JOURNAL OF CLINICAL MICROBIOLOGY, Dec. 2005, p. 5978–5982 Vol. 43, No. 12
0095-1137/05/$08.00+0 doi:10.1128/JCM.43.12.5978–5982.2005

Analysis of Genomic Downsizing on the Basis of Region-of-Difference Polymorphism Profiling of *Mycobacterium tuberculosis* Patient Isolates Reveals Geographic Partitioning

K. Rajender Rao,[1] Farhana Kauser,[1] Sriramula Srinivas,[1] Stefania Zanetti,[2] Leonardo A. Sechi,[2] Niyaz Ahmed,[1] and Seyed E. Hasnain[1,3]*

Pathogen Evolution Group, Laboratory of Molecular and Cell Biology, Centre for DNA Fingerprinting and Diagnostics (CDFD), Nacharam, Hyderabad, India[1]; Department of Biomedical Sciences, University of Sassari, Sassari, Italy[2]; and Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Jakkur, Bangalore, India[3]

Received 10 April 2005/Returned for modification 24 June 2005/Accepted 2 September 2005

***Mycobacterium tuberculosis*, the etiological agent of tuberculosis, has lost many coding and noncoding regions in its genome during the course of evolution. We performed region-of-difference (RD) analysis using PCR-based genotyping of 131 *M. tuberculosis* clinical isolates obtained from four different countries, namely, India, Peru, Libya, and Angola. Our studies revealed that RD patterns are often distinct for strains circulating in specific geographical regions and can be used to trace the descent and spread of an isolate from its original reservoir. We describe our findings, which show that no single isolate from the four countries ($n = 131$) had all the 15 RDs either deleted or retained. Tuberculosis-specific deletion 1 (TbD1) was found to be conserved in 23% of the Indian isolates, indicating their possible ancient origin. RD9 was the most conserved region, RD11 was predominantly deleted, and RD6 was the most variable among the isolates in our collection irrespective of their geographic region. In contrast to earlier reports, our results demonstrate that the deletion of RD1 does not correlate with a decrease in the virulence potential of *M. tuberculosis*, as Indian isolates ($n = 30$) examined by us were from diseased individuals and yet had lost the RD1 region. Our results further illustrated that the intactness of the RD5 region may be associated with increased virulence of the organism. This study highlights that the RDs in *M. tuberculosis* genomes are geographically distributed and specific and may possibly be associated with virulence spectrum.**

Tuberculosis (TB) remains an important cause of mortality worldwide; treatment of the disease has been further complicated by the increased number of human immunodeficiency virus-TB (HIV-TB)-coinfected patients. The future does not appear to be bright, as no breakthrough in the development of drugs and vaccines has been reported even after 8 years of availability of the complete genome sequence of *Mycobacterium tuberculosis*. Knowledge of genetic variability, mycobacterial population dynamics, and evolutionary genetics holds a lot of promise for understanding epidemic potentials of strains, their host tropism, differences in virulence, antibiotic susceptibility, etc., and will be possibly important for the treatment and control of the disease (26, 30).

Sequence analyses of the *M. tuberculosis* complex isolates have revealed that allelic polymorphism is rare, occurring at the rate of ~1 in 10,000 bp (28). As a result, the genomic fluidity in *M. tuberculosis* complex is largely due to nonrandom deletions apart from mobility of insertion sequences and transduction mediated by mycobacteriophages. Several PCR-based approaches have been developed to analyze the genomic diversity of the *M. tuberculosis* complex (1, 2, 12, 15, 16, 19, 22, 25, 27, 29, 31). The available strain-typing data revealed that regions of deletion are not randomly distributed throughout the genome but instead tend to aggregate, suggesting that a genomic region is vulnerable to changes which may prove deleterious to the pathogen. Genomic deletions provide information on the diversity and frequency of polymorphism in the mycobacterial populations (9). These could therefore be harnessed in conjunction with other genotyping techniques to study the evolution of the *M. tuberculosis* complex and to predict the clonal descent of geographically partitioned strain groups.

