is the frequency of the i-th allele at the locus, n is the number of isolates and $n/(n-1)$ is a correction for bias in small samples (Selander et al. 1986).

RESULTS

Spoligotyping results are summarised in Table I. Using this method, a total of 49 different patterns were observed for the 101 strains studied; seven patterns corresponded to orphan patterns not yet reported to the SITVIT2 database, whereas the remaining 42 patterns (n = 94 strains) corresponded to SITs. Among these 42 patterns, 35 SITs (n = 83 strains) matched a pre-existing shared type in the SITVIT2 database, whereas seven SITs (n = 11 strains) were newly created. Of these seven SITs, two (SIT2757, SIT2761) were created within the present study and five were created after a match with an orphan in the database (SIT2621 and SIT2756 from

TABLE I

Description of the *Mycobacterium tuberculosis* spoligotypes

SIT[a]	Spoligotype description	Number in this study	Percentual in this study	Percentual in database	Clade[c]
1	▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫■■■■■■■	5	5	10.70	Beijing
11	■■▫▫■■■■■■■■■■■■■■■■■■▫▫▫▫■▫■▫▫▫■■■	11	10.9	0.81	EAI3-IND
25	■■■▫▫▫▫■■■■■■■■■■■■▫▫▫▫▫▫▫▫■■▫▫■■■■	1	1.0	0.47	CAS1-Delhi_var
26	■■■▫▫▫▫■■■■■■■■■■■■▫▫▫▫▫▫▫▫▫■■■■■■■	21	20.8	1.68	CAS1-Delhi
42	■■■■■■■■■■■■■■▫▫▫▫■■■▫▫▫▫■■■■	1	1	3.68	LAM
48	■■■■■■■■■■■■■■■■■■▫▫▫▫■▫■■■■■▫■	3	3	0.58	EAI
50	■■■■■■■■■■■■■■■■■■■■■■▫▫▫▫■■■■■	1	1	4.14	H
53	■■■■■■■■■■■■■■■■■■■■■■■■■▫▫▫▫■■■■	2	2	7.08	T
60	■■■■■■■■■■■▫▫▫▫■■■■■■■■▫▫▫▫■■■■	1	1	0.34	LAM
64	■■■■■■■■■■■■▫▫▫▫■■▫■■■■▫▫▫▫■■■■	1	1	0.45	LAM
78	■■■■■■■■■■■■■■■■■■■■■■■▫▫▫▫■■▫■	1	1	0.09	T
126	■▫▫■■■■■■■■■■■■■■■■▫▫▫▫■■■▫■■■■	1	1	0.14	EAI3-IND_var
236	■■■■■■■■■■■■■■■■■■■■▫▫▫▫■■■■■■■	2	2	0.25	EAI
244	■■■■■■■■■■■■■■■■■■■■▫▫▫▫■▫▫▫▫■	1	1	0.14	T
278	■■■■■■■■■■■■■■■■■■■■■■■▫▫▫▫■■■■▫■	2	2	0.02	T
283	■■■■■■■■■■■■■■▫▫▫■▫▫▫▫▫■▫▫▫▫■■■■■	1	1	0.08	H
288	■■■▫▫▫▫▫▫▫▫■■■■■■▫▫▫▫▫▫▫▫▫■■■■▫■	2	2	0.15	CAS2
292	■■■■■■■■■■■■▫■▫▫■■▫■■■▫■■▫■▫■■	1	1	0.20	EAI
334	■▫■■■■■■■■■■■■■■■■■■■▫▫▫▫■■■■■	1	1	0.11	T
357	■■■▫▫▫▫■■■■■■■■▫▫▫▫▫▫▫▫▫▫▫■■■■■	3	3	0.11	CAS1-Delhi_var
381	■■■▫▫▫▫■■■■■■■■▫▫▫▫▫▫▫▫▫■■■■▫▫■■■	1	1	0.04	CAS1-Delhi_var
402	■■■■■■■■■■■■■▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫	1	1	0.05	Unknown[d]
591	■■■■■■■■■■■■■■▫■■▫▫▫▫■■■■■■■	1	1	0.08	EAI
620	■■■■■■■■■■■■▫▫■▫▫▫▫▫■▫▫▫▫■■■■	1	1	0.02	H
1091	■■■▫▫▫▫■■■■■■■▫▫▫▫▫▫▫▫▫■■▫■■■■	1	1	0.02	CAS1-Delhi_var
1264	■■■▫▫▫▫■■■■■■▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫	3	3	0.03	CAS
1311	▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫▫■■■▫▫▫▫	1	1	0.01	Beijing
1316	■■■▫▫■■■■■■■■■■■■■▫▫▫▫■■■■▫■	1	1	0.01	EAI
1342	■▫▫■■■■■■■■▫■■■▫▫▫▫■■■▫▫■■■	2	2	0.01	EAI3-IND_var
1344	■■■▫▫▫▫■■■■■■■▫▫■▫▫▫▫▫▫▫▫▫■■■■	1	1	0.01	CAS1-Delhi_var
1378	■■■■■■■■■■■■■■■■■■■▫▫▫■■■■■	4	4	0.05	MANU
1401	■■■▫▫▫▫■■■■■■■▫▫▫▫▫▫▫▫▫▫■■■■■	1	1	0.02	CAS1-Delhi_var
1947	■■■▫▫▫▫■■■■■■▫■▫▫■▫▫▫▫▫▫■■■■■	1	1	0.01	CAS1-Delhi_var
1951	■■■▫▫▫▫■■■■■■■▫■▫▫▫▫▫▫▫▫▫■■■■■	1	1	0.01	CAS1-Delhi_var
1970	■■■■■■■■■■▫■■■▫▫■■▫▫▫■■■■■	1	1	0.01	EAI
2621[b]	■■■▫▫▫▫■■■■■■▫▫▫▫▫▫▫▫▫▫▫■■■■■	1	1	New	CAS1-Delhi_var
2756[b]	■■■▫▫▫▫■■■■■■■▫▫▫▫▫▫▫▫▫▫▫▫▫■■■	1	1	New	CAS
2757[b]	■■■▫▫▫▫▫■▫▫▫■▫■■■■■■■▫▫▫▫■■■	3	3	New	Unknown[d]
2758[b]	■■■▫▫▫▫▫▫▫▫▫■■▫▫▫▫▫▫▫▫▫▫▫■■■	1	1	New	CAS2
2759[b]	■■■▫▫▫▫■■■■■▫▫▫▫▫▫▫▫▫▫▫▫▫▫■■■	1	1	New	CAS
2760[b]	■■■■■■■■■■■■■▫▫▫▫■■■■■▫▫▫■	1	1	New	Unknown[d]
2761[b]	▫■■■■■■■■■■■■■■■■■▫▫▫▫■▫■■■■■▫▫▫	3	3	New	EAI3-IND_var
Orphan 00008	■▫▫■■■■■■■■■■■■■■■■■■▫▫▫■■▫■■■■	1	1	Orphan	EAI
Orphan 00438	■■■▫▫▫▫▫▫▫▫■■■■■▫▫▫▫▫▫▫▫▫▫▫▫▫	1	1	Orphan	CAS-like
Orphan 00409	■■■▫▫▫▫■▫■▫▫▫▫▫■■■▫▫▫▫▫▫▫▫▫▫▫■■	1	1	Orphan	CAS1-Delhi_var
Orphan 00453	■■■■■■■■■■■■■■■■■■■■▫▫▫▫■▫▫■	1	1	Orphan	Unknown[d]
Orphan 00643	■■■■■■■■■■■■■■■■■■■■■■■▫▫▫■■▫■■	1	1	Orphan	Manu
Orphan 03398	■■■▫▫▫▫■■■■■■■▫▫■■■■■■▫▫▫▫■■■■	1	1	Orphan	CAS
Orphan 05122	■■■▫▫▫▫▫▫▫■■■■■■▫▫▫■▫▫▫▫▫▫▫■■■■■■	1	1	Orphan	CAS1-Delhi_var

