

# Global Phylogeny of *Mycobacterium tuberculosis* Based on Single Nucleotide Polymorphism (SNP) Analysis: Insights into Tuberculosis Evolution, Phylogenetic Accuracy of Other DNA Fingerprinting Systems, and Recommendations for a Minimal Standard SNP Set†

Ingrid Filliol,<sup>1</sup>‡ Alifiya S. Motiwala,<sup>1</sup>‡ Magali Cavatore,<sup>1</sup> Weihong Qi,<sup>2</sup> Manzour Hernando Hazbón,<sup>1</sup> Miriam Bobadilla del Valle,<sup>3</sup> Janet Fyfe,<sup>4</sup> Lourdes García-García,<sup>5</sup> Nalin Rastogi,<sup>6</sup> Christophe Sola,<sup>6</sup> Thierry Zozio,<sup>6</sup> Marta Inírida Guerrero,<sup>7</sup>§ Clara Inés León,<sup>7</sup>§ Jonathan Crabtree,<sup>8</sup> Sam Angiuoli,<sup>8</sup> Kathleen D. Eisenach,<sup>9</sup> Riza Durmaz,<sup>10</sup> Moses L. Joloba,<sup>11</sup> Adrian Rendón,<sup>12</sup> José Sifuentes-Osornio,<sup>3</sup> Alfredo Ponce de León,<sup>3</sup> M. Donald Cave,<sup>9</sup> Robert Fleischmann,<sup>8</sup> Thomas S. Whittam,<sup>2</sup> and David Alland<sup>1\*</sup>

Division of Infectious Disease, Department of Medicine and the Ruy V. Lourenço Center for the Study of Emerging and Re-emerging Pathogens, New Jersey Medical School, University of Medicine and Dentistry of New Jersey, New Jersey,<sup>1</sup> National Food Safety and Toxicology Center, Michigan State University, East Lansing, Michigan,<sup>2</sup> The Institute for Genomic Research, Rockville, Maryland,<sup>3</sup> and Central Arkansas Veterans Healthcare System (CAVHS), Departments of Pathology, Microbiology-Immunology, and Neurobiology and Developmental Sciences, University of Arkansas for Medical Sciences, Little Rock, Arkansas<sup>4</sup>; Department of Infectious Diseases, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán,<sup>5</sup> and Unidad de Tuberculosis Instituto Nacional de Salud Pública,<sup>5</sup> Mexico City, and Pulmonary Services, University Hospital of Monterrey, Universidad Autónoma de Nuevo León, Nuevo León, Mexico<sup>12</sup>; Victorian Mycobacterium Reference Laboratory, Victorian Infectious Diseases Reference Laboratory, North Melbourne Victoria 3051, Australia<sup>4</sup>; Unité de la Tuberculose et des Mycobactéries, Institut Pasteur de Guadeloupe, Morne Jolivière, F-97183 Abymes-Cedex, Guadeloupe<sup>6</sup>; Grupo de Micobacterias, Subdirección de Investigación, Instituto Nacional de Salud, Bogotá, Colombia<sup>7</sup>; Department of Clinical Microbiology, Faculty of Medicine, Inonu University, Malatya, Turkey<sup>10</sup>; and Departments of Medicine and Medical Microbiology, Makerere University, Kampala, Uganda<sup>11</sup>

Received 17 August 2005/Accepted 25 October 2005

We analyzed a global collection of *Mycobacterium tuberculosis* strains using 212 single nucleotide polymorphism (SNP) markers. SNP nucleotide diversity was high (average across all SNPs, 0.19), and 96% of the SNP locus pairs were in complete linkage disequilibrium. Cluster analyses identified six deeply branching, phylogenetically distinct SNP cluster groups (SCGs) and five subgroups. The SCGs were strongly associated with the geographical origin of the *M. tuberculosis* samples and the birthplace of the human hosts. The most ancestral cluster (SCG-1) predominated in patients from the Indian subcontinent, while SCG-1 and another ancestral cluster (SCG-2) predominated in patients from East Asia, suggesting that *M. tuberculosis* first arose in the Indian subcontinent and spread worldwide through East Asia. Restricted SCG diversity and the prevalence of less ancestral SCGs in indigenous populations in Uganda and Mexico suggested a more recent introduction of *M. tuberculosis* into these regions. The East African Indian and Beijing spoligotypes were concordant with SCG-1 and SCG-2, respectively; X and Central Asian spoligotypes were also associated with one SCG or subgroup combination. Other clades had less consistent associations with SCGs. Mycobacterial interspersed repetitive unit (MIRU) analysis provided less robust phylogenetic information, and only 6 of the 12 MIRU microsatellite loci were highly differentiated between SCGs as measured by  $G_{ST}$ . Finally, an algorithm was devised to identify two minimal sets of either 45 or 6 SNPs that could be used in future investigations to enable global collaborations for studies on evolution, strain differentiation, and biological differences of *M. tuberculosis*.

Compared to many bacterial species, *Mycobacterium tuberculosis* harbors relatively little genetic diversity (21, 34, 37);

however, there is increasing evidence that the interstrain variation that exists is biologically significant. Clinical *M. tuberculosis* isolates have variable gene expression profiles (25) and have different numbers of genes deleted from their chromosome (32). In animal models, *M. tuberculosis* appears to engender a range of immune responses and variable degrees of virulence depending on the infecting strain (5, 7, 47, 55). In human infections, molecular epidemiological studies have suggested that certain *M. tuberculosis* types, identified by DNA fingerprinting, can be especially prone to drug resistance acquisition (17, 59, 65) or to global dissemination (3, 9, 27, 40, 66,

\* Corresponding author. Mailing address: Division of Infectious Disease, University of Medicine and Dentistry of New Jersey, 185 South Orange Ave., MSB A920C, Newark, NJ 07103. Phone: (973) 972-2179. Fax: (973) 972-0713. E-mail: allandda@umdnj.edu.

† Supplemental material for this article may be found at <http://jb.asm.org/>.

‡ These authors contributed equally to this work.

§ Present address: Unidad de Genética Facultad de Medicina, Universidad del Rosario, Bogotá, Colombia.

69). Some related types of *M. tuberculosis* also appear to be strongly associated with specific geographic locations (20, 22, 32). This observation may be another indication of underlying biological differences among clinical strains, including an adaptation to a specific host range or a response to variations in vaccination practices (68).

It has been difficult to directly link differences in the infecting *M. tuberculosis* bacterium to variations in the course and outcome of human tuberculosis. Clinical tuberculosis is influenced by numerous factors unrelated to the pathogen, including the infected host's genetic background, underlying diseases, immune status, diet, and social and economic environment (6, 42, 46, 71, 72). Identifying the bacterial contributions to clinical phenotypes requires a method to categorize *M. tuberculosis* isolates into groups that are likely to share most genotypic and phenotypic traits. Phylogenetic techniques facilitate such studies by organizing clinical isolates into genetically related groups and by providing an evolutionary framework for investigating polymorphisms with potential biological relevance (2). However, the appropriateness of the available phylogenetic tools has not been well characterized, and few reliable evolutionary studies have been performed in *M. tuberculosis*.

The *M. tuberculosis* genome is highly conserved, with only 1,075 single nucleotide polymorphisms (SNPs) discovered between the genomes of *M. tuberculosis* strains H37Rv and CDC1551 and only 2,437 SNPs discovered between the genomes of H37Rv and *M. bovis* strain AF2122/97 (21, 26, 50, 58, 63), making phylogenetic analysis by multilocus sequencing of housekeeping genes uninformative and impractical. Instead, *M. tuberculosis* has been genotyped by measuring genetic variation in the number of insertion elements, such as IS6110 (1, 18, 67), repetitive genetic elements in the direct repeat region (spoligotyping) (14, 36), a number of variable microsatellites (mycobacterial interspersed repetitive unit [MIRU] analysis) (24, 48, 64), and large sequence polymorphism analysis (8, 32, 49). These techniques have succeeded in identifying large groups of isolates that each appear to be related through a common ancestor. However, these methods have not been measured against a single gold standard, making it difficult to fully assess their phylogenetic informativeness. These approaches also have a common drawback in that the rate of change of each phylogenetic marker is unlikely to be uniform across all markers. The diversity of markers used can further complicate analysis (61). These limitations have made it difficult to estimate evolutionary distances among and between *M. tuberculosis* strains using current techniques.

SNPs are likely to be a more exact tool for phylogenetic studies (28, 39, 45, 51, 57). SNP-based analysis is less prone to distortion by selective pressure than genetic markers such as large sequence polymorphisms (although even synonymous SNPs may affect RNA translation rates), and SNPs are also unlikely to converge, as can be the case with spoligotype or MIRU markers (33, 57). In addition, selectively neutral SNPs should accumulate at a uniform rate and thus can be used to measure divergence (i.e., they can act as molecular clocks). Only a limited number of SNP-based studies have been performed in *M. tuberculosis* to date. The *M. tuberculosis* species was initially divided into three "major genetic" groups using a combination of two alleles at *katG463* and *gyrA95* (63). Subsequently, the complete genome sequences of two *M. tubercu-*

*losis* strains were used to identify initially 12 (21) and later 77 SNPs (2) to provide additional phylogenetic detail to a larger sample of isolates. Gutacker et al. (29) identified eight families of related isolates using SNPs identified by comparing whole genomes of the four *M. tuberculosis* complex isolates. However, these studies did not systematically identify the phylogenetic groups through model-based clustering analysis; thus, the composition and number of SNP cluster groups (SCGs) within *M. tuberculosis* has been unclear. Relationships between strains and country of origin were also not examined. Furthermore, previous studies have not taken full advantage of the power of SNP-based phylogenies to serve as a "gold standard" for examining the accuracy of the other DNA typing methods.

Here we use a combination of previously and newly identified SNPs based on whole-genome comparisons of *M. tuberculosis* strains H37Rv, CDC1551, and 210 and *M. bovis* AF2122/97 (12, 21, 26) (<http://tigrblast.tigr.org>) to investigate *M. tuberculosis* evolution and phylogeny. Our goals were to identify distinct groups of *M. tuberculosis* isolates that might have common phenotypic traits and to define the relationships between these groups. We also aimed to test for the presence of recurrent mutations and for recombination between isolates that could distort phylogenetic relationships and render individual grouping less distinct. The inferred SNP-based phylogeny was used to examine associations between genetically related groups and specific geographic regions and to reexamine the evolutionary relationships between spoligotype-defined clades and the *M. tuberculosis* species as a whole. Finally, we recommend a limited SNP set that could easily be used by international laboratories to perform SNP-based studies without repeating the large-scale SNP analysis presented here.

