Genome-Wide Analysis of Synonymous Single Nucleotide Polymorphisms in *Mycobacterium tuberculosis* Complex Organisms: Resolution of Genetic Relationships Among Closely Related Microbial Strains

Michaela M. Gutacker,* James C. Smoot,*,[1,2] Cristi A. Lux Migliaccio,*,[1] Stacy M. Ricklefs,* Su Hua,* Debby V. Cousins,[†] Edward A. Graviss,[‡] Elena Shashkina,[§] Barry N. Kreiswirth[§] and James M. Musser*,[3]

**Laboratory of Human Bacterial Pathogenesis, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, Montana 59840,* [†]*Department of Agriculture, Australian Reference Laboratory for Bovine Tuberculosis, South Perth 6151, Australia,* [‡]*Department of Pathology, Baylor College of Medicine, Houston, Texas 77030 and* [§]*Public Health Research Institute Tuberculosis Center, New York, New York 10016*

Manuscript received May 7, 2002
Accepted for publication September 24, 2002

ABSTRACT

Several human pathogens (*e.g.*, *Bacillus anthracis*, *Yersinia pestis*, *Bordetella pertussis*, *Plasmodium falciparum*, and *Mycobacterium tuberculosis*) have very restricted unselected allelic variation in structural genes, which hinders study of the genetic relationships among strains and strain-trait correlations. To address this problem in a representative pathogen, 432 *M. tuberculosis* complex strains from global sources were genotyped on the basis of 230 synonymous (silent) single nucleotide polymorphisms (sSNPs) identified by comparison of four genome sequences. Eight major clusters of related genotypes were identified in *M. tuberculosis* sensu stricto, including a single cluster representing organisms responsible for several large outbreaks in the United States and Asia. All *M. tuberculosis* sensu stricto isolates of previously unknown phylogenetic position could be rapidly and unambiguously assigned to one of the eight major clusters, thus providing a facile strategy for identifying organisms that are clonally related by descent. Common clones of *M. tuberculosis* sensu stricto and *M. bovis* are distinct, deeply branching genotypic complexes whose extant members did not emerge directly from one another in the recent past. sSNP genotyping rapidly delineates relationships among closely related strains of pathogenic microbes and allows construction of genetic frameworks for examining the distribution of biomedically relevant traits such as virulence, transmissibility, and host range.

A common theme that has emerged from molecular population genetic analysis of pathogenic bacteria is that biomedically relevant traits, such as host range and virulence, are nonrandomly distributed among phylogenetic lineages (Musser 1996). Trait-lineage relationships exist, in part, because pathogenic bacterial species often have a level of genetic diversity far in excess of the variation present in higher eukaryotic organisms such as humans. Genetic variation in bacteria is due to differences in gene content and nucleotide variation in or between structural genes. Differences in gene content are caused primarily by gene deletion or by acquisition of mobile elements such as plasmids and bacteriophages (Ochman *et al.* 2000). Allelic variation arises from random nucleotide mutation, sometimes followed by selection, and from horizontal gene transfer and intragenic recombination events (Reid *et al.* 2001). Two classes of substitutions, referred to as synonymous and nonsynonymous single nucleotide polymorphisms (SNPs), can occur in genes that encode proteins (Kimura 1983; Schork *et al.* 2000; Gut 2001). Nonsynonymous SNPs (nsSNPs) result in amino acid replacements and hence provide substrate for evolutionary selection. In contrast, synonymous SNPs (sSNPs) do not alter the structure of proteins and are evolutionarily neutral or nearly so (Kimura 1983; Schork *et al.* 2000; Gut 2001). Because sSNPs are functionally neutral and easy to assay (Kimura 1983; Schork *et al.* 2000; Gut 2001), they provide useful targets for large-scale molecular population genetic studies examining evolutionary relationships among bacterial strains, especially in strongly clonal species. Moreover, sSNP genotypes can be readily analyzed with computational tools developed over decades of population genetic research. In the postgenomic era, SNPs provide a simple and fast way to compare entire genomes in many bacterial strains.

Mycobacterium tuberculosis is the most successful human pathogen worldwide, responsible for 3 million deaths each year and extensive morbidity and mortality (World Health Organization 1998). *M. tuberculosis* is a mem-

[1]These authors contributed equally to this work.

[2]*Present address:* Civil and Environmental Engineering, University of Washington, Seattle, WA 98195.