a: forty nine spoligotype patterns from 101 isolates: seven patterns corresponded to orphan patterns not yet reported to the SITVIT database (shown by their individual strain number). The remaining 42 patterns (n = 94 strains) corresponded to Spoligotype International Types (SITs): 35 SITs containing 83 isolates matched a pre-existing shared type in the SITVIT2 database, whereas seven SITs (*b*) (containing 11 isolates) were newly-created either within the present study (SIT2757, SIT2761) or after a match with an orphan in the database (SIT2621 and SIT2756 from Pakistan, SIT2758 from USA, SIT2759 from Italy and SIT2760 from India); *c:* clade designations according to SITVIT database using revised SpolDB4 rules; *d*: unknown patterns within any of the major clades described; CAS: Central-Asian clade; EAI: East-African-Indian clade; H: Haarlem clade; IND: Indian clade; LAM: Latin-American-Mediterranean clade; T: ill-defined T clade.

Pakistan, SIT2758 from the USA, SIT2759 from Italy and SIT2760 from India). A total of 66 strains (65% of the studied strains) were clustered into 14 clusters (2-21 isolates per cluster), whereas the remaining 35 strains (34.5% of the studied strains) were unique. The largest cluster, containing 21 isolates, was composed of SIT26 (CAS1-Delhi lineage), followed by SIT11 [EAI3-Indian (IND) lineage, n = 11 strains]. Of the 35 unique strains, seven were orphan strains and 28 were pseudo-orphan strains. Five of the pseudo-orphan strains were present as single isolates in the five newly-created SITs mentioned previously.

Interestingly, as many as 75.2% (76) of the 101 isolates in our study belonged to *M. tuberculosis* complex lineages, such as CAS, EAI and Manu, that are known to have a high specificity for the Indian subcontinent (Bhanu et al. 2002, Mistry et al. 2002, Kulkarni et al. 2005, Brudey et al. 2006, Mathuria et al. 2008, Sharma et al. 2008). The remaining 24.8% (25/101) of isolates belonged to the following clades: T (7/101), Beijing (6/101), LAM (3/101), H (3/101) and unknown (6/101). These observations emphasise the complex diversity of the phylogenetically coherent *M. tuberculosis* in our study sample from North Delhi, an observation that was further corroborated by the MST illustrated in A in Figure. Indeed, as many as 82/101 (81.2%) strains from our study belonged to Principal Genetic Group (PGG) 1, according to the *katG-gyrA* polymorphism-based classification of Sreevatsan et al. (1997), and these strains represented all of the known lineages associated with PGG1 isolates, i.e., EAI 26.7%, CAS 43.6%, Beijing 5.94% and Manu 4.95%. However, evolutionarily modern PGG2/3 strains were limited to LAM, H and T; these strains accounted for only 12.9% (13) of the 101 strains studied.

The MIRU-VNTR genotyping data for all of the 101 isolates in relation to the aforementioned spoligotyping results are shown in Table II. Interestingly, despite the 11-loci format used, 99 of 101 isolates were split as individual patterns in our study. The only exception concerned a cluster that exhibited an identical profile (pattern 3373*7143423) for isolates 5622 and 5942. Our study samples were not epidemiologically related, as evidenced by the contact details provided by the patients. Hence, such a high diversity of strains was not surprising. Thus, even the least diverse patterns, which concerned four isolates numbered 195, 495, 5109 and 5442, still differed from each other by two loci (MIRU loci 16 and 23). In fact, even strains 5622 and 5942, which were clustered by MIRU-VNTR genotyping, belonged to two different spoligotypes. Regarding the allelic diversities of the MIRU loci studied, locus 31 was the most discriminative, with nine alleles in our study ($h \geq 0.76$), followed by loci 26, 40, 10, 16, 31, 23, 39 and 24 ($h \geq 0.6$), whereas loci 2, 4 and 27 were moderately discriminative ($h \geq 0.3$ to < 0.6). Interestingly, although ancestral strains (such as EAI) are characterised by two or more copies of locus 24 and modern strains (such as CAS) are often characterised by one copy (and rarely 0), this pattern was not observed in our study (Table II). A larger survey may be needed in this part of the world to confirm this finding.

The IS*6110* fingerprints for representative clinical isolates of *M. tuberculosis* (n = 80, including all of the 66 isolates clustered by spoligotyping), are summarised in an UPGMA tree constructed by combined numerical analysis of IS*6110*-RFLP, spoligotyping and MIRU data in B in Figure, and the details of spoligotyping-defined clusters and RFLP patterns observed for each of the *M. tuberculosis* complex genotypic lineages in the study sample are summarised in Table III. Our data revealed that 18/80 (22.5%) of the strains had low copy numbers of IS*6110* bands ($\leq$ 6 copies). Among the low-copy-number strains, seven isolates (8.75%) had only a single IS*6110* band, while 11 isolates (13.75%) had two-six copies of IS*6110*. The remaining 62 strains (77.5%) contained > 6 copies of IS*6110*. The lineage with the maximum number of low-copy-number strains (8/21 or 38.1%) was EAI; half of these strains contained a single copy of IS*6110* (Table III). The largest clade in our study, CAS, had 7/38 (18.4%) low-copy-number isolates. No statistically-significant correlation was found between the lineage and the incidence of low-copy-number isolates.