Recent developments in comparative genomics revealed differences in the presence or absence of regions of differences (RDs) and tuberculosis-specific deletion 1 (TbD1) between the *M. tuberculosis* H37Rv and the *M. bovis* BCG. These genomic landmarks have been described as potential markers for understanding the historical origins and genealogy of the present-day mycobacterial pathogens (3, 5, 10). These regions harbor several important genes and virulence factors, and their presence or absence could help identify lineages of different isolates in a particular geographical region on an evolutionary time scale. RD patterns have emerged as important for typing systems for epidemiological and evolutionary studies of *M. tuberculosis*. Subsequent studies found that some of these deletions were variable and were also absent from other subspecies of the genus *Mycobacterium* (14). Of these RDs, RD3 and RD11 are now classified as mobile genetic elements whereas RD5 and RD6 are the insertion sequences. Some of these RDs, namely, RD1, RD2, RD4, RD7, RD8, RD9, RD10, RD12, RD13, RD14, and TbD1, contain the coding sequences. Deletion or truncation of these RDs may be due to rare strand-slippage errors of DNA polymerase, although the exact mechanism of such events is still elusive (5). These deletion patterns

* Corresponding author. Mailing address: Centre for DNA Fingerprinting and Diagnostics, Hyderabad 500 076, India. Phone: 91 40 2715 5604. Fax: 91 40 2715 5610. E-mail: hasnain@cdfd.org.in.

promise to be an important source of variability among the members of the *M. tuberculosis* complex and represent unidirectional genetic events having a major role in the evolution of *Mycobacterium* species (3).

In this study, we attempted to analyze the distribution and genotypic characteristics of different RDs in the genome of *M. tuberculosis* isolates derived from four different geographic regions. Although there is a need to analyze isolates from all over the world to understand current global scenario of strain variation, data from some of the highly burdened countries will nonetheless be significant from the point of view of tuberculosis control efforts in these countries.

MATERIALS AND METHODS

Bacterial isolates. The genomic DNAs of 131 *M. tuberculosis* clinical isolates from different geographic regions were obtained from Angola ($n = 15$), Peru ($n = 38$), Libya ($n = 48$), and India ($n = 30$). The Peruvian isolates represented independent patient populations with and without AIDS (isolates 1 to 19 were from HIV-negative TB patients and isolates 20 to 38 were from HIV-positive TB patients) and were cultured at the Hospital Dos de Mayo and Hospital Maria Auxiliadora in Lima, Peru. The Libyan strains were isolated and cultured at the Departimento di Scienze Biomediche, Sassari, Italy, whereas the Indian isolates ($n = 30$) were collected from the Mahavir Hospital and Research Centre, Hyderabad, India. Genomic DNA was extracted according to methods described in earlier studies (24). Isolates were coded numerically according to the patient's geographic origin.