## MATERIALS AND METHODS

**Bacterial isolates.** A total of 323 clinical *M. tuberculosis* and *M. bovis* isolates were tested in this study. Two-hundred ninety-four *M. tuberculosis* isolates were obtained from reference laboratories or large medical centers located in the United States (Arizona, Arkansas, Colorado, and Oregon), Australia, Brazil, Colombia, Guadeloupe, India, Mexico (Monterrey, Orizaba, and Huauchinango), Turkey, and Uganda. The Australian samples included isolates from patients originating throughout Asia (Afghanistan, China, Hong Kong, India, Indonesia, Korea, Laos, Pakistan, Philippines, and Vietnam). The Arkansas, Orizaba, and Huauchinango isolates were obtained as part of community-based studies (10, 15, 38). The Uganda isolates were obtained from an urban medical center in Kampala; however, virtually all of these samples originated from patients who were indigenous to the local area. The Australian samples were obtained as part of a previous study on isoniazid (INH) resistance (31) and consisted of a diverse group of INH-resistant isolates and an equal number of pansusceptible controls. Twenty-nine clinical *M. bovis* strains were obtained from both human and animal infections isolated in the United States and Colombia. DNA from *M. tuberculosis* strains H37Rv and CDC1551 (gifts from John Belisle at Colorado State University), strain 210 (obtained from Kathleen Eisenach at the University of Arkansas for Medical Sciences), and *M. bovis* strain AF2122/97 (a gift from Stewart Cole at the Institute Pasteur) were analyzed in parallel with the clinical samples as internal assay controls.

**SNP analysis.** We examined 212 SNPs (159 synonymous SNPs [sSNPs], 35 nonsynonymous SNPs [nsSNPs], and 18 intergenic SNPs [igSNPs]) discovered through pairwise comparisons of the *M. tuberculosis* H37Rv, CDC1551, and strain 210 and *M. bovis* AF2122/97 genomes. The SNPs had either been identified previously (2, 29) or were newly identified by performing intergenome comparisons as described previously (2). SNPs were selected if they were well distributed across the *M. tuberculosis* genome and if they were unique to one of the four strains (i.e., allele A in one strain and allele B in the other three). Seventy-five SNPs were unique to H37Rv, 56 SNPs were unique to CDC1551, 38 SNPs were unique to strain 210, and 42 SNPs were unique to *M. bovis*. One SNP was mistakenly common to both *M. tuberculosis* strain 210 and *M. bovis*. The SNPs

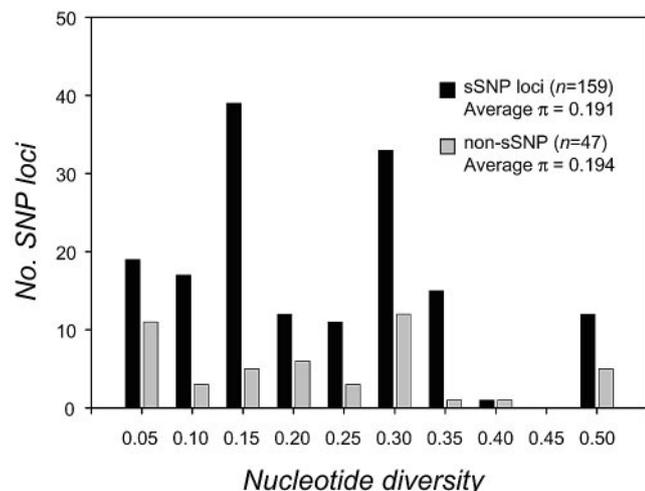


FIG. 1. Nucleotide diversity among SNPs identified through intergenomic comparisons. The distribution of nucleotide diversity for 159 sSNPs and 47 non-sSNPs among 219 strains of *M. tuberculosis* and *M. bovis* for which complete SNP data were available.

were detected in clinical strains using hairpin primer (HP) assays as described previously (30). All HP assays were also tested on *M. tuberculosis* strains H37Rv, CDC1551, and strain 210 and *M. bovis* strain AF2122/97 to confirm the presence of each allele and to verify the performance of the SNP assays. Assays on the clinical DNA samples were considered reliable only if the cycle thresholds generated in the paired wells differed by three or more cycles. Assays with fewer than three cycle differences were repeated and, if necessary, were confirmed by DNA sequencing (using the ABI Dye Terminator kit and analyzed using an ABI 3100 Genetic analyzer) or labeled as “unknown” in the final database. It was possible to assign an allele to every SNP locus in 219 of the 327 *M. tuberculosis* and *M. bovis* isolates (including the reference strains) using this approach. A small number of SNP loci had indeterminate alleles in the remaining 104 isolates. Primers for the HP assays and a complete list of the SNP alleles are included as online supplemental materials (Table S1).

**Population genetic and phylogenetic analysis.** Nucleotide diversity and linkage disequilibrium were analyzed using the computer program DNA Sequence Polymorphism (DnaSP) (56). The primary analysis was performed using the complete data for the sSNPs at 159 loci and the 219 isolates. The concatenated sSNP data were analyzed by the parsimony method with 500 bootstrap replicates, and the results were used to generate a consensus parsimony tree (43). A similar distance-based analysis was conducted by the neighbor-joining algorithm using the number of nucleotide differences (43). Model-based clustering analysis was done using STRUCTURE (54), in which isolates are assigned to clusters probabilistically, assuming Hardy-Weinberg equilibrium and linkage equilibrium within populations. The sSNP haplotypes were analyzed using the no-admixture model (each individual comes purely from one of the clusters) with 30,000 burn-in length and 100,000 replicates. Simulations were run by setting the *K* value from 1 to 10 to estimate the cluster number (*K*). A second analysis was performed including all 212 SNPs (both sSNPs and nsSNPs) and 327 isolates.

**Other phylogenetic markers.** Spoligotypes were derived and assigned to spoligotype clades (or identified as either “orphan” or “unknown” type) as described previously (20, 36, 58, 61). MIRUs were derived by PCR amplification of the 12 variable *M. tuberculosis* microsatellites and assigned an allele number based on the number of repeats as described previously (48). A dendrogram of MIRU genotypes was constructed from a distance matrix based on the proportion of allele differences between MIRU genotypes using the neighbor-joining algorithm (43).

**Determination of minimal SNP set.** A computer program called “SNPT” (single nucleotide polymorphism typing) was developed to identify the minimal number of sSNP loci required to resolve the same sSNP types (STs) as the entire SNP set. The minimal number of SNP loci was determined recursively as follows.

(i) The program identifies the number of SNP haplotypes defined in original input data on the basis of all SNP loci and then calculates the corresponding *D* value with the following equation (35):

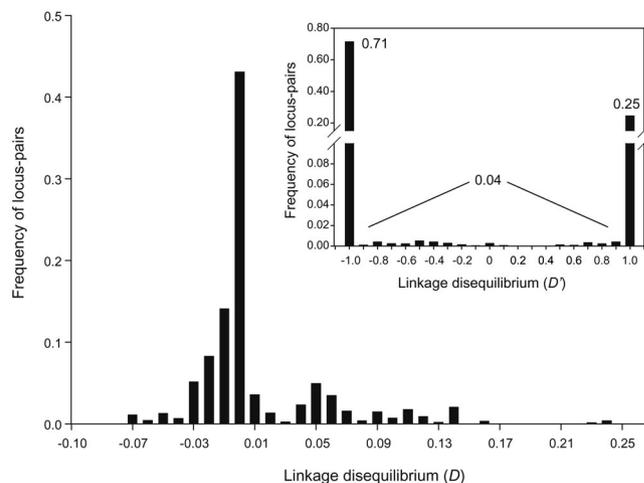


FIG. 2. Linkage disequilibrium among study loci. The distribution of the linkage disequilibrium coefficient (*D*) for 12,246 pairwise comparisons of alleles at 159 sSNP loci. A total of 5,822 (47%) of these comparisons are significant by a chi-squared test, and 3,008 (52%) remained significant using a Bonferroni correction for multiple tests. The insert shows the distribution of the standardized coefficient of linkage disequilibrium (*D'*). Ninety-six percent of the locus pairs are in complete linkage disequilibrium (i.e., *D'* = 1).

$$D = 1 - \frac{1}{N(N-1)} \sum_{j=1}^s n_j(n_j - 1)$$

where *N* is the total number of isolates in the input data set, *s* is the number of SNP haplotypes, and *n<sub>j</sub>* is the number of isolates in the *j*th SNP type. (ii) The program simulates genotyping of the input isolate population on the basis of a single locus, calculates the *D* value for the simulated typing scheme, and then identifies the loci producing the greatest *D* values by iterating through all input loci. (iii) If the highest *D* value from the previous simulated typing scheme is less than that of the original input scheme, for every locus/loci combination producing the highest simulated *D* value the program generates a new loci combination by adding one new locus, based on which genotyping of the input isolate population is simulated. This process is repeated until the simulated *D* value for a subset of SNP loci equals the *D* value of the input typing scheme.

To estimate the number of sSNPs needed to resolve the same phylogeny as the comprehensive set, simulated haplotypes containing 5 to 158 sSNP loci were generated by randomly sampling 100 times from the original ST profile of the 159 sSNPs loci. For each genotype length (number of sSNPs), cluster structure was inferred by the model-based method in STRUCTURE and the number of sSNP sets that gave the same cluster structure inferred from the comprehensive sSNP set was counted.

## RESULTS

**Levels of variation.** In the first step of the analysis, we examined 215 isolates of *M. tuberculosis* and *M. bovis* in which the nucleotides at all 159 sSNP loci were resolved. This yielded a total of 219 isolates with complete data when the four sequenced genomes were included. The frequency of the rare allele at each sSNP locus ranged from 1 to 82, with an average of 25.7 (11.8%) out of the 219 isolates. The nucleotide diversity ranged from 0.05 to 0.50 across the 159 loci with an average diversity of 0.19 (Fig. 1). This level of nucleotide diversity means that two isolates selected at random from this collection will differ at an sSNP locus 19% of the time.