A summary of IS*6110*-RFLP-defined clusters (4 clusters containing 13 strains; analysis performed using BioNumerics software at 5% tolerance with the same number of IS*6110* copies and band positions) and the corresponding spoligotyping and MIRU-VNTR data are summarised in Table IV. These data comprised a one-banded pattern (n = 6 strains with a band of 1.45 kbp), a four-banded pattern (n = 3 strains, bands of 1.4, 2.2, 3.2 and 5.2 kbp), a seven-banded pattern (n = 2 strains, bands of 1.1, 1.5, 1.9, 2.1, 3.0, 4.1 and 5.5 kbp) and a 12-banded pattern (n = 2 strains, bands of 1.0, 1.3, 1.45, 2.2, 2.45, 3.1, 3.4, 4.9, 6.2, 6.8, 7.8 and 8.7 kbp). Note that all of the 13 strains were clearly discriminated by MIRUs and all but two (with the exception of strains 5 and 938, which belonged to the EAI3-IND sublineage) were also discriminated by spoligotyping.

DISCUSSION

Although various studies have been performed in India using spoligotyping of *M. tuberculosis* (Mistry et al. 2002, Singh et al. 2004, 2007, Kulkarni et al. 2005, Mathuria et al. 2008, Sharma et al. 2008), very few studies have reported the use of MIRU-VNTRs for molecular typing of *M. tuberculosis* isolates (Gutierrez et al. 2006, Sharma et al. 2008, Stavrum et al. 2009, Narayanan et al. 2010). Therefore, the present study validates the use of repetitive elements for molecular typing of *M. tuberculosis* in Indian isolates and further underlines the usefulness of MIRU-VNTRs for discriminating low-IS*6110*-copy isolates, which accounted for more than one-fifth of the strains in the present study. IS*6110*-RFLP provided insufficient discrimination because of its inherently limited polymorphism among isolates with low IS*6110* copy numbers (< 6 copies). This deficiency is partially solved with spoligotyping as a secondary method (Cowan et al. 2005). Our results also corroborate a previous proposition that it is ultimately MIRU typing that will resolve the problems stemming from low-IS*6110*-copy strains (Supply et al. 2000); indeed, IS*6110*-RFLP typing in other parts of India has revealed the presence of a
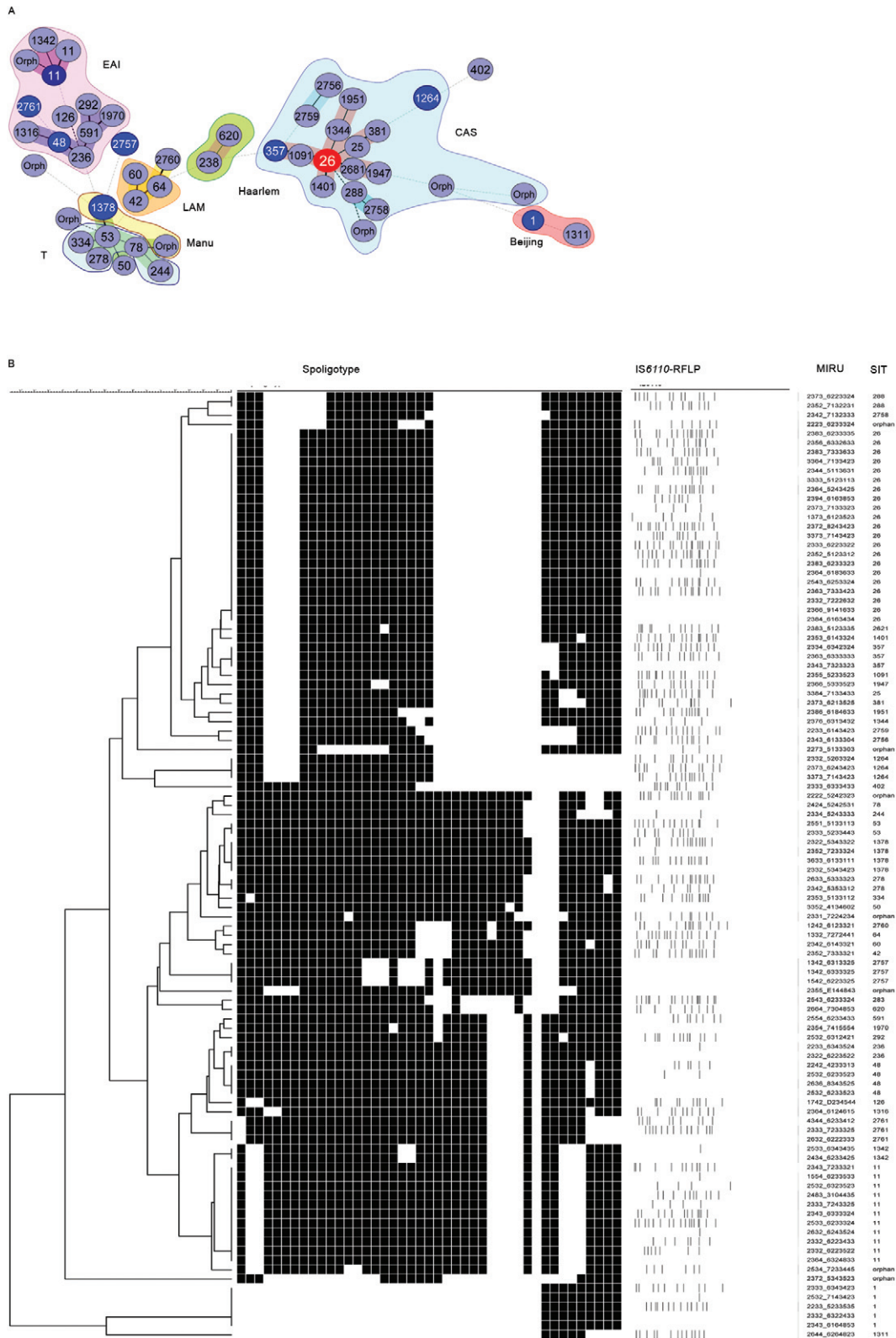
A summary of the population structure of *Mycobacterium tuberculosis* clinical isolates from North India. A: unrooted minimum spanning tree (MST) showing evolutionary relationships of *M. tuberculosis* spoligotypes. The length of the branches indicates the level of change required by loss of spacer to go from one allele to another. The colour of circles is proportional to number of clinical isolates in the study (blue: 1 strain; dark blue: 2-5 strains; red: 6 or more strains); B: combined numerical analysis of IS*6110*-restriction fragment length polymorphism (RFLP), spoligotyping and mycobacterial interspersed repetitive units (MIRUs) data using the unweighted-pair group method using average linkages; CAS: Central-Asian clade: EAI: East-African-Indian clade; H: Haarlem clade; IND: Indian clade; LAM: Latin-American-Mediterranean clade; SIT: Spoligotype Internacional Types; T: ill-defined T clade.