PCR amplification and deletion analysis. To amplify the regions of differences in the genome of *M. tuberculosis* clinical isolates, PCR primers for all the 14 RDs and TbD1 regions were designed according to the method of Brosch et al. (5). Genomic DNA of each of the isolates was amplified with all the primer sets specific to RDs and TbD1. The primer sets used were as follows: for RD1 (9.5-kb genes Rv3871-Rv3879c), the internal primer pair RD1in-Rv3878F (GTC AGC CAA GTC AGG CTA CC) and RD1in-Rv3878R (CAA CGT TGT GGT TGT TGA GG) and the flanking primer pair RD1-flank.left (GAA ACA GTC CCC AGC AGG T) and RD1-flank.right (TTC AAC GGG TTA CTG CGA AT) were used; for RD2 (10.8-kb genes Rv1978-Rv1988), the internal primer pair RD2-Rv1979.int.F (TAT AGC TCT CGG CAG GTT CC) and RD2-Rv1979-int.R (ATC GGC ATC TAT GTC GGT GT) and the flanking primer pair RD2-flank.F (CTC GAC CGC GAC GAT GTG C) and RD2-flank.R (CCT CGT TGT CAC CGC GTA TG) were used; for RD3 (9.2-kb genes Rv1573-Rv1586c), the internal primer pair RD3-Rv1586.int.F (TTA TCT TGG CGT TGA CGA TG) and RD3-Rv1586.int.R (CAT ATA AGG GTG CCC GCT AC) and the second internal primer pair RD3-int-REP.F (CTG ACG TCG TTG TCG AGG TA) and RD3-int-REP.R (GTA CCC CCA GGC GAT CTT) were used; for RD4 (12.7-kb genes Rv1505c-Rv1516c), the internal primer pair RD4-Rv1516.int.F (CAA GGG GTA TGA GGT TCA CG) and RD4-Rv1516.int.R (CGG TGA TTC GTG ATT GAA CA) and the flanking primer pair RD4-flank.F (CTC GTC GAA GGC CAC TAA AG) and RD4-flank.R (AAG GCG AAC AGA TTC AGC AT) were used; for RD5 (9.0-kb genes Rv2346c-Rv2353c), the internal primer pair RD5A-Rv2348.int.F (AAT CAC GCT GCT GCT ACT CC) and RD5A-Rv2348.int.R (GTG CTT TTG CCT CTT GGT C) and the second internal primer pair RD5B-plcA.int.F (CAA GTT GGG TCT GGT CGA AT) and RD5B-plcA.int.R (GCT ACC CAA GGT CTC CTG GT) were used; for RD6 (4.9-kb genes Rv3425-Rv3428c), the internal primer pair RD6-IS1532F (CAG CTG GTG AGT TCA AAT GC) and RD6-IS1532R (CTC CCG ACA CCT GTT CGT) was used (no flanking or second internal primer pair was used); for RD7 (12.7-kb genes Rv1964-Rv1977), the internal primer pair RD7-Rv1976.int.F (TGG ATT GTC GAC GGT ATG AA) and RD7-Rv1976.int.R (GGT CGA TAA GGT CAC GGA AC) and the flanking primer pair RD7-flank.F (GGT AAT CGT GGC CGA CAA G) and RD7-flank.R (CAG CTC TTC CCC TCT CGA C) were used; for RD8 (5.9-kb genes *ephA*-*lpqG*), the internal primer pair RD8-ephA.F (GGT GTG ATT TGG TGA GAC GAT G) and RD8-ephA.R (AGT TCC TCC TGA CTA ATC CAG GC) and the flanking primer pair RD8-flank.F (CAA TCA GGG CTG TGC TAA CC) and RD8-flank.R (CGA CAG TTG TGC GTA CTG GT) were used; for RD9 (2.0-kb genes *cobL*-Rv2075), the internal primer pair RD9-intF (CGA TGG TCA ACA CCA CTA CG) and RD9-intR (CTG GAC CTC GAT GAC CAC TC) and the flanking primer pair RD9-flankF (GTG TAG GTC AGC CCC ATC C) and RD9-flankR (GCC CAA CAG CTC GAC ATC) were used; for RD10 (1.9-kb genes Rv0221-Rv0223), the internal primer pair RD10-intF (GTA ACC GCT TCA CCG GAA T) and RD10-intR (GTC AAC TCC ACG GAA AGA CC) and the flanking primer pair RD10-flankF (CTG CAA CCA TCC GGT ACA C) and RD10-flankR (GTC ATG AAC GCC GGA CAG) were used; for RD11 (11.0-kb genes Rv2645-Rv2659c), the internal primer pair RD11-Rv2646F (CGG CAG CTA GAC GAC CTC) and RD11-Rv2646R (AAC GTG CTG CGA TAG GTT TT) and the flanking primer pair RD11-fla-F (TCA CAT AGG GGC TGC GAT AG) and RD11-fla-R (AGA GGA ACC TTT CGG TGG TT) were used; for RD12 (2.8-kb genes *sseC*-Rv3121), the internal primer pair RD12-Rv3120.int.F (GAA ATA CGA GTG CGC TGA CC) and RD12-Rv3120.int.R (CTC TGA ACC ATC GGT GTC G) and the flanking primer pair RD12-flank.F (GCC ATC AAC GTC AAG AAC CT) and RD12-flank.R (CGG CCA GGT AAC AAG GAG T) were used; for RD13 (3.0-kb genes Rv1255c-Rv1257c), the internal primer pair RD13intF (GGA TGT CAC TCG GAA CGG CA) and RD13intR (CAC CGG GCT GAT CGA GCG A) and the flanking primer pair RD13-flank.F (CGA TGG TGT TTC TTG GTG AG) and RD13-flank.R (GGA TCG GCT CAG TGA ATA CC) were used; for RD14 (9.0-kb genes Rv1765c-Rv1773c), the internal primer pair RD14-Rv1769.int.F (GTG GAG CAC CTT GAC CTG AT) and RD14-Rv1769.int.R (CGT CGA ATA CGA GTC GAA CA) and the flanking primer pair RD14-flankF (TTG ATT CGC CAA CAA CTG AA) and RD14-flankR (GGG CTG GTT AGT GTC GAT TC) were used; and for TbD1 (genes *mmpS6* and *mmpL6*), the internal primer pair TBD1intS.F (CGT TCA ACC CCA AAC AGG TA) and TBD1intS.R (AAT CGA ACT CGT GGA ACA CC) and the flanking primer pair TBD1fla1-F (CTA CCT CAT CTT CCG GTC CA) and TBD1fla1-R (CAT AGA TCC CGG ACA TGG TG) were used.