The alleles at the sSNP loci are highly nonrandom in their haplotype distribution. This statistical association can be seen in the distribution of the linkage disequilibrium coefficient (*D*)

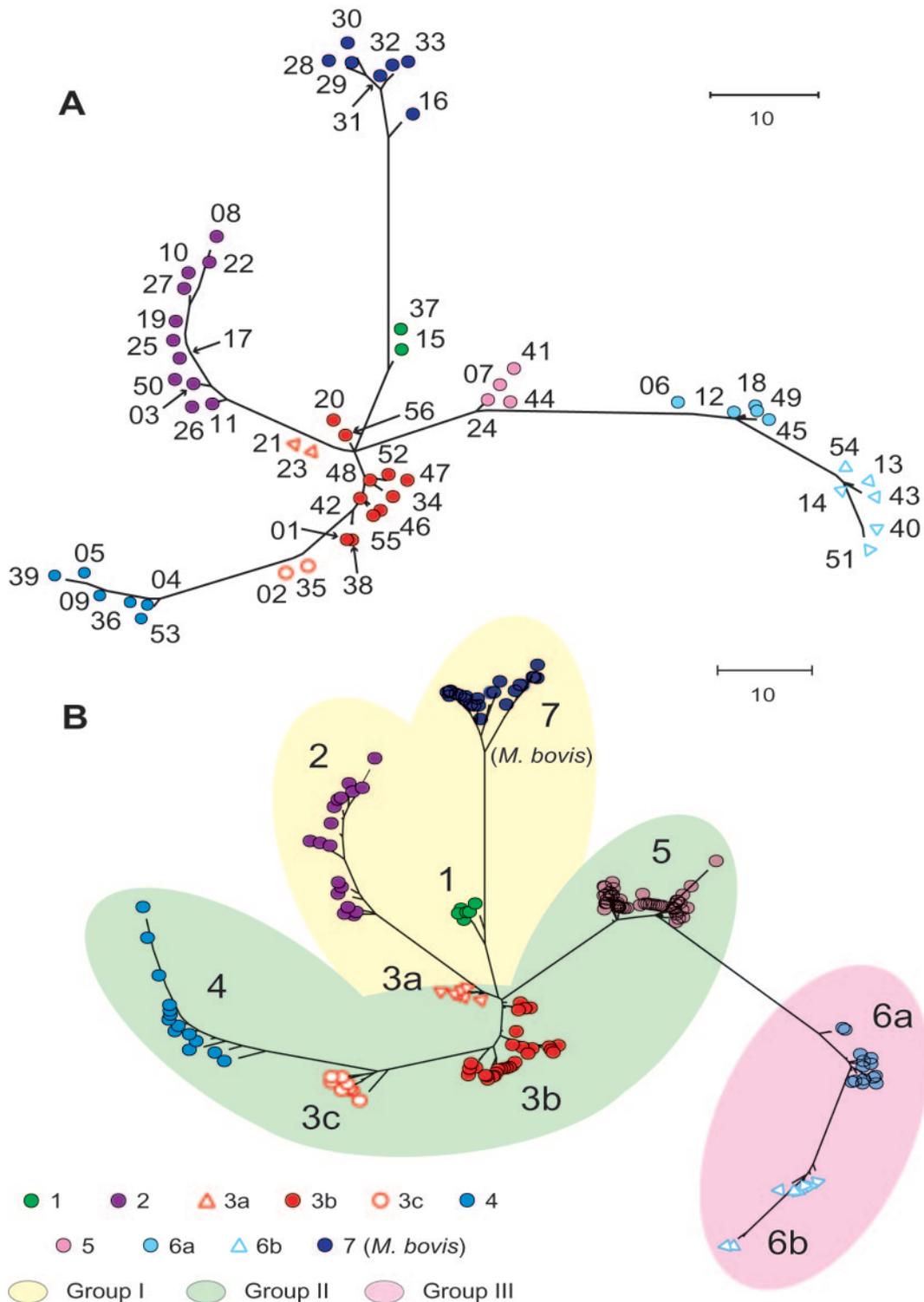


FIG. 3. Phylogenetic trees of a global collection of *M. tuberculosis* isolates. A. Distance-based neighbor joining tree of *M. tuberculosis* and *M. bovis* based on 159 sSNPs resolved the 219 isolates into 56 sequence types (STs). Each ST is indicated by a dot color coded for membership in an SCG and a numerical value. The bootstrap values associated with SCG-1 and SCG-2 to SCG-7 are 84% and 99%, respectively. Reference strains *M. tuberculosis* 210, *M. tuberculosis* CDC1551, *M. tuberculosis* H37Rv, and *M. bovis* AF2122/97 are present in SCG-2, -4, -6b, and -7, respectively. The modeling-based clustering analysis of the same comprehensive set of the genotype data (159 sSNPs in 219 isolates) identified the same seven clusters. B. Model-based neighbor-joining tree based on a comprehensive data set of 212 SNPs resolved the 327 isolates into 182 STs and presented identical clusters as shown in panel A. Numbers designate each SCG; SC subgroups are indicated by colored symbols. The SNP lineages that belong to the three “major genetic” groups based on combination of two alleles at *katG463* and *gyrA95* are also highlighted. The scale bar indicates the number of SNP differences.

for 12,246 pairwise comparisons of alleles at 159 sSNP loci (Fig. 2). A total of 5,822 (47%) of these comparisons are significant by a chi-squared test, and 3,008 (52%) are significant using a highly conservative Bonferroni correction for multiple tests (56). The standardized coefficient of linkage disequilibrium ( $D'$ ) is strongly U-shaped and shows that 96% of the locus pairs are in complete linkage disequilibrium (i.e.,  $D' = 1$ ), that is, there are at most three out of the four possible haplotypes for most locus pairs. The fact that only 4% of the locus pairs have  $-1 < D' < 1$  suggests that recurrent mutation and recombination have played only a minor role in generating haplotype diversity.

**Phylogenetic analysis of clinical isolates.** The 159 sSNPs resolved 212 isolates into 56 haplotypes or STs (Fig. 3A). Both the parsimony analysis and distance-based neighbor-joining method grouped these strains into seven SNP cluster groups (SCGs) (bootstrap values,  $>80\%$ ) comprised of six distinct *M. tuberculosis* SCGs and a seventh group that contained all the *M. bovis* isolates but no *M. tuberculosis* isolates. Five subgroups were also identified. The model-based clustering recovered the same seven SCG groups (Fig. 4). All known *M. bovis* strains were placed in the *M. bovis* group by both clustering methods. We next performed a second analysis using all 323 *M. tuberculosis* and *M. bovis* study isolates and all 212 SNP loci (including sSNPs, nsSNPs, and igSNPs). This resolved the isolates into 182 SNP types and the same seven SCGs and five subgroups that were identified using only the sSNPs, although the subgroups became more apparent with the increased number of SNP types (Fig. 3B). All but four nsSNPs and igSNPs were phylogenetically informative. However, all isolates that were assigned to a particular group using sSNPs were assigned to the same group using all SNP markers without exception.

Our results showed that most clusters were separated from each other by multiple mutations. It was particularly striking that we did not detect any *M. tuberculosis* isolates that were situated intermediately between cluster types. This distribution is consistent with past subdivision of the *M. tuberculosis* species through geographic barriers or evolutionary bottlenecks as has been suggested by others (2, 4, 23, 63), and it demonstrates the degree to which *M. tuberculosis* appears to be diverging into lineages containing distinct subspecies.

Previous studies based on frequency and distribution of chromosomal deletions have convincingly argued that *M. bovis* and *M. tuberculosis* share a recent common ancestor (8). Using this information, we rooted the total SNP tree with *M. bovis* haplotypes to infer an evolutionary timeline for the common ancestors of each SCG. SCG-1 appears to have diverged early after the most recent common ancestor with *M. bovis*; this was followed, in order, by SCG-3abc, then SCG-2 and SCG-5, and then by SCG-4 and SCG-6ab. These results largely agree with the assignment of isolates to one of the three major genetic groups defined by *katG463* and *gyrA95* SNP polymorphisms (63). All major genetic group I isolates fell into SCG-1, -3a, and -2; all major genetic group II isolates fell into SCG-3b, -3c, -4, and -5; and all major genetic group III isolates fell into SCG-6a and -6b (Fig. 3B). Thus, none of the major genetic group assignments contradicted the phylogeny elucidated by the SNP trees, although SCG-3abc can be viewed as “transitional” SCGs where the major genetic groups diverged.

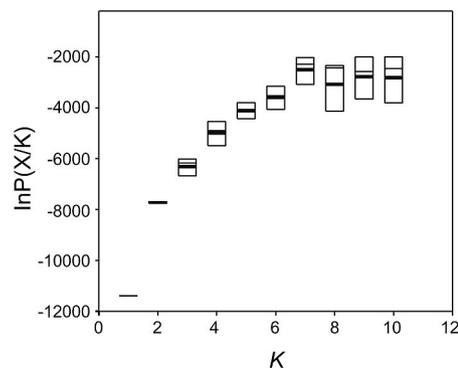


FIG. 4. Inferring the number of clusters in an *M. tuberculosis* phylogeny. The number of clusters ( $K$ ) were inferred for the genotype data of 219 *M. tuberculosis* and *M. bovis* strains on the basis of 159 sSNPs using the model-based clustering method implemented in STRUCTURE. Five independent simulations were run for  $K$  values ranging from 1 to 10. The top and bottom boundaries of each box indicate the 25th percentile and the 75th percentile, respectively. The thick black line and the thin black line within the box mark the median and the mean, respectively. The estimate of the posterior probability of  $K$ ,  $P(X|K)$ , was the highest when  $K = 7$ .

**Geographic distribution of SNP clusters.** Several recent investigations have suggested that, on a global scale, the geographic origin of the human host has a strong influence on the type of infecting *M. tuberculosis* strain (32). To address this hypothesis we examined the distribution of SCGs by the country where each *M. tuberculosis* isolate was identified (Fig. 5A). Only countries contributing 40 or more isolates were included in this analysis. Each country did indeed appear to have one or two predominant SCGs, with the exception of the United States and Guadeloupe. United States and Guadeloupe isolates appeared to have relatively equal numbers of isolates from most SCGs. These results are consistent with the epidemiology of tuberculosis in the United States where a large proportion of tuberculosis cases occur among the foreign born (11); Guadeloupe is an island of the lesser Antilles of a mixed African, European, and Indian descent and where today's tuberculosis epidemiology is also partly driven by the foreign born. Australia, which also contains large immigrant populations, appeared to have two predominant SCGs. Subdividing the Australian isolates by the country of birth for each human host permitted resolution of the samples into isolates from the Indian subcontinent, which were predominantly SCG-1, and isolates from East Asia, which were predominantly SCG-2 followed by SCG-1 (Fig. 5B). The ancestral placement of SCG-1 suggests that *M. tuberculosis* originated in the Indian subcontinent.