TABLE II

Mycobacterial interspersed repetitive units (MIRUs)-variable-number tandem repeats (VNTRs)
typing in relation to the spoligotypes of the samples

Serial number	Sample number	Spoligotype clade	12-loci MIRU patterns[a]											
1	00001	CAS1-Delhi	2	3	6	4	*5	2	4	3	4	2	5
2	00007	CAS1-Delhi	2	3	8	3	*6	2	3	3	3	2	3
3	00009	CAS1-Delhi	2	3	6	3	*7	3	3	3	4	2	3
4	00D13	CAS1-Delhi	2	3	8	4	*6	1	6	3	4	3	4
5	00019	CAS1-Delhi	2	3	8	3	*6	2	3	3	3	3	5
6	00023	CAS1-Delhi	2	3	8	3	*7	3	3	3	6	3	3
7	00115	CAS1-Delhi	2	3	9	4	*6	1	6	3	8	5	3
8	00232	CAS1-Delhi	2	3	3	2	*7	2	2	2	6	3	2
9	00292	CAS1-Delhi	3	3	3	3	*5	1	2	3	1	1	3
10	00413	CAS1-Delhi	2	3	7	3	*7	1	3	3	3	2	3
11	00456	CAS1-Delhi	1	3	7	3	*6	1	2	3	5	2	3
12	00533	CAS1-Delhi	2	3	5	6	*6	3	3	2	6	3	3
13	03328	CAS1-Delhi	2	3	6	4	*6	1	8	3	6	3	3
14	03439	CAS1-Delhi	2	3	4	4	*5	1	1	3	6	3	1
15	03472	CAS1-Delhi	2	3	6	6	*9	1	4	1	6	3	3
16	04882	CAS1-Delhi	2	3	3	3	*6	2	2	3	3	2	2
17	05109	CAS1-Delhi	2	3	7	2	*8	2	4	3	4	2	3
18	05119	CAS1-Delhi	3	3	6	4	*7	1	3	3	4	2	3
19	05188	CAS1-Delhi	2	5	4	3	*6	2	5	3	3	2	4
20	05605	CAS1-Delhi	2	3	5	2	*5	1	2	3	3	1	2
21	05622	CAS1-Delhi	3	3	7	3	*7	1	4	3	4	2	3
22	00002	CAS1-Delhi_var	2	3	5	5	*5	2	3	3	5	2	3
23	00003	CAS1-Delhi_var	2	3	6	6	*5	3	3	3	5	2	3
24	00006	CAS1-Delhi_var	2	3	6	3	*6	3	3	3	3	3	3
25	00015	CAS1-Delhi_var	2	3	8	3	*5	1	2	3	3	3	5
26	00018	CAS1-Delhi_var	3	3	8	4	*7	1	3	3	4	3	3
27	00061	CAS1-Delhi_var	2	3	3	4	*6	3	4	2	3	2	4
28	00221	CAS1-Delhi_var	2	3	5	3	*6	1	4	3	3	2	4
29	00333	CAS1-Delhi_var	2	3	7	6	*6	3	1	3	4	3	2
30	00409	CAS1-Delhi_var	2	2	7	3	*5	1	3	3	3	0	3
31	00586	CAS1-Delhi_var	2	3	4	3	*7	3	2	3	3	2	3
32	00587	CAS1-Delhi_var	2	3	7	3	*6	2	1	3	5	2	5
33	03365	CAS1-Delhi_var	2	3	8	6	*6	1	8	4	6	3	3
34	05122	CAS1-Delhi_var	2	2	2	3	*6	2	3	3	3	2	4
35	00234	CAS2	2	3	5	2	*7	1	3	2	2	3	1
36	04735	CAS2	2	3	4	2	*7	1	3	2	3	3	3
37	05389	CAS2	2	3	7	3	*6	2	2	3	3	2	4
38	00030	CAS	2	3	3	2	*5	2	6	3	3	2	4
39	00377	CAS	2	3	4	3	*6	1	3	3	3	0	4
40	00438	CAS	2	3	7	2	*5	3	4	3	5	2	3
41	03398	CAS	2	3	5	5	*F	1	4	4	8	4	3
42	05137	CAS	2	2	3	3	*6	1	4	3	4	2	3
43	05442	CAS	2	3	7	3	*6	2	4	3	4	2	3
44	05942	CAS	3	3	7	3	*7	1	4	3	4	2	3
45	00022	Beijing	2	2	3	3	*5	2	3	3	5	3	5
46	00179	Beijing	2	6	4	4	*6	2	6	4	8	2	3
47	00195	Beijing	2	3	3	3	*6	3	4	3	4	2	3
48	00205	Beijing	2	3	3	2	*6	3	2	2	4	3	3
49	00284	Beijing	2	5	3	2	*7	1	4	3	4	2	3
50	03288	Beijing	2	3	4	3	*6	1	6	4	8	5	3
51	00404	Unknown	1	3	4	2	*6	3	1	3	3	2	5