Each PCR assay was carried out with a 20-μl reaction mixture containing 50 nM of each primer, 10 mM Tris-HCl (pH 8.8), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 10% dimethyl sulfoxide, 200 μM nucleotide mix, 1 U *Taq* polymerase, and 1 to 10 ng of template DNA, and the volume was made up with sterile distilled water. After denaturation at 95°C for 3 min, the reaction mixture was processed for another 35 cycles at 95°C for 30 s, 58°C for 1 min, and 72°C for 4 min followed by 72°C for 10 min in a Gene Amp 9700 system (Perkin-Elmer). The amplicons were visualized by electrophoresis, which was carried out at 125 V for 1 h on a 1% agarose gel in Tris-acetate-EDTA buffer, and the amplicon sizes were estimated by comparison with a 1-kb Plus DNA ladder (Gibco BRL, Life Technologies). Deletion polymorphism were scored based on the size of different amplicons on an agarose gel. The RD region was amplified with flanking primers; for confirmation purposes, amplification using internal primers was also performed. All the methods in this regard were optimized as suggested by Brosch et al. (6).

RESULTS

RDs are geographically partitioned. The genomic DNAs of 131 *M. tuberculosis* isolates belonging to four countries were screened for 15 loci (14 RDs and TbD1). On the basis of deletion markers, a dendrogram was constructed to understand the genetic relationships and distances among the isolates from the four countries. Data of all 131 isolates were used in a binary format for construction of an unrooted tree based on the information of RD patterns (Fig. 1). The tree thus obtained clearly revealed four distinct clusters. As expected, isolates belonging to a specific geographical region clustered together as a single group. They grouped into four clusters, namely, Angola (isolates 1 to 15), Peru (isolates 1 to 19 and 20 to 38), Libya (isolates 1 to 48), and India (isolates 1 to 30). Each cluster was found to be geographically specific. Each region-specific major cluster was divided into subclusters on the basis of the variation in the RD patterns. The 15 RDs analyzed for 131 isolates formed 62 distinct patterns. No similar RD patterns were found among isolates belonging to different countries.

Distribution of deletion polymorphism among our isolates. Out of the total of 131 isolates tested, only 8 (6.1%) retained the TbD1 region (Fig. 2). The TbD1 region was present in seven (23%) of the Indian isolates and accounts for 5.3% of