Our study also included samples from one indigenous community in Uganda and one indigenous community in Mexico (Huauchinango) that were unlikely to have a large number of imported tuberculosis cases. A second Mexican community in Orizaba is predominately mestizo in origin; however, as with Huauchinango, this community has not experienced any recent migration to or from the region. The Ugandan samples showed the largest predominance of a single SCG (SCG-5). However, this predominance is unlikely to represent recent epidemic spread of a single strain, because we found that there was significant diversity among the Ugandan SCG-5 isolates as

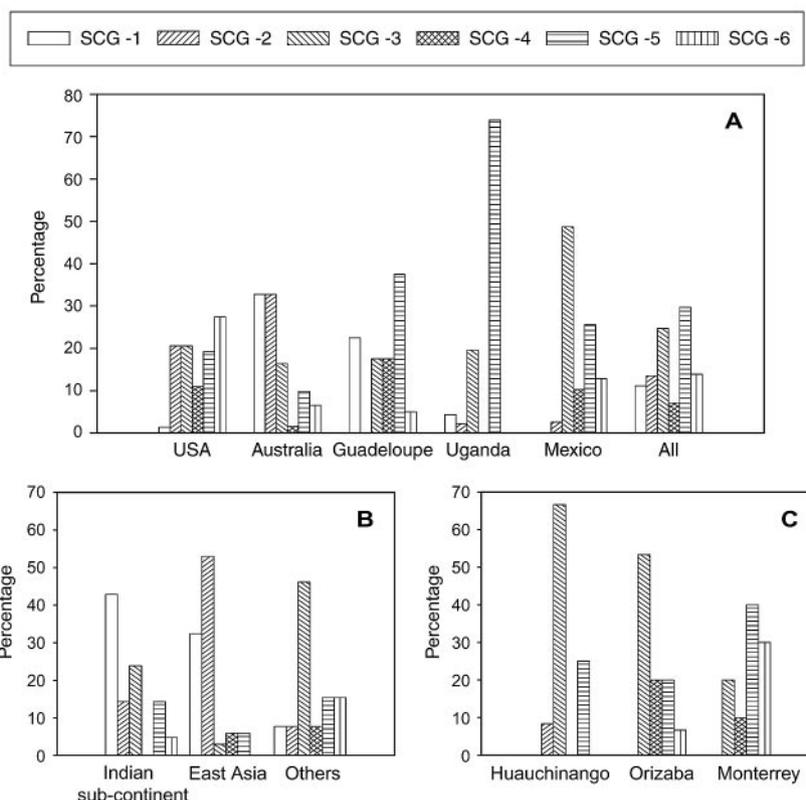


FIG. 5. Distribution of the SCGs by country of origin. The proportion of *M. tuberculosis* isolates with a particular SCG is shown for various geographic regions. A. The country where the isolate was cultured. B. The place of birth of the human host from which the isolate was cultured for isolates obtained from Australia. C. The place of birth of the human host from which the isolate was cultured for isolates obtained from Mexico. "Indian subcontinent" includes *M. tuberculosis* isolates from India, Pakistan, and Afghanistan and seven additional *M. tuberculosis* isolates obtained independently from India. "East Asia" includes *M. tuberculosis* isolates from Vietnam, China, Philippines, Indonesia, Korea, Hong Kong, and Laos. "Others" includes *M. tuberculosis* isolates from Australia, Europe, and Africa. Huauchinango and Orizaba represent relatively closed populations, while Monterrey represents a more urban population. Two Mexican isolates for which the origin of host was unknown were not included.

determined by SNP and MIRU type diversity. The Mexican isolates were obtained from three different geographic locations. *M. tuberculosis* isolates from communities in the adjoining districts of Orizaba and Huauchinango predominantly belonged to SCG-3, while isolates from a more urban Monterrey population were predominantly SCG-5 and SCG-6 (Fig. 5C). Interestingly, substantial numbers of SCG-6 isolates were only detected in the United States and Monterrey. Monterrey is located approximately 112 miles south of the United States borders; this geographic pattern suggests cross-border spread of SCG-6. The predominance of different SCGs in the Mexican regions demonstrates that different SCGs can achieve high frequencies even within a single country. In the case of Mexico, it suggests that *M. tuberculosis* has been introduced several times into the population from different geographic sources.

**Phylogenetic composition and relationships of spoligotype-defined clades.** Worldwide collections of *M. tuberculosis* have been classified into approximately nine different clades using the spoligotyping system (20, 58). Thus, spoligotyping has also been used to formulate hypotheses about the global evolution and spread of *M. tuberculosis* and to identify groups of isolates that may have common virulence or other biological attributes. However, multiple and repeated genetic events can result in indistinguishable spoligotypes, which can obscure phylogenetic

signals. Indeed, genetic convergence of spoligotypes already has been demonstrated (70). To assess the robustness of the spoligotype typing system, we compared the SCG and spoligotype clade assignments for the 246 isolates for which spoligotypes were available (Fig. 6). The results show that some spoligotype clades were phylogenetically accurate while others were less so. In particular, the spoligotype-defined Beijing clade, a group overrepresented in drug-resistant isolates and which appears to have unique virulence properties (5, 16, 47, 55), was exclusively present in SCG-2 (i.e., no other clade was assigned to this SCG, and all of the spoligotyped isolates within this SCG were members of the Beijing clade). The SNP tree also demonstrates the relatively large genetic distance among Beijing isolates, suggesting that this group is relatively diverse despite their common membership in a single SCG and clade. Furthermore, although the Beijing clade is often thought ancestral (belonging to major genetic group I), SCG-1 and -3 appear to have more ancestral roots than Beijing, and SCG-5 appears to be equidistant with the Beijing clade from SCG-1 (even though all SCG-5 isolates were classified into major genetic group II). Thus, the SNP tree both confirms the spoligotype Beijing clade assignment and clarifies the evolutionary relationships between this clade and other members of the *M. tuberculosis* species.

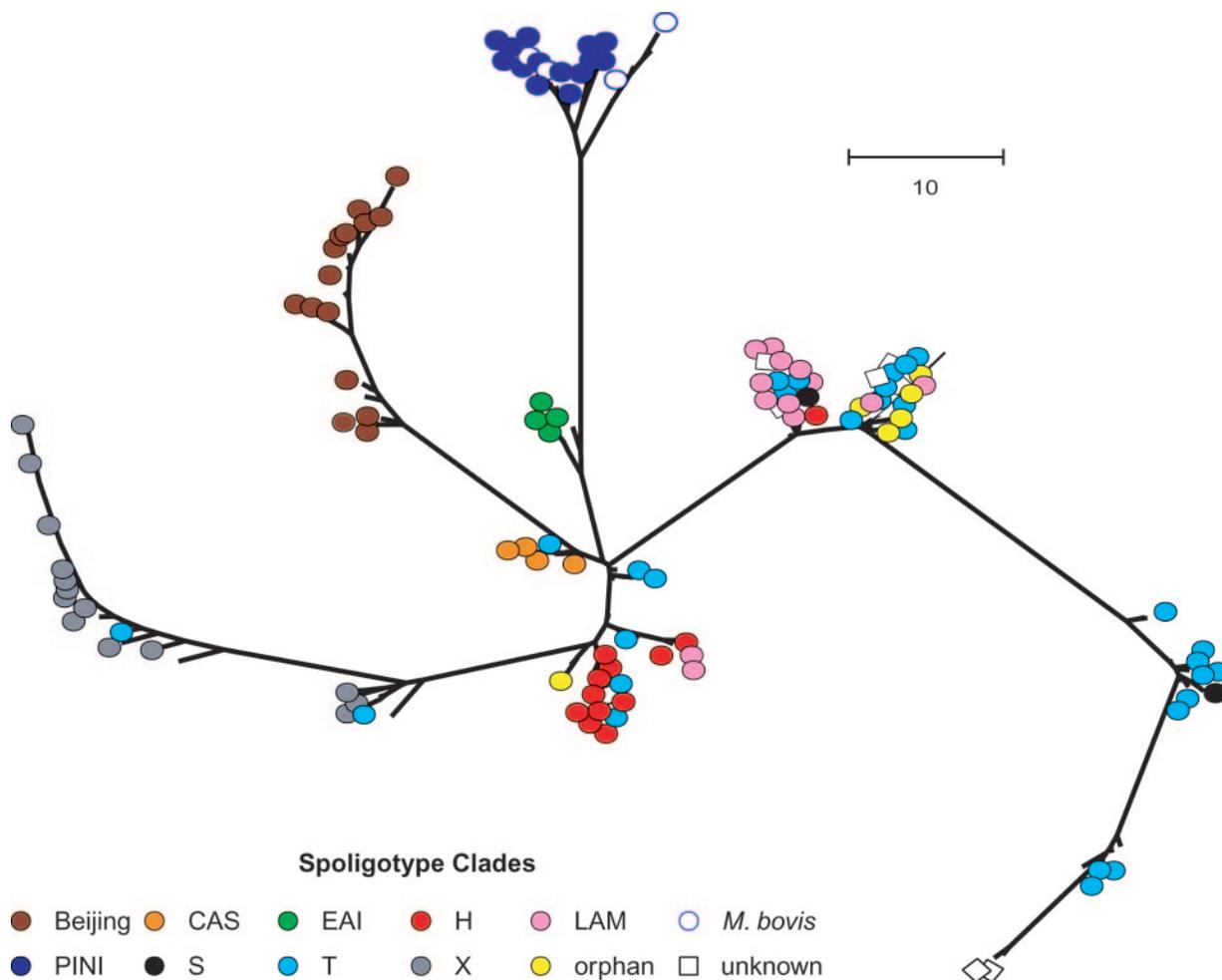


FIG. 6. Distribution of the spoligotype clades on the SNP-based phylogeny. Each isolate is indicated by a dot, which is color coded according to the spoligotype clade assignment. CAS, Central Asian clade; EAI, East African-Indian clade; H, Haarlem clade; LAM, Latin American and Mediterranean clade; PINI, *Mycobacterium pinnipedii* clade; S, S clade; T, T clade; X, X clade.

Several other spoligotype clade assignments are concordant with the SNP tree. The East African-Indian (EAI) clade was uniquely present in SCG-1, and the Central Asian (CAS) clade was confined to SCG-3a (although not uniquely present), suggesting that the EAI clade has the most ancestral roots of all clades, followed by CAS. The PINI clade (a spoligotype highly similar to a classic *M. bovis* pattern but initially identified in a new species of mycobacteria [13]) was contained within SCG-7 (*M. bovis*), and the X clade was well associated with SCG-3c and -4. Other clades associated less well with either individual or adjoining SCGs and SC subgroups. For example, the T superfamily is an ill-defined group of isolates with similar spoligotype patterns (58). Our results showed that the T clade comprised the majority of SCG-6a and -6b, but it could also be found in SCG-3abc, -4, and -5, suggesting that the T clade classification in fact includes a number of phylogenetically distinct subgroups. The Haarlem (H) clade is another spoligotype group that occurs at high frequency in human tuberculosis studies (20, 44). Our results showed that the H clade could be found in two distinct genetic clusters, SCG-3b and SCG-5, although neither of these SCGs contained only H clade isolates. The Latin American and Mediterranean (62) clade iso-

lates were also associated with SCG-3b and SCG-5, suggesting that some H and LAM clade isolates are more closely related to each other than to other isolates within their respective clades. The relatively high concentration of “orphan” and “unknown” spoligotypes in SCG-5 points to other isolates that are also likely to have common ancestry with the LAM and possibly the H clades.