↳

Serial number	Sample number	Spoligotype clade	12-loci MIRU patterns[a]											
52	00418	Unknown	1	3	4	2	*6	3	3	3	3	2	5
53	00420	Unknown	1	5	4	2	*6	2	2	3	3	2	5
54	05399	LAM	2	3	4	2	*6	1	4	3	3	2	1
55	05443	LAM	2	3	5	2	*7	3	3	3	3	2	1
56	05706	LAM	1	3	3	2	*7	2	7	2	4	4	1
57	00011	Manu	2	3	5	2	*7	2	3	3	3	2	4
58	00474	Manu	2	3	2	2	*5	3	4	3	3	2	2
59	00495	Manu	2	3	3	2	*5	3	4	3	4	2	3
60	00643	Manu	2	2	2	2	*5	2	4	2	3	2	3
61	06262	Manu	3	6	3	3	*6	1	3	3	1	1	1
62	00012	T	2	3	3	3	*5	2	3	3	4	4	3
63	00025	T	2	3	4	2	*5	3	5	3	3	1	2
64	00240	T	2	6	3	3	*5	3	3	3	3	2	3
65	01109	T	2	3	3	4	*5	2	4	3	3	3	3
66	02186	T	2	3	5	3	*5	1	3	3	1	1	2
67	03915	T	2	5	5	1	*5	1	3	3	1	1	3
68	03928	T	2	4	2	4	*5	2	4	2	5	3	1
69	00004	EAI	2	2	3	3	*6	3	4	3	5	2	4
70	00008	EAI	2	5	3	4	*7	2	3	3	4	4	5
71	00013	EAI	2	5	3	2	*6	2	3	3	5	2	3
72	00017	EAI	2	5	5	4	*6	2	3	3	4	3	3
73	00028	EAI	2	6	3	6	*8	3	4	3	5	2	5
74	00077	EAI	2	2	4	2	*4	2	3	3	3	1	3
75	00298	EAI	2	3	2	2	*6	2	2	3	5	2	2
76	00406	EAI	2	5	3	2	*6	3	1	2	4	2	1
77	03322	EAI	2	3	5	4	*7	4	1	5	5	5	4
78	03493	EAI	2	3	6	4	*6	1	2	4	6	1	5
79	00005	EAI3-IND	2	6	3	2	*6	2	4	3	5	2	4
80	00050	EAI3-IND	2	4	8	3	*3	1	0	4	4	3	5
81	00130	EAI3-IND	2	3	4	3	*7	2	3	3	3	2	1
82	00134	EAI3-IND	2	3	4	3	*6	3	3	3	3	2	4
83	00374	EAI3-IND	2	5	3	3	*6	2	3	3	3	2	4
84	00378	EAI3-IND	2	5	3	2	*6	3	2	3	5	2	3
85	00652	EAI3-IND	2	3	3	2	*6	2	2	3	5	2	2
86	00660	EAI3-IND	2	3	3	2	*6	2	2	3	4	3	3
87	00938	EAI3-IND	1	5	5	4	*6	2	3	3	5	3	3
88	03390	EAI3-IND	2	3	6	4	*6	3	2	4	8	3	3
89	05431	EAI3-IND	2	3	3	3	*7	2	4	3	3	2	5
90	00058	EAI3-IND_var	1	7	4	2	*E	2	3	4	5	4	4
91	01255	EAI3-IND_var	2	4	3	4	*6	2	3	3	4	2	5
92	01263	EAI3-IND_var	2	5	3	3	*6	3	4	3	4	3	5
93	05396	EAI3-IND_var	2	3	3	3	*7	2	3	3	3	2	5
94	05438	EAI3-IND_var	2	6	3	2	*6	2	2	2	3	3	3
95	05468	EAI3-IND_var	4	3	4	4	*6	2	3	3	4	1	2
96	00169	H	3	3	5	2	*4	1	3	4	6	0	2
97	03510	H	2	6	6	4	*7	3	0	4	8	5	3
98	05256	H	2	5	4	3	*6	2	3	3	3	2	4
99	00020	Unknown	2	3	3	3	*6	3	3	3	4	3	3
100	00453	Unknown	2	3	3	1	*7	2	2	4	2	3	4
101	05176	Unknown	1	2	4	2	*6	1	2	3	3	2	1

MIRU patterns are shown in the order (2, 4, 10, 16, 20, 23, 24, 26, 27, 31, 39 and 40). Note that MIRU loci 20 was not amplified in our study (shown as asterisk) since it did not show promising discriminatory power. Note that number of copies > 9 is shown by a letter (10 = A, 11 = B, 12 = C etc.). Strains 5622 (serial 21) and 5942 (serial 44) have the same MIRU pattern, but different spoligotypes. CAS: Central-Asian clade: EAI: East-African-Indian clade; H: Haarlem clade; IND: Indian clade; LAM: Latin-American-Mediterranean clade; T: ill-defined T clade.

TABLE III

Details of spoligotype clusters and IS*6110* restriction fragment length polymorphism (RFLP) patterns observed
for each of the *Mycobacterium tuberculosis* complex genotypic lineage in the study sample

Clade	Total (n)	Isolates typed by IS*6110*-RFLP (n)	Isolates with a single copy of IS*6110* n (%)	Isolates with two-six copies of IS*6110* n (%)	Isolates with > six copies of IS*6110* n (%)
EAI	27	21	4 (19.05)	4 (19.05)	13 (61.90)
CAS	44	38	1 (2.60)	6 (15.80)	31 (81.60)
Beijing	6	4	1 (25)	0	3 (75)
Manu	5	4	1 (25)	0	3 (75)
T	7	6	0	1 (16.67)	5 (83.33)
LAM	3	3	0	0	3 (100)
H	3	2	0	0	2 (100)
Undefined	6	2	0	0	2 (100)

CAS: Central-Asian clade: EAI: East-African-Indian clade; H: Haarlem clade; IND: Indian clade; LAM: Latin-American-Mediterranean clade; T: ill-defined T clade.

significant number of *M. tuberculosis* isolates with low copy numbers (15, 67%) or zero copies (1, 24%) of IS*6110* bands (Das et al. 1995, Mathuria et al. 2008, Narayanan et al. 2008, Radhakrishnan et al. 2001). We did not find any zero-copy strains in the present study.

Spoligotyping, in the present study, revealed 49 SIT patterns. Of these patterns, seven SITs were newly created. This observation might suggest the introduction of new genotypes due to casual contact and/or increased international travel. The most common spoligotype found in our study was SIT26 (CAS1-Delhi, n = 21, 20.8% of all isolates), followed by SIT11 (EAI3-IND lineage, n = 11 strains, 10.9% of all isolates). Reportedly a predominant spoligotype in Northern India (Bhanu et al. 2002, Singh et al. 2004, 2007, Suresh et al. 2006, Mathuria et al. 2008, Stavrum et al. 2009), SIT26 was found at a lower frequency (7.4% of all strains) in a study from Mumbai (Kulkarni et al. 2005) and represents only 1.7% of strains in the SITVIT2 database. SIT26 is limited mainly to the Middle East and to Central Asia. It has also been found in regions in which frequent migration to and from the Indian subcontinent occurs, e.g., Kenya, South Africa, Malaysia, Myanmar, Australia, USA and parts of Europe. The second-largest spoligotype found in our study (SIT11, EAI3-IND) has also been found to be a major spoligotype in other studies from Delhi (Singh et al. 2004). The Beijing spoligotype (SIT1) was found in 5% of cases in the present study. SIT1 has also been found to be one of the major spoligotypes in other parts of North India, varying from 3.8-9.6% of the isolates studied (Singh et al. 2004, Suresh et al. 2006, Mathuria et al. 2008). A study from Mumbai reported SIT1 to be the major spoligotype, along with SIT26 (7.4% of the isolates for each lineage) (Kulkarni et al. 2005). In the international database, SIT1 contains 10.7% of the strains reported. Interestingly, one of the Beijing isolates found in our study had a single IS*6110* band, even though the Beijing genotype is also defined on the basis of a characteristic multibanded IS*6110* RFLP pattern. However,