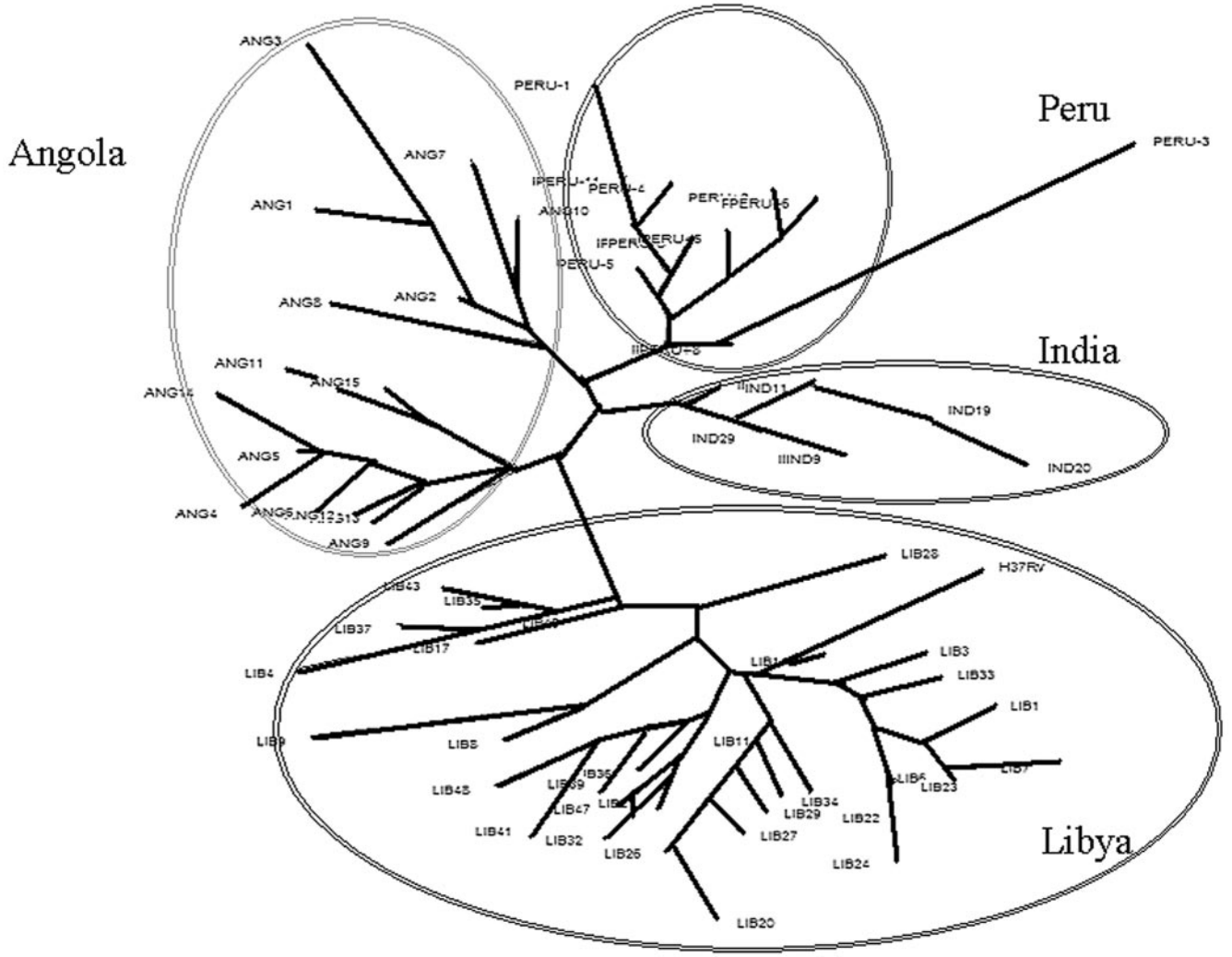

FIG. 1. A radial dendrogram depicting four *M. tuberculosis* clusters belonging to four different countries constructed based on binary data using the neighbor joining method.

the total isolates analyzed, indicating that these isolates probably represent the ancient gene pool. A single Angolan isolate (Angola 7) retained this region. All the isolates from Peru and Libya showed the absence of the TbD1. All the clinical isolates of Indian origin analyzed by us lacked the RD1 region. However, the extent of RD1 deletion in strains of other regions was comparatively less. Forty percent of Angolan isolates had deletions of RD1. The RD1 deletion was also observed in 21% of Libyan and 2.6% of Peruvian isolates.

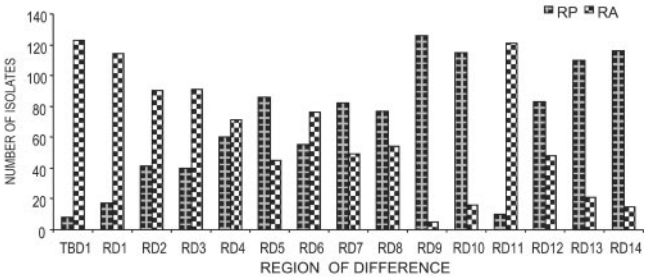

FIG. 2. RD distribution. The overall scenario for the presence (RP) or absence (RA) of a specific region is shown.

Among the 131 isolates we tested, RD9 was deleted only in five isolates (Angola 1, 3, and 11 and Peru 1 and 3) and present in 126 (96%) isolates. Therefore, RD9 was found to be the most conserved RD among the isolates tested. The RD11 region was absent in 98% of Libyan, 84% of Peruvian, and 80% of Angolan isolates, and all the Indian *M. tuberculosis* isolates tested lacked this region. The RD10 and RD14 regions were present in 100% of Indian and Angolan isolates. The RD10, RD13, and RD14 regions were present in 97% of Peruvian isolates. The RD7 region was absent in about 94% of Angolan and 84% of Peruvian isolates. The RD2, RD3, and RD4 regions were totally absent in all the Peruvian and Indian isolates and also absent in 80%, 87%, and 20% of the Angolan isolates, respectively. The RD8 region was absent in all the Indian isolates, whereas RD12 and RD5 were absent in Peruvian and Angolan isolates, respectively. On the other hand, RD2, RD3, RD4, RD5, RD6, RD8, and RD12 were retained in the majority of Libyan isolates.

Overall, we found RD9 to be the most conserved, RD11 to be predominantly deleted, and RD6 as to be the most variable one, irrespective of the origin of isolates. No single isolate among all the four countries ($n = 131$) had all the 15 RDs either deleted or retained. One of the Peruvian isolates (Peru 3) had lost all the RDs except RD5, whereas the three Libyan isolates (Libya 19, Libya 26, and Libya 30) retained 12 RDs and lost only three RDs (TbD1, RD11, and RD13). Each of these isolates was found to have a deletion of at least 3 RDs out of 15 RDs tested; these deletions account for 20% of the total region of differences.