**Evolutionary dynamics of the principal *M. tuberculosis* microsatellites.** The *M. tuberculosis* genome contains a number of polymorphic microsatellites (i.e., short tandem repeats) that can be used to generate highly diverse DNA fingerprints in a technique called MIRU typing (48). As with spoligotyping, variation at these loci mimics a stepwise mutation model so that the extent to which this variation retains a phylogenetic signal is not known. To address this, we MIRU-typed 263 study isolates, including all of the isolates that had been spoligotyped, producing 183 MIRU types (MTs). We then examined the distribution of SCGs onto a MIRU tree (Fig. 7). We plotted the distribution of SCGs onto the MIRU tree instead of plotting the MIRU groups onto the SNP cluster tree (as was the case for the spoligotyping analysis), because MIRU groups have not yet been rigorously defined in a fashion similar to that

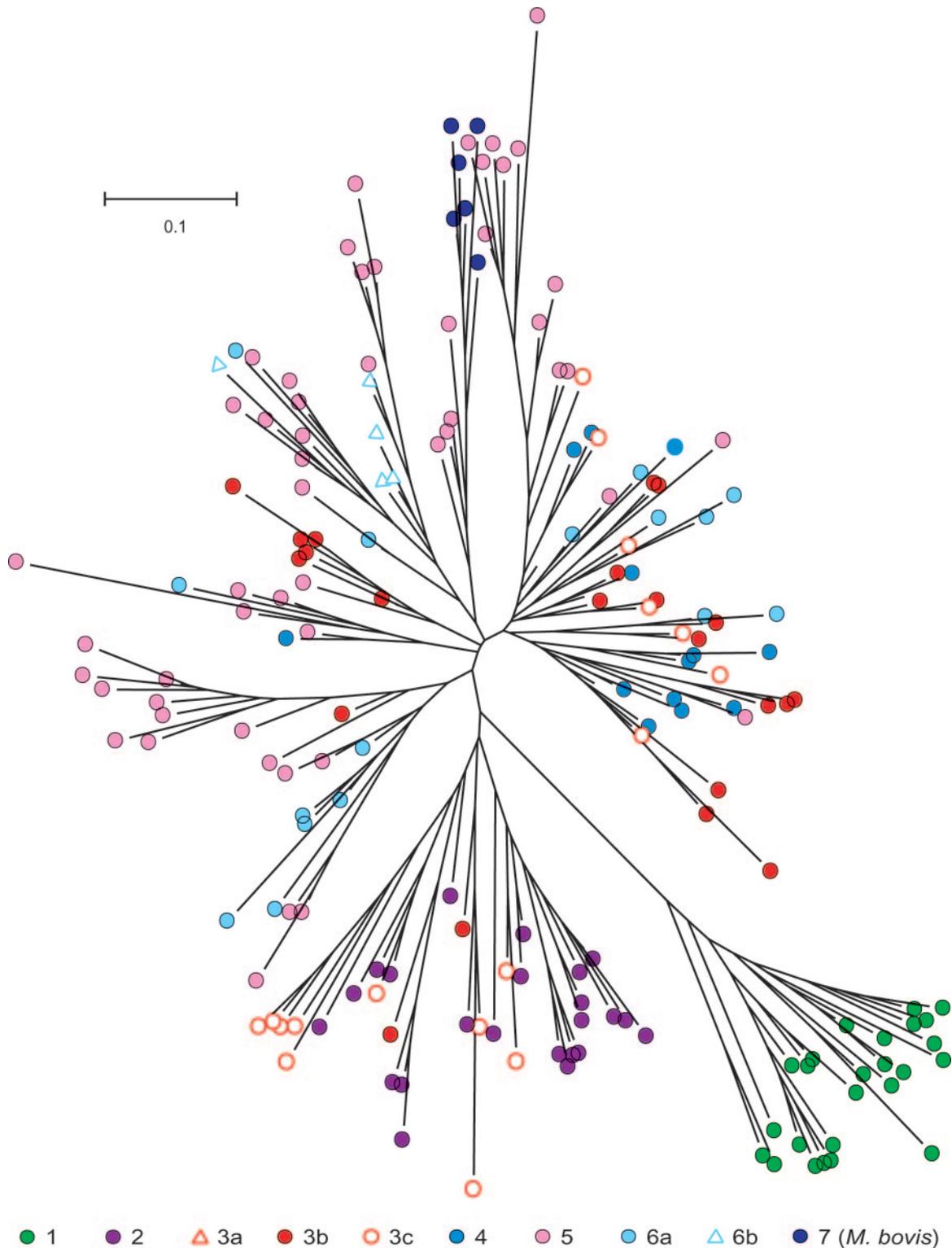


FIG. 7. Distribution of SCGs on the distance-based neighbor-joining tree representing the MIRU phylogeny. SCG assignment (SCG-1 through SCG-7) of each MIRU type is based on the same colors and symbols used in Fig. 3B. The location(s) of the SCG-3a isolates (red triangles) are not shown, because MIRU results were not available for this group.

for spoligotype clades. As with spoligotyping, MIRU typing was completely concordant with SNP analysis for *M. bovis* isolates and for *M. tuberculosis* isolates belonging to SCG-1 (EAI clade). MIRU typing also had relatively good concor-

dance with SNP analysis for SCG-2 (Beijing clade) isolates, although SCG-3a and -3b isolates were also found intermingled with SCG-2 isolates on the same MIRU branch. Other SCGs were generally more mixed on the branches of the

TABLE 1. Measures of diversity and genetic differentiation of mycobacterial interspersed repetitive unit (MIRU) loci

MIRU loci	No. of alleles	$H_S^a$	$H_T^b$	$G_{ST}^c$
2	4	0.056	0.058	0.046
4	7	0.223	0.427	0.478
10	8	0.428	0.751	0.431
16	6	0.435	0.508	0.143
20	3	0.071	0.077	0.080
23	7	0.242	0.581	0.583
24	3	0.053	0.317	0.832
26	8	0.544	0.758	0.282
27	5	0.270	0.265	0.000
31	6	0.338	0.572	0.409
39	6	0.201	0.456	0.559
40	7	0.548	0.779	0.297
Avg	5.8	0.284	0.462	0.386

<sup>a</sup>  $H_S$  is the within population diversity.

<sup>b</sup>  $H_T$  is the diversity in the total population.

<sup>c</sup>  $G_{ST}$  is Nei's coefficient of genetic differentiation.

MIRU tree, although some SCGs appeared to be concentrated in specific regions. These results indicate that MIRU typing, in its current 12-allele format, is a poor tool for phylogenetic analysis despite the extensive pattern diversity that is produced by this method.

**Informativeness and diversity of individual MIRU microsatellites.** Two variables affect the ability of microsatellite loci to identify specific SCGs: (i) the allelic diversity within an SCG, expressed as the coefficient of diversity  $H_S$ , and (ii) the allelic diversity between SCGs, expressed as the difference between total diversity ( $H_T$ ) and  $H_S$ . Microsatellites with divergent allele frequencies across SCGs will have low  $H_S$  values and high  $H_T$  values (where the theoretically ideal microsatellite would have  $H_S = 0$  and  $H_T =$  the maximum possible value). Conversely, loci with high  $H_S$  values will harbor most MIRU allelic diversity within SCGs and thus provide little useful phylogenetic information but may generate a high degree of diversity, which can be useful for distinguishing between closely related isolates.  $H_S$  and  $H_T$  can also be combined to derive a coefficient of differentiation ( $G_{ST}$ ) (52) which measures the overall ability of a microsatellite to differentiate among SCGs and ranges from 0 (no subpopulation differentiation) to 1.0 (complete differentiation) (52).

We examined the distribution of alleles within and between

each SCG for each of the microsatellite loci used in MIRU typing. The allele frequencies and single-locus diversities were calculated for each cluster, and then  $H_S$ ,  $H_T$ , and  $G_{ST}$  were derived (Table 1). Using a  $G_{ST}$  of 0.4 as a cutoff, the results show that MIRU microsatellite loci 4, 10, 23, 24, 31, and 39 (especially loci 23, 24, and 39) with high  $G_{ST}$  values were likely to be useful phylogenetic markers. Conversely, the loci 2, 16, 20, and 27 did not appear to contribute significant phylogenetic information. Other microsatellite loci had intermediate values. Further examination of  $H_T$  and  $H_S$  indicated that loci 16 and 27 were likely to provide substantial diversity to MIRU patterns, while loci 2 and 20 provided little phylogenetic information or diversity even though all four of these loci had similar low  $G_{ST}$  values.

**Diversity within SNP cluster groups.** Phylogenetic trees constructed with previously identified SNP markers can be biased by branch collapse (2). Branch collapse hides strain diversity that would normally be observed if an SNP tree was constructed with multilocus sequence typing results. The wide range of DNA patterns generated by the MIRU typing permitted us to analyze each SCG for genetic diversity that might be hidden within collapsed branches (Table 2). The results show that SCG-3a and SCG-5 contained more sequence diversity than the other clusters. Notably, SCG-1 showed sequence diversity that was equal to that of most of the other SCGs, even though it had a limited number of SNP types assigned to it. The *M. bovis* isolates showed diversity that was similar to one of the *M. tuberculosis* clusters, even though these isolates represent a geographically diverse collection of an entire species. This suggests that *M. bovis* is even more clonal than the *M. tuberculosis* species or that microsatellite evolution occurs at a slower pace in *M. bovis*.

**Determination of the minimal SNP set.** SNP analysis is relatively expensive, and it would be advantageous to define a minimal SNP set that could robustly identify all known SNP types or at least all known SCGs. The 219 isolates used to generate the initial SNP tree (Fig. 3A) were analyzed using SNPT to determine a minimal SNP set that could be used to identify all 56 STs. Phylogenetic simulations of haplotypes containing 5 to 158 sSNP loci were performed using STRUCTURE to determine a minimal SNP set that could be used to identify all seven SCGs. This work identified a minimal set of 45 sSNPs that could identify all 56 STs (Table 3) and a minimal set of 16 sSNPs that could be used to group all 219 isolates into

TABLE 2. Estimates of genetic diversity within SNP cluster groups (SCGs) using mycobacterial interspersed repetitive unit (MIRU) typing

SCG	No. of MTs	Polymorphic loci	Mean no. of alleles	MT <sup>a</sup>		Isolate	
				$H$	SE ( $H$ )	$H$	SE ( $H$ )
1	25	0.67	2.75	0.300	0.083	0.281	0.082
2	23	0.83	2.58	0.269	0.057	0.215	0.047
3a	10	0.92	2.92	0.431	0.081	0.431	0.081
3b	24	0.75	2.58	0.281	0.075	0.252	0.062
3c	9	0.42	2.08	0.229	0.090	0.212	0.086
4	17	0.58	2.25	0.218	0.080	0.207	0.078
5	52	0.92	3.58	0.427	0.072	0.414	0.071
6a	17	0.75	2.58	0.291	0.075	0.285	0.075
6b	5	0.42	1.42	0.183	0.067	0.158	0.060
7 ( <i>M. bovis</i> )	6	0.50	1.58	0.211	0.069	0.060	0.022

<sup>a</sup> Mean average genetic diversity ( $H$ ) and standard errors (SE) of single locus diversity within each population across the MIRU types (MTs) and isolates.