outliers in the Beijing family have also been identified by other investigators (Kurepina et al. 1998). Lastly, we also found that 4.95% of the strains in this study belonged to the Manu lineage; however, unlike other reports (Singh et al. 2004, Suresh et al. 2006), we did not find strains belonging to SIT54. A study from New Delhi reported SIT54 strains in 12% of the isolates studied (Singh et al. 2004). In a subsequent report from the same group, SIT54, representing a recently-defined Manu clade, was found in only 0.55% of the isolates studied (Singh et al. 2007). We did not find a single isolate belonging to SIT54 in our study. Other recent studies from North India have also failed to find isolates belonging to SIT54 (Mathuria et al. 2008, Sharma et al. 2008). Thus, even though the majority of the isolates in our study (76/101) belonged to *M. tuberculosis* complex lineages, such as CAS, EAI and Manu that are frequently reported in the Indian subcontinent (Bhanu et al. 2002, Singh et al. 2004, 2007, Kulkarni et al. 2005, Brudey et al. 2006, Gutierrez et al. 2006, Suresh et al. 2006, Mathuria et al. 2008, Sharma et al. 2008, Stavrum et al. 2009), changes occurring within clades reemphasise the need for continuous cluster investigation to detect any change in the circulating spoligotypes.

The phylogenetically-coherent *M. tuberculosis* complex diversity in our study was corroborated by the MST shown in A in Figure. The PGG1 strains in our study (81.2% of all isolates) represented all known lineages associated with the PGG classification scheme, i.e., EAI 26.7%, CAS 43.6%, Beijing 5.94% and Manu 4.95%. Based on the presence or absence of a specific deletion (TbD1, a 2 kb sequence), *M. tuberculosis* can be divided into "ancient" TbD1-positive and "modern" TbD1-negative strains, a classification scheme that corresponds well with the PGG classification in that both EAI and *Mycobacterium africanum* are TbD1 positive (Rastogi & Sola 2007). Interestingly, the nodes for each of the PGG1 groups in the tree shown in A in Figure were composed of the prototypes that characterise the

TABLE IV

Summary of IS*6110*-restriction fragment length polymorphism (RFLP)-defined clusters using a 5% tolerance rate with the same number of IS*6110* copies and band positions[a] and parallel spoligotyping and mycobacterial interspersed repetitive units (MIRUs)-variable-number tandem repeats (VNTRs) data

IS*6110*-RFLP clusters	Sample number	IS*6110*-RFLP patterns	SIT (clade)	Spoligotype description	MIRU
1 band	4		236 (EAI5)		2233*6343524
	5		11 (EAI3-IND)		2632*6243524
	938		11 (EAI3-IND)		1554*6233533
	284		1 (Beijing)		2532*7143423
	1263		1342 (EAI3-IND)		2533*6343435
	3328		26 (CAS1-Delhi)		2364*6183633
4 bands	5431		11 (EAI3-IND)		2333*7243325
	413		26 (CAS1-Delhi)		2373*7133323
	3333		1344 (CAS1-Delhi_var)		2376*6313432
7 bands	25		278 (T1)		2342*5353312
	660		11 (EAI3-IND)		2332*6223433
12 bands	5399		60 (LAM4)		2342*6143321
	5389		288 (CAS2)		2373*6223324

a: four clusters containing 13 strains. These concerned identical one banded pattern (n = 6), four banded pattern (n = 3), seven banded pattern (n = 2) and 12 banded pattern (n = 2). Note that all the strains were clearly discriminated at individual level by MIRUs and all but two [with the exception of strains 5 and 938 belonging to East-African-Indian (EAI)3-Indian (IND) sublineage] were also discriminated by spoligotyping; CAS: Central-Asian clade; H: Haarlem clade; LAM: Latin-American-Mediterranean clade; SIT: Spoligotype International Types; T: ill-defined T clade.

respective lineages in the SpolDB4 database. This finding, combined with the frequency of isolation of each strain (denoted by the colour of the circle in the MST), showed that CAS (prototype SIT26 for CAS1-DEL) is the predominant lineage in North India, followed by EAI (prototype SIT11 for EAI3-IND) and Manu (prototype SIT1378 for Manu-3), and that CAS could have been present historically in this subcontinent. The Manu lineage, considered the ancestor of *M. tuberculosis* complex isolates (Brudey et al. 2006), was originally described in India (Singh et al. 2004). In our study, both EAI and CAS lineage strains occurred more frequently (75%) in patients between the ages of 18-44 years of age, reflecting ongoing transmission in North India. Our observations of the prevalence and proportion of CAS1-DEL and EAI3-IND corroborate recent findings from India, that the bulk of TB infections in North India is caused by the TbD1-negative/PGG1 genogroup (Gutierrez et al. 2006, Stavrum et al. 2009) as opposed to infections in South India, where the TbD1-positive/PGG1 EAI lineage predominates (Narayanan et al. 2008). Lastly, the high diversity of *M. tuberculosis* strains in our study, despite the fact that they were isolated from patients attending health facilities within a 10 km radius in a country with a high level of TB endemicity, was remarkable.

Although the 101 isolates of *M. tuberculosis* in our study had been differentiated into 49 patterns, including 14 clusters, by spoligotyping, MIRU-VNTR typing differentiated all of the clustered isolates into unique types. In fact, MIRU-VNTRs also differentiated two isolates present in a single 12-banded IS*6110*-RFLP cluster into unique types. Though unusual for high-banded strains, similar findings have been reported by other investigators who reported isolates within IS*6110*-RFLP clusters that differed by two-seven loci in a 24 loci MIRU-VNTR assay (Shamputa et al. 2010). We used primers for 11 loci from the original set of 12 loci used in earlier studies (Supply et al. 2000). Valcheva et al. (2008) compared the use of 12, 15 and 24 loci to genotype 73 strains of *M. tuberculosis*. They could achieve a discriminatory index of 0.994, 0.996 and 0.997 with primers for the 12, 15 and 24-loci sets, respectively. They also recommended using a set of five MIRU loci (Mtub04, Mtub21, QUB11b, QUB26 and MIRU 40) to achieve a good discrimination for use in the first-line typing of *M. tuberculosis* strains in Bulgaria, although this discrimination was lower than that achieved with the 12, 15 and 24-loci sets.