[3]*Corresponding author:* Laboratory of Human Bacterial Pathogenesis, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 903 S. 4th St., Hamilton, MT 59840. E-mail: jmusser@niaid.nih.gov

ber of the *M. tuberculosis* complex, a group of five closely related "sibling" species [*M. tuberculosis* sensu stricto (s.s.), *M. africanum*, *M. microti*, *M. bovis*, and *M. canettii*] that cause tuberculosis in humans and animals (Baess 1979; Kapur *et al.* 1994; Nolte and Metchock 1995; Feizabadi *et al.* 1996; Sreevatsan *et al.* 1997a; Musser *et al.* 2000). Study of 56 genes in several hundred *M. tuberculosis* complex strains suggested that there is about one synonymous nucleotide substitution per 10,000 nucleotide sites (Kapur *et al.* 1994; Sreevatsan *et al.* 1997a; Musser *et al.* 2000). Restricted allelic variation limits the utility of multilocus sequence analysis (Maiden *et al.* 1998) for estimating genetic relationships among *M. tuberculosis* strains and for studying relationships between strain genotype and patient phenotype.

Although many molecular methods have been used for categorizing *M. tuberculosis* strains (Ross *et al.* 1992; Groenen *et al.* 1993; van Embden *et al.* 1993; Frothingham and Meeker-O'Connell 1998; Kurepina *et al.* 1998; Ramaswamy and Musser 1998; Kremer *et al.* 1999; Soini *et al.* 2000; Mazars *et al.* 2001; Supply *et al.* 2001), phylogenetic relationships among this group of organisms have not been resolved. nsSNPs located in codon 463 of the *katG* gene and codon 95 of the *gyrA* gene permit all *M. tuberculosis* strains to be assigned to three principal genetic groups (Sreevatsan *et al.* 1997a). The distribution of these nsSNPs provided evidence supporting the hypothesis that principal genetic group 1 is ancestral to group 2 and that group 2 is ancestral to group 3. However, because the original genetic groupings are based on only two nsSNPs, the evolutionary hypothesis of *M. tuberculosis* outlined by Sreevatsan *et al.* (1997a) requires additional investigation. Moreover, the worldwide threat of *M. tuberculosis* to human health emphasizes the need to develop rapid methods to identify genetic relationships among all strains, especially among organisms responsible for large infection outbreaks, drug-resistant strains, and strains that cause severe clinical disease. The availability of genome sequences of two *M. tuberculosis* strains (Cole *et al.* 1998; http://www.tigr.org) and of partial genome sequences of a third *M. tuberculosis* strain (http://www.tigrblast.tigr.org) and an *M. bovis* strain (http://www.sanger.ac.uk/Projects/M_bovis) facilitated resolution of genetic relationships among *M. tuberculosis* complex organisms by large-scale sSNP analysis. Our results indicate that sSNP genotyping permits all strains of closely related pathogens to be assigned to lineages that are identical or related by descent and removes a critical barrier to population-based studies of the relationships between strain genotypes and patient phenotypes.

MATERIALS AND METHODS

Bacterial strains: *M. tuberculosis* complex isolates ($n = 432$) were recovered worldwide from patients with pulmonary and extrapulmonary tuberculosis and from diseased animals (see Table 1 in the supplementary material at http://www.genetics.org/supplemental). Many of the organisms studied were characterized previously by IS*6110* profiling and spoligotyping (Bifani *et al.* 1996; Sreevatsan *et al.* 1997a; Kurepina *et al.* 1998; Bifani *et al.* 1999; Yaganehdoost *et al.* 1999; Soini *et al.* 2000, 2001; Mathema *et al.* 2002). The strain collection includes 112 isolates representing drug-susceptible and multidrug-resistant strains causing outbreaks in New York City, New Jersey, and Houston, and abundantly occurring strains ($n = 104$) of the same IS*6110* profiles isolated from patients in the United States, Latin America, Africa, and Asia. The 112 isolates represent the known breadth of IS*6110* and spoligotype diversity present in the species (see Table 1 in the supplementary material at http://www.genetics.org/supplemental) and will be referred to as the core group of isolates. We also analyzed 90 isolates recovered from patients in Houston enrolled in a population-based study of tuberculosis as well as five pairs of strains cultured from patients with reactivation tuberculosis. These 90 isolates include all organisms with low copy numbers of IS*6110* (five or fewer copies) and all strains with unique IS*6110* profiles obtained over a 6-month period.