Region-wise distribution of signature RDs. The distribution of deletion polymorphism in our international collection of isolates, based on the presence or absence of 15 regions of differences, revealed significant regional diversity. The presence of RD9, RD10, andRD14 could be regarded as the signature of the Indian isolates. The presence of RD10, RD13, and RD14 was, however, specific to Angolan isolates. Libyan isolates almost always revealed a signature represented by the intactness of RD4, RD9, and RD12. Interestingly, the missing RD5 and RD6 regions in all the Indian isolates correlated with low or null copy numbers of insertion element IS*6110*, whereas these two RDs were present in majority of the Libyan isolates.

DISCUSSION

The comparative genomics tools have created an unprecedented opportunity to undertake large-scale sequence comparisons among the *M. tuberculosis* complex organisms. Various approaches such as bacterial artificial chromosome arrays (4, 10), subtractive hybridization (18), and DNA microarrays (3, 20) have identified several small or large regions of differences among the members of the *M. tuberculosis* complex. Genomic deletions have been an important driving force in the chromosomal evolution of the *M. tuberculosis* complex (5, 14, 23). These deletion events often arise due to recombination between adjacent insertion sequence (IS) elements (6, 8). The genome of tubercle bacilli houses more than 40 IS and mobile genetic elements that could mediate deletion events (7) independent of some of the rare strand-slippage errors of DNA polymerases. Genomic deletion analysis of the *M. tuberculosis* complex bacilli confirmed its usefulness in deducing the origins and descent of circulating strains of *M. tuberculosis* complex in different countries (5, 13, 20, 30). The genes from RD regions showed great variability among certain members of the *M. tuberculosis* complex, and some new specific deletions in *M. canettii*, *M. microti*, and seal isolates have been identified (20). Genomic deletions represent unidirectional genetic events in the evolution of *M. bovis* BCG (3) and have been described to be potentially significant in documenting clonal descent of the strains.

In the present study, we sought to validate the use of genomic deletion markers to characterize a collection of strains originating from four different countries of the world for which such an analysis has not been done previously.

RD analysis of 131 isolates revealed four major geographical clusters, making this perhaps the first report that demonstrates specific, biogeographic distribution of RD signatures in the chromosomes of *M. tuberculosis* isolates from different countries. The distribution of the observed deletions may help in understanding their order of occurrence during the course of bacterial evolution (21, 26). The TbD1 junction sequence conservation is a perfect indicator that differentiates ancient and modern isolates. This study demonstrated that 23% of the Indian *M. tuberculosis* isolates have a conserved TbD1 region and thus may be regarded as "ancient" in origin. This observation further substantiates the results of a recent study performed in our laboratories (2).

The genes present in the RD1 region have been described to be essential for *M. tuberculosis* virulence, and loss of RD1 was previously shown to be associated with BCG attenuation (17). Our study provide a further indication that virulence may not be necessarily dictated by the presence of RD1 region, as all the Indian clinical isolates analyzed by us showed a deleted RD1, given the fact that they were obtained from patients with active tuberculosis.

All but five of the isolates from our collection harbored RD9, the deletion of which is considered a signature of *M. africanum* and *M. tuberculosis* complex bacteria other than *M. tuberculosis* (3, 22). Our isolates that revealed RD9 deletion were probably *M. tuberculosis* complex bacteria and not *M. tuberculosis.*

We also attempted to differentiate Peruvian isolates recovered from HIV-seropositive and -seronegative patients. All the Peruvian isolates analyzed with 15 RDs did not show any significant difference between these two categories. However, one of these isolates (Peru 3) recovered from an HIV-seronegative patient showed deletions of 14 RDs from its genome but not deletion of RD5, which is comprised of eight genes, two of which code for ESAT-6-like proteins with unknown functions (5). All other Peruvian and Indian isolates, however, showed deletions of most of the RDs. In this connection, it was interesting that low-copy-number IS*6110* isolates from India did not carry RD5. Recent observations revealed that the low-copy-number IS*6110* strains of principle genetic group 1 predominate south Asian regions and are described as low in virulence (11). However, it is too early to speculate on an association of RD5 with virulence. The presence or absence of different genomic deletions in different strains might possibly be dictated by the selection pressures exerted by the host immune mechanisms (and the genetic background) determining the repertoire of various RDs in different strains. Further, these mechanisms might actually shape the genome content of the bacterium over time, which in turn suggests that successive loss of DNA facilitates appearance of more-successful pathogens in certain kind of hosts (5).