TABLE 3. A minimal set of 45 sSNPs that can be used to resolve *M. tuberculosis* and *M. bovis* isolates into the 56 sSNP types and 10 SNP cluster-subgroups<sup>a</sup>

IS	SCG	SNP position and allele in H37Rv
1	3b	4280708 G
2	3c	4280708 G
3	2	4255919 G
4	4	4255919 G
5	4	4254003 G
6	9a	4137829 C
7	5	4119243 L
8	2	4024270 C
9	4	3783054 A
10	4	3544707 C
11	2	3455683 C
12	9a	3450722 C
13	9b	3440539 A
14	9c	3440539 A
15	1	3440461 G
16	2	3438383 G
17	2	3207247 A
18	2	2990037 C
19	2	2891265 L
20	3a	2825759 L
21	3a	2627946 A
22	3a	2532614 G
23	3a	2462869 A
24	3a	2376133 A
25	3a	2223680 G
26	3a	2158580 G
27	3a	1952599 C
28	3a	1892015 L
29	3a	1884695 A
30	3a	1692683 A
31	3a	1692067 A
32	3a	1548147 G
33	3a	1477596 C
34	3a	1294396 C
35	3a	1191859 L
36	3a	1163132 C
37	3a	1068149 C
38	3a	949219 C
39	3a	923063 L
40	3a	918314 L
41	3a	909164 C
42	3a	797736 C
43	3a	519804 G
44	3a	311611 L
45	3a	220048 C
46	3a	92197 G
47	3a	43943 A
48	3a	37031 G
49	3a	37031 G
50	3a	37031 G
51	3a	37031 G
52	3a	37031 G
53	3a	37031 G
54	3a	37031 G
55	3a	37031 G
56	3a	37031 G

<sup>a</sup> ST, sequence type; SCG, SNP cluster group and subgroup.

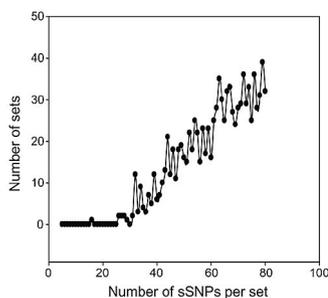


FIG. 8. Determining the size of the minimal SNP set. Graph representing the number of sSNP sets that give the same cluster structure as that inferred from the comprehensive set of the 159 sSNPs versus the number of sSNP loci per set (ranging from 5 to 80). For each genotype length, sSNPs were generated by 100 random samples from the original source of 159 sSNPs. Cluster structure was inferred by the model-based method in STRUCTURE. In this simulation study, the minimum number of sSNP loci needed to resolve the same clusters was 16.

the same six SGCs (plus an additional *M. bovis* cluster) as the comprehensive set of 159 sSNPs (Fig. 8, Table 4). Visual inspection of the 16 sSNP set permitted us to further reduce this number to a minimal subset of 6 SNPs that could be used to classify all isolates into all SCGs (Table 4). These two limited SNP sets should make it possible to perform significantly larger studies which will further define the evolution and global distribution of the *M. tuberculosis* species.

DISCUSSION

Our SNP-based phylogenetic analysis of a global collection of *M. tuberculosis* isolates suggests that this species can be divided into six phylogenetically distinct groups (with a seventh group containing all *M. bovis* isolates), and that two of these SCGs can be further subdivided into five subgroups, resulting in a combination of nine SCGs and subgroups plus an *M. bovis* SCG. These results differ somewhat from an earlier study, by Gutacker et al. (29), that identified eight major groups of *M. tuberculosis* without subgroups as well as a three-branch tree rather than the four-branch tree that we obtained in our analysis. This discrepancy may be explained by our use of a model-based clustering analysis using the program STRUCTURE (54) to identify and assign SCGs. In contrast, Gutacker et al. appear to have made visual SCG assignments based on a phy-

logenetic tree constructed using MEGA software. Our study also differed in the use of SNPs that were unique to one of the four sequenced strains, while Gutacker et al. may have included SNPs shared by two of the four strains. This fact could accentuate the degree of divergence that we found. Interestingly, both studies demonstrate that all of the *M. tuberculosis* SCGs are distinct and deeply branching. Thus, the *M. tuberculosis* species appears to consist of very distinct strain “families.” We propose that these families (or SCGs) are the ideal units to investigate biological variations among clinical isolates.

Our results provide additional insights into prior observations that *M. tuberculosis* strains have stable associations with their human host populations (32). We found that the predominant SCG varied depending on the country where *M. tuberculosis* isolates were collected and, in the case of Australia, with the birth country of the patient. Isolates from countries with large immigrant populations, such as the United States, represented a mixture of most SCGs. This result is consistent with observations that United States tuberculosis is due to a mixture of strains imported from around the world (11). The isolates from Uganda and the provinces of Orizaba and Huachuquingo in Mexico were obtained predominantly from indigenous or relatively closed populations, and the distribution of SCGs in these populations is particularly noteworthy. These isolates had the least cluster group diversity compared to isolates from other countries, suggesting that tuberculosis began in each of these regions with the introduction of a single, founding cluster group. The predominance of a different SCG in Monterrey, Mexico, suggests that tuberculosis was introduced separately into this region, where it became the most commonly observed SCG. Isolates originating from the Indian subcontinent predominantly belonged to the most ancestral cluster SCG-1, and isolates from East Asia predominately belonged to SCG-1 and SCG-2. These results suggest that the evolutionary radiation of human tuberculosis began on the Indian subcontinent, spreading to East Asia and then elsewhere.

The availability of a “gold standard” phylogeny makes it possible to evaluate the informativeness of other DNA fingerprinting systems for phylogenetic analysis. Our results support the major genetic group phylogeny assigned by *katG463* and *gyrA95* gene polymorphisms (63). However, the SNP tree revealed much more phylogenetic detail, including the group where major phylogenetic groups I and II diverge. This is not

Downloaded from <http://jtb.asm.org/> on December 4, 2018 by guest

TABLE 4. Sixteen sSNPs identified by STRUCTURE and a minimal subset of six sSNPs that can be used to resolve *M. tuberculosis* and *M. bovis* isolates into the seven SNP cluster groups (including the *M. bovis* cluster)

SCG	SNP position and allele in H37Rv <sup>d</sup>															
	192197	473687	519804	1830293	1920118	2156023	2361602	2376133	2460626	2598398	3057134	3111473	3352929	3440539	3550786	4314642
1	G	G	g	g	T	g	g	a	C	g	c	C	G	g	c	g
2	G	G	g	g	G	g	g	a/g <sup>a</sup>	C	g	c	T	G	a	g	g
3	G	G	g	g	G	g	g	a	C	g	c	C	G	a	c	g
4	G	G	g	g	G	g	g	a	A	g	t	C	G	a	c	g
5	G	G	g	g	G	g	g	a	C	g	c	C	C	a	c	a
6	T	G	g	g	G	g	c	a	C	a/g <sup>b</sup>	c	C	C	a	c	a
7	G	A	a/g <sup>c</sup>	a	T	t	g	a	C	g	c	C	G	g	c	g

<sup>a</sup> Either allele defines SCG-2.  
<sup>b</sup> Either allele defines SCG-6.  
<sup>c</sup> Either allele defines SCG-7.  
<sup>d</sup> Capital letters indicate the minimal six-SNP subset.

surprising given that both systems are based on SNP markers. The SNP tree also revealed a number of strong concordances with spoligotyping. EAI and Beijing clades were strongly associated with SCG-1 and -2, respectively. Moreover, our results conclusively show that EAI clade isolates are the most ancestral, as has been suggested by previous observations that the TbD1 deletion is absent in EAI isolates (8, 19). CAS, PINI, and X clades also correspond relatively well to a small number of SCGs, although intermixing with other SCGs is apparent. Other clades appeared to be less successful at grouping isolates into genetically related clusters. It is possible that a more detailed comparison of spoligotype patterns and the SNP tree might result in more accurate clade assignments. MIRU typing appeared to be the least successful at accurately assigning isolates to genetically related groups. A MIRU tree correctly classified all SCG-1 isolates together, and other MIRU branches showed some concordance with the SNP tree. However, no other MIRU branch corresponded either to a single SCG or to a logical combination of neighboring SCGs. These results are supported by our analysis of individual MIRU microsatellites, which demonstrated that a number of loci are not phylogenetically informative. MIRU remains an attractive method to distinguish among potentially related isolates, because this typing system generates a large and diverse set of DNA fingerprints. Phylogenetic analysis using this system could potentially be improved using a staged approach. We suggest that MIRU typing be modified so that an initial analysis classifies *M. tuberculosis* isolates into clusters using the most phylogenetically informative microsatellites. This process could be followed by secondary, and perhaps tertiary, classification using the microsatellites that were less phylogenetically informative but that provided fingerprint diversity.

There are a few potential limitations to this study. The phylogenetic analysis was performed using SNP markers that were obtained by comparing only four sequenced *M. tuberculosis* genomes. Trees generated using this type of phylogenetic marker may be distorted by branch collapse, which may cause distinct lineages to become combined into a single cluster group (2). However, it has been demonstrated that branch collapse will not alter the position of each ancestral branch point on the overall tree, so these results remain highly informative (2, 53). Furthermore, our analysis of MIRU diversity demonstrated that most SCGs on the tree were equally diverse, with the exceptions of SCG-3a and SCG-5, which had only moderately increased diversity. This suggests that no SCG contains an excessive number of collapsed branches, although SCG-3a and SCG-5 are worthy of further study. Future genomic sequencing studies will be required to identify additional SNP markers to expand the branches that are collapsed. Although the *M. tuberculosis* isolates analyzed in this study were obtained from diverse locations around the world, the number of geographic locations and the number of isolates sampled from each location remain limited. However, several of our conclusions are supported by previous studies. For example, our observation that SCG-1 isolates predominated in the Indian subcontinent isolates and that SCG-2 isolates predominated in East Asia is supported by similar observations based on the geographic distribution of EAI, CAS, and Beijing clades (41, 60; K. Brudey and N. Rastogi, unpublished observations). However, a much larger study will be required to

firmly delineate the origin and global spread of *M. tuberculosis* and to definitively identify all *M. tuberculosis* SCGs. Our identification of a smaller number of highly informative SNPs that can replace the 212 SNPs used in this study should make such investigations possible. Investigations of this worldwide disease will be best performed through collaborations between many different centers internationally. A single basic SNP set will need to be agreed upon for this work to progress. We suggest that the minimal SNP set presented here should comprise the core markers for these future studies.

#### ACKNOWLEDGMENTS

We thank Karine Brudey for her assistance with the spoligotyping experiments. We also thank Helen Billman-Jacobe for *M. tuberculosis* isolates from Australia and Patricia del Portillo and Juan Germán Rodríguez for the *M. bovis* strains from Colombia.