Very few studies on MIRU-VNTR typing have been conducted in India (Gutierrez et al. 2006, Sharma et al. 2008, Stavrum et al. 2009, Narayanan et al. 2010). Gutierrez et al. (2006) studied 91 isolates of *M. tuberculosis* from 12 different regions in India, using 12 MIRU-VNTR loci and nine additional VNTR loci. They obtained 78 distinct genotypes, including six clusters and 72 unique patterns. Stavrum et al. (2009) analysed 65 isolates of *M. tuberculosis* using 15 loci. Of the 65 isolates, 21 were distributed into 10 clusters. Isolates in five of these clusters could be discriminated by spoligotyping and/or SNP analysis of the *katG/rpoB* genes. A recent study from India (Sharma et al. 2008) distinguished 97 isolates of *M. tuberculosis* into unique isolates by using primers for

six MIRU loci. However, when we used these six loci to analyse our isolates, we obtained 98 patterns with three clusters of two strains each and 95 unique isolates. Further analysis with the additional five loci used in the present study differentiated the six isolates into unique types, forming three clusters. In another recent study, Narayanan et al. (2010) used 15 MIRU loci to study paired isolates from 25 HIV-positive and 23 HIV-negative patients with TB. Their main aim was to study the recurrence of TB in South India in these patients. However, in spite of using 15 loci, they found six MIRU clusters of two isolates each, which could be differentiated by spoligotyping or IS*6110* typing. Similarly, the single MIRU cluster in our study was differentiated by spoligotyping.

As in the study by Sharma et al. (2008), we found MIRU locus 26 to be highly polymorphic with an allelic diversity of 0.75. Stavrum et al. (2009) also found MIRU 26 to be the most discriminatory locus for isolates belonging to the CAS1_DELHI sub-lineage. However, in our study, locus 31 was equally discriminatory with an allelic diversity of 0.76. Unlike studies by other groups (Cowan et al. 2002, 2005, Sola et al. 2003, Sun et al. 2004, Supply et al. 2006), MIRU locus 24 was polymorphic in the current study, with an allelic diversity of 0.6 and four alleles. Valcheva et al. (2008) found MIRU locus 24 to have only a single allele, whereas Sun et al. (2004) found a maximum of two alleles at this locus. This reinforces the observation that first-line typing schemes for different countries may require the use of different loci because the population structures of circulating *M. tuberculosis* strains vary across different parts of the world (Supply et al. 2006). Moreover, the relative discriminatory powers of particular VNTR loci vary depending on the strain in question (Murase et al. 2008, Comas et al. 2009).

In conclusion, the DR locus typing resulted in 14 clusters (2-21 isolates per cluster) with a clustering rate of 65%, which is in agreement with its slow molecular clock (Kremer et al. 1999). Thus, spoligotyping alone is not discriminatory enough for use in our setting, which exhibits a highly conserved population structure of the tubercle bacilli. At the same time, of the 80 strains on which IS*6110* typing was performed, as many as 18 (22.5%) had ≤ 6 copies of IS*6110* bands, a fact that led to insufficient discrimination by this method alone. By contrast, MIRU-VNTR typing differentiated all 101 strains into unique isolates with the exception of two strains. Thus, although our results clearly show the discriminatory potential of spoligotyping in conjunction with MIRU typing compared to IS*6110*-RFLP alone, further studies on a larger number of *M. tuberculosis* isolates from India are needed to optimise the MIRU loci useful for large-scale genotyping.

REFERENCES

Bhanu NV, van Soolingen D, van Embden JD, Dar L, Pandey RM, Seth P 2002. Predominance of a novel *Mycobacterium tuberculosis* genotype in the Delhi region of India. *Tuberculosis (Edinb) 82*: 105-112.

Brudey K, Driscoll J, Rigouts L, Prodinger W, Gori A, Al-Hajoj S, Allix C, Aristimuno L, Arora J, Baumanis V, Binder L, Cafrune P, Cataldi A, Cheong S, Diel R, Ellermeier C, Evans J, Fauville-Dufaux M, Ferdinand S, de Viedma DG, Garzelli C, Gazzola L,

Gomes H, Gutierrez MC, Hawkey P, van Helden P, Kadival G, Kreiswirth B, Kremer K, Kubin M, Kulkarni S, Liens V, Lillebaek T, Minh Ly H, Martin C, Martin C, Mokrousov I, Narvskaia O, Ngeow YF, Naumann L, Niemann V, Parwati I, Rahim M, 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 P, Tortoli E, Tracevska T, Vincent V, Victor T, Warren R, Yap SF, Zaman K, Portaels F, Rastogi N, Sola C 2006. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol 6*: 6-23.

Chauhan A, Chauhan DS, Parashar D, Gupta P, Sharma VD, Sachan AS, Gupta R, Agarawal BM, Katoch VM 2004. DNA fingerprinting of *Mycobacterium tuberculosis* isolates from Agra region by is *6110* probe. *Indian J Med Microbiol 22*: 238-240.

Comas I, Homolka S, Niemann S, Gagneux S 2009. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS ONE 4*: e7815.

Cowan LS, Diem L, Monson T, Wand P, Temporado D, Oemig TV, Crawford JT 2005. Evaluation of a two-step approach for large-scale, prospective genotyping of *Mycobacterium tuberculosis* isolates in the United States. *J Clin Microbiol 43*: 688-695.

Cowan LS, Mosher L, Diem L, Massey JP, Crawford JT 2002. Variable-number tandem repeat typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110* by using mycobacterial interspersed repetitive units. *J Clin Microbiol 40*: 1592-1602.

Das S, Paramasivan CN, Lowrie DB, Prabhakar R, Narayanan PR 1995. IS*6110* restriction fragment length polymorphism typing of clinical isolates of *Mycobacterium tuberculosis* from patients with pulmonary tuberculosis in Madras, South India. *Tuber Lung Dis 76*: 550-554.

Gutierrez MC, Ahmed N, Willery E, Narayanan S, Hasnain SE, Chauhan DS, Katoch VM, Vincent V, Locht C, Supply P 2006. Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India. *Emerg Infect Dis 12*: 1367-1374.

Helal ZH, Ashour MS, Eissa SA, Abd-Elatef G, Zozio T, Babapoor S, Rastogi N, Khan MI 2009. Unexpectedly high proportion of ancestral Manu genotype *Mycobacterium tuberculosis* strains cultured from tuberculosis patients in Egypt. *J Clin Microbiol 47*: 2794-2801.

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 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol 35*: 907-914.

Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PW, Martin C, Palittapongarnpim P, Plikaytis BB, Riley LW, Yakrus MA, Musser JM, van Embden JD 1999. 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 37*: 2607-2618.

Kulkarni S, Sola C, Filliol I, Rastogi N, Kadival G 2005. Spoligotyping of *Mycobacterium tuberculosis* isolates from patients with pulmonary tuberculosis in Mumbai. *India Res Microbiol 156*: 588-596.

Kurepina NE, Sreevatsan S, Plikaytis BB, Bifani PJ, Connell ND, Donnelly RJ, van Sooligen D, Musser JM, Kreiswirth BN 1998. Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS*6110* elements in *Mycobacterium tuberculosis*: non-random integration in the dnaA-dnaN region. *Tuber Lung Dis 79*: 31-42.