In conclusion, we find evidence of an association between RD patterns and the biogeographic evolution of strains. The TbD1 region was more frequently present in most of the Indian isolates, supporting the concept that Indian *M. tuberculosis* strains are more "ancient." Our study also provokes second thoughts on the deletion of RD1 and suggests that this may not decrease the virulence potential of *M. tuberculosis*. This is because a majority of the Indian patient isolates from diseased individuals lacked this region. Similarly, other genes present within the RD 5 loci may also be important from the perspective of geographic evolution and impact of host factors. In the future, it may be of interest to carefully compare the virulence properties of strains with different RD profiles by using a variety of available animal models to obtain novel insights into

the evolutionary dynamics and virulence mechanisms of this major pathogen. Secondly, our study demonstrated the potential of deletion markers in lineage identification, phylogeny, and molecular epidemiology and may therefore be used in conjunction with the existing typing methods. We suggest that RD analysis of a larger number of *M. tuberculosis* isolates from various geographic regions may give more insights and contribute to our understanding of the evolution, epidemiology, and pathogenicity of *M. tuberculosis*.

ACKNOWLEDGMENTS

We are thankful to Robert H. Gilman for providing *M. tuberculosis* genomic DNA samples of the strains from Peru. K. J. R. Murthy is gratefully acknowledged for providing Indian isolates. We thank A. P. Suresh Kumar and R. Sugandhan for their help.

Financial support in the form of core grants to the Centre for DNA Fingerprinting and Diagnostics by the Department of Biotechnology (DBT), Government of India, is gratefully acknowledged.