This work was supported by Public Health Service grants AI-46669 and AI-49352 from the National Institutes of Health. This work was also supported by grant 176W009, from The Welcome Trust, G26264M and 30987M from the Mexican Council of Science and Technology, and grant 6570-12-636-95 from COLCIENCIAS. The Institute Pasteur of Guadeloupe is supported by the Délégation Générale au Réseau International des Instituts Pasteur and is a member of the international network of Pasteur Institutes.

#### REFERENCES

- Alland, D., G. E. Kalkut, A. R. Moss, R. A. McAdam, J. A. Hahn, W. Bosworth, E. Drucker, and B. R. Bloom. 1994. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. *N. Engl. J. Med.* **330**:1710–1716.
- Alland, D., T. S. Whittam, M. B. Murray, M. D. Cave, M. H. Hazbon, K. Dix, M. Kokoris, A. Dueterhoeft, J. A. Eisen, C. M. Fraser, and R. D. Fleischmann. 2003. Modeling bacterial evolution with comparative-genome-based marker systems: application to *Mycobacterium tuberculosis* evolution and pathogenesis. *J. Bacteriol.* **185**:3392–3399.
- Anh, D. D., M. W. Borgdorff, L. N. Van, N. T. Lan, T. van Gorkom, K. Kremer, and D. van Soolingen. 2000. *Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. *Emerg. Infect. Dis.* **6**:302–305.
- Baker, L., T. Brown, M. C. Maiden, and F. Drobniowski. 2004. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. *Emerg. Infect. Dis.* **10**:1568–1577.
- Barczak, A. K., P. Domenech, H. I. Boshoff, M. B. Reed, C. Manca, G. Kaplan, and C. E. Barry III. 2005. In vivo phenotypic dominance in mouse mixed infections with *Mycobacterium tuberculosis* clinical isolates. *J. Infect. Dis.* **192**:600–606.
- Bellamy, R., C. Ruwende, T. Corrah, K. P. McAdam, H. C. Whittle, and A. V. Hill. 1998. Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. *N. Engl. J. Med.* **338**:640–644.
- Bishai, W. R., A. M. Dannenberg, Jr., N. Parrish, R. Ruiz, P. Chen, B. C. Zook, W. Johnson, J. W. Boles, and M. L. Pitt. 1999. Virulence of *Mycobacterium tuberculosis* CDC1551 and H37Rv in rabbits evaluated by Lurie's pulmonary tubercle count method. *Infect. Immun.* **67**:4931–4934.
- 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.
- Caminero, J. A., M. J. Pena, M. I. Campos-Herrero, J. C. Rodriguez, I. Garcia, P. Cabrera, C. Lafoz, S. Samper, H. Takiff, O. Afonso, J. M. Pavon, M. J. Torres, D. van Soolingen, D. A. Enarson, and C. Martin. 2001. Epidemiological evidence of the spread of a *Mycobacterium tuberculosis* strain of the Beijing genotype on Gran Canaria Island. *Am. J. Respir. Crit. Care Med.* **164**:1165–1170.
- Cave, M. D., Z. H. Yang, R. Stefanova, N. Fomukong, K. Ijaz, J. Bates, and K. D. Eisenach. 2005. Epidemiologic import of tuberculosis cases whose isolates have similar but not identical IS6110 restriction fragment length polymorphism patterns. *J. Clin. Microbiol.* **43**:1228–1233.
- Centers for Disease Control and Prevention. 2005. Trends in tuberculosis—United States, 2004. *Morb. Mortal. Wkly. Rep.* **54**:245–249.
- Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry III, F. Tekaiia, 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, B. G. Barrell, et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537–544.

13. Cousins, D. V., R. Bastida, A. Cataldi, V. Quse, S. Redrobe, S. Dow, P. Duignan, A. Murray, C. Dupont, N. Ahmed, D. M. Collins, W. R. Butler, D. Dawson, D. Rodriguez, J. Loureiro, M. I. Romano, A. Alito, M. Zumarraga, and A. Bernardelli. 2003. Tuberculosis in seals caused by a novel member of the *Mycobacterium tuberculosis* complex: *Mycobacterium pinnipedii* sp. nov. *Int. J. Syst. Evol. Microbiol.* **53**:1305–1314.
14. Dale, J. W., D. Brittain, A. A. Cataldi, D. Cousins, J. T. Crawford, J. Driscoll, H. Heersma, T. Lillebaek, T. Quitugua, N. Rastogi, R. A. Skuce, C. Sola, D. Van Soolingen, and V. Vincent. 2001. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardised nomenclature. *Int. J. Tuberc. Lung Dis.* **5**:216–219.
15. DeRiemer, K., L. Garcia-Garcia, M. Bobadilla-del-Valle, M. Palacios-Martinez, A. Martinez-Gamboa, P. M. Small, J. Sifuentes-Osornio, and A. Ponce-de-Leon. 2005. Does DOTS work in populations with drug-resistant tuberculosis? *Lancet* **365**:1239–1245.
16. Drobniewski, F., Y. Balabanova, V. Nikolayevsky, M. Ruddy, S. Kuznetsov, S. Zakharova, A. Melentyev, and I. Fedorin. 2005. Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia. *JAMA* **293**:2726–2731.
17. Drobniewski, F., Y. Balabanova, M. Ruddy, L. Weldon, K. Jeltkova, T. Brown, N. Malomanova, E. Elizarova, A. Melentyev, E. Mutovkin, S. Zhakharova, and I. Fedorin. 2002. Rifampin- and multidrug-resistant tuberculosis in Russian civilians and prison inmates: dominance of the Beijing strain family. *Emerg. Infect. Dis.* **8**:1320–1326.
18. Eisenach, K. D., J. T. Crawford, and J. H. Bates. 1988. Repetitive DNA sequences as probes for *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **26**:2240–2245.
19. Ferdinand, S., G. Valetudie, C. Sola, and N. Rastogi. 2004. Data mining of *Mycobacterium tuberculosis* complex genotyping results using mycobacterial interspersed repetitive units validates the clonal structure of spoligotyping-defined families. *Res. Microbiol.* **155**:647–654.
20. Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. A. Dang, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, E. Kassa-Kelembho, M. L. Ho, A. Makristathis, C. Mammina, G. Martin, P. Mostrom, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasolofoa-Razanamparany, M. Ridell, M. L. Rossetti, F. Stauffer, P. N. Suffys, H. Takiff, J. Texier-Maugein, V. Vincent, J. H. de Waard, C. Sola, and N. Rastogi. 2003. Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. *J. Clin. Microbiol.* **41**:1963–1970.
21. Fleischmann, R. D., D. Alland, J. A. Eisen, L. Carpenter, O. White, J. Peterson, R. DeBoy, R. Dodson, M. Gwinn, D. Haft, E. Hickey, J. F. Kolonay, W. C. Nelson, L. A. Umayam, M. Ermolaeva, S. L. Salzberg, A. Delcher, T. Utterback, J. Weidman, H. Khouri, J. Gill, A. Mikula, W. Bishai, W. R. Jacobs, Jr., J. C. Venter, and C. M. Fraser. 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J. Bacteriol.* **184**:5479–5490.
22. Friedman, C. R., G. C. Quinn, B. N. Kreiswirth, D. C. Perlman, N. Salomon, N. Schluger, M. Lutfey, J. Berger, N. Poltoratskaia, and L. W. Riley. 1997. Widespread dissemination of a drug-susceptible strain of *Mycobacterium tuberculosis*. *J. Infect. Dis.* **176**:478–484.
23. Frothingham, R. 1999. Evolutionary bottlenecks in the agents of tuberculosis, leprosy, and paratuberculosis. *Med. Hypotheses* **52**:95–99.
24. Frothingham, R., and W. A. Meeker-O'Connell. 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology* **144**(Part 5):1189–1196.
25. Gao, Q., K. E. Kripke, A. J. Saldanha, W. Yan, S. Holmes, and P. M. Small. 2005. Gene expression diversity among *Mycobacterium tuberculosis* clinical isolates. *Microbiology* **151**:5–14.
26. Garnier, T., K. Eiglmeier, J. C. Camus, N. Medina, H. Mansoor, M. Pryor, S. Duthoy, S. Grondin, C. Lacroix, C. Monsempe, S. Simon, B. Harris, R. Atkin, J. Doggett, R. Mayes, L. Keating, P. R. Wheeler, J. Parkhill, B. G. Barrell, S. T. Cole, S. V. Gordon, and R. G. Hewinson. 2003. The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl. Acad. Sci. USA* **100**:7877–7882.
27. Glynn, J. R., J. Whiteley, P. J. Bifani, K. Kremer, and D. van Soolingen. 2002. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. *Emerg. Infect. Dis.* **8**:843–849.
28. Goujon, C. P., V. M. Schneider, J. Grofti, J. Montigny, V. Jeantils, P. Astagneau, W. Rozenbaum, F. Lot, C. Frochain-Herchkovitch, N. Delphin, F. Le Gal, J. C. Nicolas, M. C. Milinkovitch, and P. Deny. 2000. Phylogenetic analyses indicate an atypical nurse-to-patient transmission of human immunodeficiency virus type 1. *J. Virol.* **74**:2525–2532.
29. Gutacker, M. M., J. C. Smoot, C. A. Migliaccio, S. M. Ricklefs, S. Hua, D. V. Cousins, E. A. Graviss, E. Shashkina, B. N. Kreiswirth, and J. M. Musser. 2002. Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics* **162**:1533–1543.
30. Hazbon, M. H., and D. Alland. 2004. Hairpin primers for simplified single-nucleotide polymorphism analysis of *Mycobacterium tuberculosis* and other organisms. *J. Clin. Microbiol.* **42**:1236–1242.
31. Hazbón, M. H., M. B. d. Valle, M. I. Guerrero, M. Varma-Basil, I. Filliol, M. Cavatore, R. Colangeli, H. Safi, H. Billman-Jacobe, C. Lavender, J. Fyfe, L. García-García, A. Davidow, M. Brimacombe, C. I. León, T. Porras, M. Bose, F. Chaves, K. D. Eisenach, J. Sifuentes-Osornio, A. P. d. León, M. D. Cave, and D. Alland. 2005. The role of *embB* codon 306 mutations in *Mycobacterium tuberculosis* revisited: a novel association with broad drug resistance and IS6110 clustering rather than ethambutol resistance. *Antimicrob. Agents Chemother.* **49**:3794–3802.
32. Hirsh, A. E., A. G. Tsolaki, K. DeRiemer, 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.
33. Holmes, E. C., S. Nee, A. Rambaut, G. P. Garnett, and P. H. Harvey. 1995. Revealing the history of infectious disease epidemics through phylogenetic trees. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **349**:33–40.
34. Hughes, A. L., R. Friedman, and M. Murray. 2002. Genomewide pattern of synonymous nucleotide substitution in two complete genomes of *Mycobacterium tuberculosis*. *Emerg. Infect. Dis.* **8**:1342–1346.
35. Hunter, P. R., and M. A. Gaston. 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J. Clin. Microbiol.* **26**:2465–2466.
36. Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, and J. van Embden. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* **35**:907–914.
37. Kapur, V., T. S. Whittam, and J. M. Musser. 1994. Is *Mycobacterium tuberculosis* 15,000 years old? *J. Infect. Dis.* **170**:1348–1349.
38. Kato-Maeda, M., J. Sifuentes-Osornio, M. Bobadilla-del-Valle, G. M. Ruiz-Palacios, and A. Ponce-de-Leon. 1999. Drug resistance among acid-fast bacilli. *Lancet* **353**:1709.
39. Kersulyte, D., A. K. Mukhopadhyay, B. Velapatino, W. Su, Z. Pan, C. Garcia, V. Hernandez, Y. Valdez, R. S. Mistry, R. H. Gilman, Y. Yuan, H. Gao, T. Alarcon, M. Lopez-Brea, G. Balakrish Nair, A. Chowdhury, S. Datta, M. Shirai, T. Nakazawa, R. Ally, I. Segal, B. C. Wong, S. K. Lam, F. O. Olfat, T. Boren, L. Engstrand, O. Torres, R. Schneider, J. E. Thomas, S. Czinn, and D. E. Berg. 2000. Differences in genotypes of *Helicobacter pylori* from different human populations. *J. Bacteriol.* **182**:3210–3218.
40. Kubica, T., S. Rusch-Gerdes, and S. Niemann. 2004. The Beijing genotype is emerging among multidrug-resistant *Mycobacterium tuberculosis* strains from Germany. *Int. J. Tuberc. Lung Dis.* **8**:1107–1113.
41. Kulkarni, S., C. Sola, I. Filliol, N. Rastogi, and G. Kadival. 2005. Spoligotyping of *Mycobacterium tuberculosis* isolates from patients with pulmonary tuberculosis in Mumbai, India. *Res. Microbiol.* **156**:588–596.
42. Kumar, D., J. M. Watson, A. Charlett, S. Nicholas, J. H. Darbyshire, and Public Health Laboratory Service/British Thoracic Society/Department of Health Collaborative Group. 1997. Tuberculosis in England and Wales in 1993: results of a national survey. *Thorax* **52**:1060–1067.
43. Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* **5**:150–163.
44. Lari, N., L. Rindi, C. Sola, D. Bonanni, N. Rastogi, E. Tortoli, and C. Garzelli. 2005. Genetic diversity, determined on the basis of *katG463* and *gyrA95* polymorphisms, spoligotyping, and IS6110 typing, of *Mycobacterium tuberculosis* complex isolates from Italy. *J. Clin. Microbiol.* **43**:1617–1624.
45. Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* **95**:3140–3145.
46. Malik, A. N., and P. Godfrey-Faussett. 2005. Effects of genetic variability of *Mycobacterium tuberculosis* strains on the presentation of disease. *Lancet Infect. Dis.* **5**:174–183.
47. Manca, C., L. Tsenova, A. Bergtold, S. Freeman, M. Tovey, J. M. Musser, C. E. Barry III, V. H. Freedman, and G. Kaplan. 2001. Virulence of a *Mycobacterium tuberculosis* clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN- $\alpha$ /beta. *Proc. Natl. Acad. Sci. USA* **98**:5752–5757.
48. Mazars, E., S. Lesjean, A. L. Banuls, M. Gilbert, V. Vincent, B. Gicquel, M. Tibaeyre, C. Locht, and P. Supply. 2001. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc. Natl. Acad. Sci. USA* **98**:1901–1906.
49. 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.
50. Musser, J. M., A. Amin, and S. Ramaswamy. 2000. Negligible genetic diversity of *Mycobacterium tuberculosis* host immune system protein targets: evidence of limited selective pressure. *Genetics* **155**:7–16.