Mathuria JP, Sharma P, Prakash P, Samaria JK, Katoch VM, Anupurba S 2008. Role of spoligotyping and IS*6110*-RFLP in assessing genetic diversity of *Mycobacterium tuberculosis* in India. *Infect Genet Evol 8*: 346-351.

Mistry NF, Iyer AM, D'Souza DT, Taylor GM, Young DB, Antia NH 2002. Spoligotyping of *Mycobacterium tuberculosis* isolates from multiple-drug-resistant tuberculosis patients from Bombay, India. *J Clin Microbiol 40*: 2677-2680.

Murase Y, Mitarai S, Sugawara I, Kato S, Maeda S 2008. Promising loci of variable numbers of tandem repeats for typing Beijing family *Mycobacterium tuberculosis*. *J Med Microbiol 57*: 873-880.

Narayanan S, Das S, Garg R, Hari L, Rao VB, Frieden TR, Narayanan PR 2002. Molecular epidemiology of tuberculosis in a rural area of high prevalence in South India: implications for disease control and prevention. *J Clin Microbiol 40*: 4785-4788.

Narayanan S, Gagneux S, Hari L, Tsolaki AG, Rajasekhar S, Narayanan PR, Small PM, Holmes S, Deriemer K 2008. Genomic interrogation of ancestral *Mycobacterium tuberculosis* from South India. *Infect Genet Evol 8*: 474-483.

Narayanan S, Swaminathan S, Supply P, Shanmugam S, Narendran G, Hari L, Ramchandran R, Locht C, Jawahar MS, Narayanan PR 2010. Impact of HIV infection on the recurrence of tuberculosis in South India. *J Infect Dis 201*: 691-703.

Radhakrishnan I, Manju YK, Kumar RA, Mundayoor S 2001. Implications of low frequency of IS*6110* in fingerprinting field isolates of *Mycobacterium tuberculosis* from Kerala, India. *J Clin Microbiol 39*: 1683.

Rastogi N, Sola C 2007. Molecular evolution of the *Mycobacterium tuberculosis* complex. In JC Palomino, S Cardoso Leão, V Ritacco (eds.), *Tuberculosis-2007: from basic science to patient care*, Amadeo Online Textbooks, available from: Tuberculosis-Textbook.com.

RNTCP Status Report - Revised National TB Control Programme 2008. Central TB Division. Directorate General of Health Services, Ministry of Health and Family welfare, Nirman Bhawan, New Delhi. Available from: tbcindia.org.

Selander RK, Caugant DA, Ochman H, Musser, Gilmour MN, Whittam TS 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol 51*: 873-884.

Shamputa IC, Lee J, Allix-Béguec C, Cho EJ, Lee JI, Rajan V, Lee EG, Min JH, Carroll MW, Goldfeder LC, Kim JH, Kang HS, Hwang S, Eum SY, Park SK, Lee H, Supply P, Cho SN, Via LE, Barry CE 3rd 2010. Genetic diversity of *Mycobacterium tuberculosis* isolates from a tertiary care tuberculosis hospital in South Korea. *J Clin Microbiol 48*: 387-394.

Sharma P, Chauhan DS, Upadhyay P, Faujdar J, Lavania M, Sachan S, Katoch K, Katoch VM 2008. Molecular typing of *Mycobacterium tuberculosis* isolates from a rural area of Kanpur by spoligotyping and mycobacterial interspersed repetitive units (MIRUs) typing. *Infect Genet Evol 8*: 621-626.

Siddiqi N, Shamim M, Amin A, Chauhan DS, Das R, Srivastava K, Singh D, Sharma VD, Katoch VM, Sharma SK, Hanief M, Hasnain SE 2001. Typing of drug resistant isolates of *Mycobacterium tuberculosis* from India using the IS*6110* element reveals substantive polymorphism. *Infect Genet Evol 1*: 109-116.

Singh UB, Arora J, Suresh N, Pant H, Rana T, Sola C, Rastogi N, Pande JN 2007. Genetic biodiversity of *Mycobacterium tuberculosis* isolates from patients with pulmonary tuberculosis in India. *Infect Genet Evol 7*: 441-448.

Singh UB, Suresh N, Bhanu NV, Arora J, Pant H, Sinha S, Aggarwal RC, Singh S, Pande JN, Sola C, Rastogi N, Seth P 2004. Predominant tuberculosis spoligotypes, Delhi, India. *Emerg Infect Dis 10*: 1138-1142.

Sola C, Filliol I, Legrand E, Lesjean S, Locht C, Supply P, Rastogi N 2003. Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. *Infect Genet Evol 3*: 125-133.

Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci USA 94*: 9869-9874.

Stavrum R, Myneedu VP, Arora VK, Ahmed N, Grewal HMS 2009. In-depth molecular characterization of *Mycobacterium tuberculosis* from New Delhi - predominance of drug resistant isolates of the 'Modern' (TbD1) type. *PLoS ONE 4*: e4540.

Sun YJ, Bellamy R, Lee AS, Ng ST, Ravindran S, Wong SY, Locht C, Supply P, Paton NI 2004. Use of mycobacterial interspersed repetitive unit-variable-number tandem repeat typing to examine genetic diversity of *Mycobacterium tuberculosis* in Singapore. *J Clin Microbiol 42*: 1986-1993.

Suresh N, Singh UB, Arora J, Pant H, Seth P, Sola C, Rastogi N, Samantaray JC, Pandey JN 2006. *rpoB* gene sequencing and spoligotyping of multidrug resistant *Mycobacterium tuberculosis* isolates from India. *Infect Genet Evol 6*: 474-483.

Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rusch-Gerdes S, Willery E, Savine E, de Haas P, van Deutekom H, Roring S, Bifani P, Kurepina N, Kreiswirth B, Sola C, Rastogi N, Vatin V, Gutierrez MC, Fauville M, Niemann S, Skuce R, Kremer K, Locht C, van Soolingen D 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol 44*: 4498-4510.

Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Locht C 2001. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J Clin Microbiol 39*: 3563-3571.

Supply P, Mazars E, Lesjean S, Vincent V, Gicquel B, Locht C 2000. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol Microbiol 36*: 762-771.

Valcheva V, Mokrousov I, Narvskaya O, Rastogi N, Markova N 2008. Utility of new 24-locus variable-number tandem-repeat typing for discriminating *Mycobacterium tuberculosis* clinical isolates collected in Bulgaria. *J Clin Microbiol 46*: 3005-3011.

van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM, Small PM 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol 31*: 406-409.

van Soolingen D, Hermans PW, de Haas PE, Soll DR, van Embden JD 1991. 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 29*: 2578-2586.