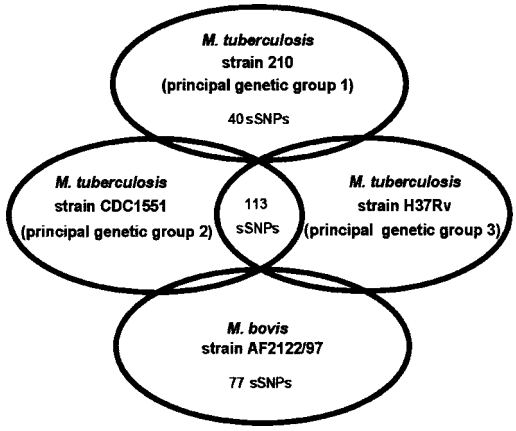


Figure 1.—Schematic showing distribution of sSNPs analyzed. Complete genome sequences are available for *M. tuberculosis* s.s. strains H37Rv (Cole *et al.* 1998) and CDC1551 (http://www.tigrblast.tigr.org) and partial genome sequences are available for *M. tuberculosis* s.s. strain 210 (http://www.tigrblast.tigr.org) and *M. bovis* AF2122/97 (http://www.sanger.ac.uk/Projects/M_bovis). sSNPs randomly distributed around the chromosome were analyzed. All sSNPs were sequence verified.

M. bovis isolates ($n = 103$) were recovered from 11 host species (cows, pigs, possum, water buffalo, deer, badger, elk, eland, yak, bison, and humans) in 14 countries (United States, Canada, Tanzania, Spain, Australia, New Zealand, Sweden, Wales, Ireland, England, the former Soviet Union, Iran, Japan, and Vietnam). Isolates of *M. africanum* ($n = 8$), *M. microti* ($n = 4$), and *M. canettii* ($n = 1$) also were studied (see Table 2 in the supplementary material at http://www.genetics.org/supplemental). All isolates were assigned to species on the basis of conventional biochemical test results (Nolte and Metchock 1995).

Identification and verification of sSNPs: Several strategies were used to identify sSNPs present in four *M. tuberculosis* complex genomes (Figure 1). First, the genomes of *M. tuberculosis* strains H37Rv (Cole *et al.* 1998) and CDC1551 (Valway *et al.* 1998; http://www.tigr.org) were aligned with CrossMatch (P. Green, http://www.genome.washington.edu/UWGC/analysistools/Swat.cfm) to identify polymorphic sites, which were differentiated as sSNPs and nsSNPs on the basis of the

inferred protein sequences. All sSNPs were verified by sequencing the relevant gene region in strains TN587, CDC1551, and H37Rv, representatives of principal genetic groups 1, 2, and 3, respectively (Sreevatsan *et al.* 1997a). To identify additional sSNPs, the two available *M. tuberculosis* genome sequences were also aligned with a partially completed *M. bovis* genome sequence (http://www.sanger.org). sSNPs were differentiated from nsSNPs by comparison of the inferred protein sequences and verified by sequencing the relevant gene region in strains TN587, CDC1551, and H37Rv and in *M. bovis* strains TN10130 and TN12465. Similar procedures were used to identify and verify sSNPs between H37Rv, CDC1551, and the partially completed genome of *M. tuberculosis* strain 210 (a principal genetic group 1 organism; Beggs *et al.* 2000). The PCR primers used to amplify and sequence the gene segments containing putative sSNPs are listed in Table 3 in the supplementary material at http://www.genetics.org/supplemental.

sSNP analysis: The SNaPshot primer extension method (Applied Biosystems, Foster City, CA) was used to analyze sSNPs with the primers listed in Table 3 in the supplementary material at http://www.genetics.org/supplemental. The SNaPshot technique is based on addition of a single fluorescently labeled ddNTP to the 3′ end of an unlabeled specific oligonucleotide primer that hybridizes to its target DNA located contiguous to the SNP of interest. Labeling reactions were performed with the SNaPshot kit according to the instructions supplied by the manufacturer, except that the reaction mixture consisted of 2 μl of PCR template, 4 μl of SNaPshot ready reaction premix, 4 μl of dH_2O, and 1 μl 0.2 mm primer. Data were generated with an ABI Prism 3700 automated sequencer (Applied Biosystems).

IS*6110* RFLP profiling and spoligotyping: IS*6110* restriction fragment length polymorphism (RFLP) profiling and spoligotyping were conducted by standard methods (van Embden *et al.* 1993; van Soolingen *et al.* 1995).

Data analysis: The sSNP data were concatenated, resulting in one character string (nucleotide sequence) for each strain. Phylogenetic and molecular evolution analyses were conducted with *MEGA* version 2.1 (http://www.megasoftware.net/), using the neighbor-joining method with 1000 bootstrap replicates and distance calculated using the number of different sSNPs.