51. Nakamura, Y. 2001. Molecular analyses of the serotype of *Cryptococcus neoformans*. *Nippon Ishinkin Gakkai Zasshi* **42**:69–74.
52. Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York, N. Y.
53. Pearson, T., J. D. Busch, J. Ravel, T. D. Read, S. D. Rhoton, J. M. U'Ren, T. S. Simonson, S. M. Kachur, R. R. Leadem, M. L. Cardon, M. N. Van Ert, L. Y. Huynh, C. M. Fraser, and P. Keim. 2004. Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proc. Natl. Acad. Sci. USA* **101**:13536–13541.
54. Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959.
55. Reed, M. B., P. Domenech, C. Manca, H. Su, A. K. Barczak, B. N. Kreiswirth, G. Kaplan, and C. E. Barry III. 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* **431**:84–87.
56. Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**:2496–2497.
57. Schork, N. J., D. Fallin, and J. S. Lanchbury. 2000. Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin. Genet.* **58**:250–264.
58. Sebban, M., I. Mokrousov, N. Rastogi, and C. Sola. 2002. A data-mining approach to spacer oligonucleotide typing of *Mycobacterium tuberculosis*. *Bioinformatics* **18**:235–243.
59. Shemyakin, I. G., V. N. Stepanshina, I. Y. Ivanov, M. Y. Lipin, V. A. Anisimova, A. G. Onasenko, O. V. Korobova, and T. M. Shinnick. 2004. Characterization of drug-resistant isolates of *Mycobacterium tuberculosis* derived from Russian inmates. *Int. J. Tuberc. Lung Dis.* **8**:1194–1203.
60. Singh, U. B., N. Suresh, N. V. Bhanu, J. Arora, H. Pant, S. Sinha, R. C. Aggarwal, S. Singh, J. N. Pande, C. Sola, N. Rastogi, and P. Seth. 2004. Predominant tuberculosis spoligotypes, Delhi, India. *Emerg. Infect. Dis.* **10**:1138–1142.
61. Sola, C., I. Filliol, M. C. Gutierrez, I. Mokrousov, V. Vincent, and N. Rastogi. 2001. Spoligotype database of *Mycobacterium tuberculosis*: biogeographic distribution of shared types and epidemiologic and phylogenetic perspectives. *Emerg. Infect. Dis.* **7**:390–396.
62. Sola, C., I. Filliol, E. Legrand, I. Mokrousov, and N. Rastogi. 2001. *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.* **53**:680–689.
63. 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.
64. 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.
65. Toungousova, O. S., D. A. Caugant, P. Sandven, A. O. Mariandyshev, and G. Bjune. 2004. Impact of drug resistance on fitness of *Mycobacterium tuberculosis* strains of the W-Beijing genotype. *FEMS Immunol. Med. Microbiol.* **42**:281–290.
66. Toungousova, O. S., P. Sandven, A. O. Mariandyshev, N. I. Nizovtseva, G. Bjune, and D. A. Caugant. 2002. Spread of drug-resistant *Mycobacterium tuberculosis* strains of the Beijing genotype in the Archangel Oblast, Russia. *J. Clin. Microbiol.* **40**:1930–1937.
67. van Embden, J. D., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. Shinnick, et al. 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J. Clin. Microbiol.* **31**:406–409.
68. van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. van Embden. 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. *J. Clin. Microbiol.* **33**:3234–3238.
69. Victor, T. C., P. E. de Haas, A. M. Jordaana, G. D. van der Spuy, M. Richardson, D. van Soolingen, P. D. van Helden, and R. Warren. 2004. Molecular characteristics and global spread of *Mycobacterium tuberculosis* with a Western Cape F11 genotype. *J. Clin. Microbiol.* **42**:769–772.
70. Warren, R. M., E. M. Streicher, S. L. Sampson, G. D. van der Spuy, M. Richardson, D. Nguyen, M. A. Behr, T. C. Victor, and P. D. van Helden. 2002. Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. *J. Clin. Microbiol.* **40**:4457–4465.
71. Weiss, R. A., and A. J. McMichael. 2004. Social and environmental risk factors in the emergence of infectious diseases. *Nat. Med.* **10**:S70–S76.
72. Wilkinson, R. J., M. Llewelyn, Z. Toossi, P. Patel, G. Pasvol, A. Lalvani, D. Wright, M. Latif, and R. N. Davidson. 2000. Influence of vitamin D deficiency and vitamin D receptor polymorphisms on tuberculosis among Gujarati Asians in west London: a case-control study. *Lancet* **355**:618–621.

## ERRATUM

### Global Phylogeny of *Mycobacterium tuberculosis* Based on Single Nucleotide Polymorphism (SNP) Analysis: Insights into Tuberculosis Evolution, Phylogenetic Accuracy of Other DNA Fingerprinting Systems, and Recommendations for a Minimal Standard SNP Set

Ingrid Filliol, Alifiya S. Motiwala, Magali Cavatore, Weihong Qi, Manzour Hernando Hazbón, Miriam Bobadilla del Valle, Janet Fyfe, Lourdes García-García, Nalin Rastogi, Christophe Sola, Thierry Zoio, Marta Inírida Guerrero, Clara Inés León, Jonathan Crabtree, Sam Angiuoli, Kathleen D. Eisenach, Riza Durmaz, Moses L. Joloba, Adrian Rendón, José Sifuentes-Osornio, Alfredo Ponce de León, M. Donald Cave, Robert Fleischmann, Thomas S. Whittam, and David Alland

*Division of Infectious Disease, Department of Medicine and the Ruy V. Lourenço Center for the Study of Emerging and Re-emerging Pathogens, New Jersey Medical School, University of Medicine and Dentistry of New Jersey, Newark, New Jersey; National Food Safety and Toxicology Center, Michigan State University, East Lansing, Michigan; The Institute for Genomic Research, Rockville, Maryland; Central Arkansas Veterans Healthcare System (CAVHS), Departments of Pathology, Microbiology-Immunology, and Neurobiology and Developmental Sciences, University of Arkansas for Medical Sciences, Little Rock, Arkansas; Department of Infectious Diseases, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, and Unidad de Tuberculosis Instituto Nacional de Salud Pública, Mexico City, and Pulmonary Services, University Hospital of Monterrey, Universidad Autónoma de Nuevo León, Nuevo León, Mexico; Victorian Mycobacterium Reference Laboratory, Victorian Infectious Diseases Reference Laboratory, North Melbourne, Victoria 3051, Australia; Unité de la Tuberculose et des Mycobactéries, Institut Pasteur de Guadeloupe, Morne Jolivère, F-97183 Abymes-Cedex, Guadeloupe; Grupo de Micobacterias, Subdirección de Investigación, Instituto Nacional de Salud, Bogotá, Colombia; Department of Clinical Microbiology, Faculty of Medicine, Inonu University, Malatya, Turkey; and Departments of Medicine and Medical Microbiology, Makerere University, Kampala, Uganda*

Volume 188, no. 2, p. 759–772, 2006. Page 762: Figure 3 should appear as shown on the following page.