REFERENCES

1. **Ahmed, N., L. Caviedes, M. Alam, K. R. Rao, V. Sangal, P. Sheen, R. H. Gilman, and S. H. Hasnain.** 2003. Distinctiveness of *Mycobacterium tuberculosis* genotypes from human immunodeficiency virus type 1-seropositive and -seronegative patients in Lima, Peru. J. Clin. Microbiol. **41:**1712–1716.
2. **Ahmed, N., M. Alam, K. R. Rao, F. Kauser, N. A. Kumar, N. N. Qazi, V. Sangal, V. D. Sharma, R. Das, V. M. Katoch, K. J. Murthy, S. Suneetha, S. K. Sharma, L. A. Sechi, R. H. Gilman, and S. E. Hasnain.** 2004. Molecular genotyping of a large, multicentric collection of tubercle bacilli indicates geographical partitioning of strain variation and has implications for global epidemiology of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **42:**3240–3247.
3. **Behr, M. A., M. A. Wilson, W. P. Gill, H. Salamon, G. K. Schoolnik, S. Rane, and P. M. Small.** 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science **284:**1520–1523.
4. **Brosch, R., S. V. Gordon, A. Billault, T. Garnier, K. Eiglmeier, C. Soravito, B. G. Barrell, and S. T. Cole.** 1998. Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. Infect. Immun. **66:**2221–2229.
5. **Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole.** 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc. Natl. Acad. Sci. USA **99:**3684–3689.
6. **Brosch, R., W. J. Philipp, E. Stavropoulos, M. J. Colston, S. T. Cole, and S. V. Gordon.** 1999. Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra strain. Infect. Immun. **67:**5768–5774.
7. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeie, S. Gas, C. E. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, et al.** 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393:**537–544.
8. **Fang, Z., C. Doig, D. T. Kenna, N. Smittipat, P. Palittapongarnpim, B. Watt, and K. J. Forbes.** 1999. IS*6110*-mediated deletions of wild-type chromosomes of *Mycobacterium tuberculosis*. J. Bacteriol. **181:**1014–1020.
9. **Goguet de la Salmoniere, Y. O., C. C. Kim, A. G. Tsolaki, A. S. Pym, M. S. Siegrist, and P. M. Small.** 2004. High-throughput method for detecting genomic-deletion polymorphisms. J. Clin. Microbiol. **42:**2913–2918.
10. **Gordon, S. V., R. Brosch, A. Billault, T. Garnier, K. Eiglmeier, and S. T. Cole.** 1999. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Mol. Microbiol. **32:**643–655.
11. **Gutierrez, M. C., N. Ahmed, E. Willery, S. Narayanan, S. E. Hasnain, D. S. Chauhan, V. M. Katoch, V. Vincent, C. Locht, and P. Supply.** Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India denotes an ancient focus of tuberculosis in South Asia. (Unpublished.)
12. **Hermans, P. W. M., D. van Soolingen, J. W. Dale, A. R. J. Schuitema, R. A. McAdam, D. Catty, and J. D. A. van Embden.** 1990. Insertion element IS*986* from *Mycobacterium tuberculosis*: a useful tool for diagnosis and epidemiology of tuberculosis. J. Clin. Microbiol. **28:**2051–2058.
13. **Hirsh, A. E., A. G. Tsolaki, K. De Riemer, M. W. Feldman, and P. M. Small.** 2004. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. Proc. Natl. Acad. Sci. USA **101:**4871–4876.
14. **Kato-Maeda, M., J. T. Rhee, T. R. Gingeras, H. Salamon, J. Drenkow, N. Smittipat, and P. M. Small.** 2001. Comparing genomes within the species *Mycobacterium tuberculosis*. Genome Res. **11:**547–554.
15. **Kox, L. F., J. van Leeuwen, S. Knijper, H. M. Jansen, and A. H. Kolk.** 1995. PCR assay based on DNA coding for 16S rRNA for detection and identification of mycobacteria in clinical samples. J. Clin. Microbiol. **33:**3225–3233.
16. **Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. van Embden.** 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.
17. **Lewis, K. N., R. Liao, K. M. Guinn, M. J. Hickey, S. Smith, M. A. Behr, and D. R. Sherman.** 2003. Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guerin attenuation. J. Infect. Dis. **187:**117–123.
18. **Mahairas, G. G., P. J. Sabo, M. J. Hickey, D. C. Singh, and C. K. Stover.** 1996. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. J. Bacteriol. **178:**1274–1282.
19. **Majeed, A. A., N. Ahmed, K. R. Rao, S. Ghousunnissa, F. Kauser, B. Bose, H. A. Nagarajaram, V. M. Katoch, D. V. Cousins, L. A. Sechi, R. H. Gilman, and S. E. Hasnain.** 2004. AmpliBASE MT: a *Mycobacterium tuberculosis* diversity knowledge base. Bioinformatics **20:**989–992.
20. **Marmiesse, M., P. Brodin, C. Buchrieser, C. Gutierrez, N. Simoes, V. Vincent, P. Glaser, S. T. Cole, and R. Brosch.** 2004. 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. Microbiology **150:**483–496.
21. **Mostowy, S., D. Cousins, J. Brinkman, A. Aranaz, and M. A. Behr.** 2002. Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. J. Infect. Dis. **186:**74–80.
22. **Parsons, L. M., R. Brosch, S. T. Cole, A. Somoskovi, A. Loder, G. Bretzel, D. van Soolingen, Y. M. Hale, and M. Salfinger.** 2002. Rapid and simple approach for identification of *Mycobacterium tuberculosis* complex isolates by PCR-based genomic deletion analysis. J. Clin. Microbiol. **40:**2339–2345.
23. **Pym, A. S., and R. Brosch.** 2000. Tools for the population genomics of the tubercle bacilli. Genome Res. **10:**1837–1839.
24. **Siddiqi, N., M. Shamim, A. Amin, D. S. Chauhan, R. Das, K. Srivastava, D. Singh, V. D. Sharma, V. M. Katoch, S. K. Sharma, M. Hanief, and S. E. Hasnain.** 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.
25. **Singh, U. B., N. Suresh, N. V. Bhanu, J. Arora, H. Pant, S. Sinha, R. C. Aggarwal, S. Singh, J. N. Pande, C. Sola, and P. Seth.** 2004. Predominant tuberculosis spoligotypes, Delhi, India. Emerg. Infect. Dis. **10:**1138–1142.
26. **Snel, B., P. Bork, and M. A. Huynen.** 1999. Genome phylogeny based on gene content. Nat. Genet. **21:**108–110.
27. **Soini, H., X. Pan, L. Teeter, J. M. Musser, and E. A. Graviss.** 2001. Transmission dynamics and molecular characterization of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110*. J. Clin. Microbiol. **39:**217–221.
28. **Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam, and J. M. Musser.** 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. USA **94:**9869–9874.
29. **Supply, P., S. Lesjean, E. Savine, K. Kremer, D. van Soolingen, and C. Locht.** 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.
30. **Tsolaki, A. G., A. E. Hirsh, K. DeRiemer, J. A. Enciso, M. Z. Wong, M. Hannan, Y. O. Goguet de la Salmoniere, K. Aman, M. Kato-Maeda, and P. M. Small.** 2004. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. Proc. Natl. Acad. Sci. USA **101:**4865–4870.
31. **van Soolingen, D., A. G. van der Zanden, P. E. de Haas, G. T. Noordhoek, A. Kiers, N. A. Foudraine, F. Portaels, A. H. Kolk, K. Kremer, and J. D. van Embden.** 1998. Diagnosis of *Mycobacterium microti* infections among humans by using novel genetic marker. J. Clin. Microbiol. **36:**1840–1845.