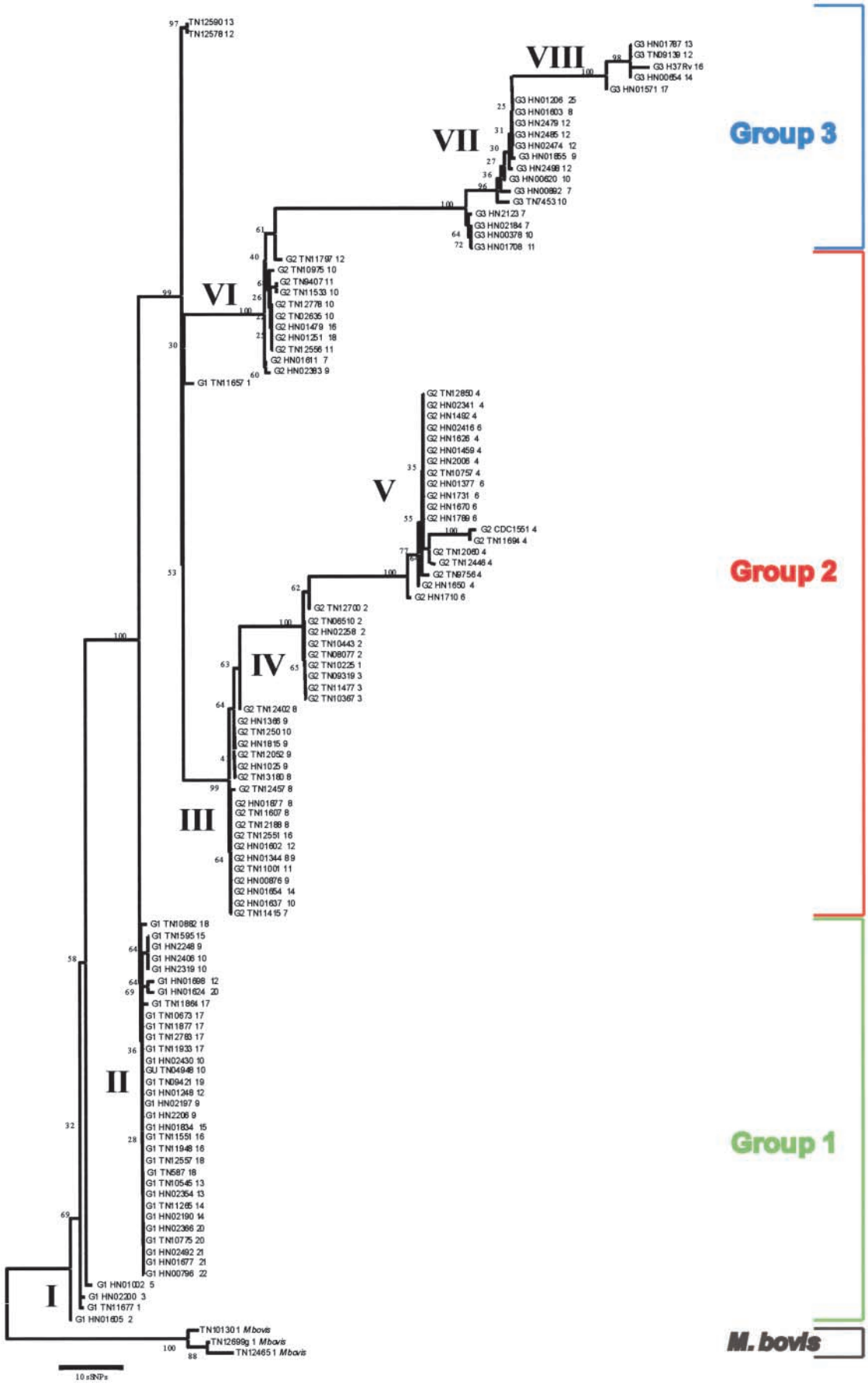
RESULTS

Confirmation of restricted allelic variation in *M. tuberculosis*: Comparison of the 4.4-Mb genome sequences of *M. tuberculosis* strain H37Rv and CDC1551 identified only ~900 SNPs located in open reading frames, confirming the restricted level of structural gene sequence variation reported previously (Kapur *et al.* 1994; Sreevatsan *et al.* 1997a; Musser *et al.* 2000). Approximately 65% of the SNPs were nsSNPs, an unanticipated result given that sSNPs are expected to outnumber nsSNPs except under situations involving positive selection (Kimura 1983). Ninety-one percent of all putative sSNPs studied were confirmed to be true sSNPs (not genome-sequencing artifacts) by sequencing genomic DNA purified from strains H37Rv and CDC1551 (data not shown). A total of 148 sequence-confirmed sSNPs distributed randomly around the chromosome were used for initial analysis.

***M. tuberculosis* s.s. isolates comprise eight major clusters of genotypes:** A core group of 112 isolates representing the breadth of diversity in the species, as assessed by IS*6110* copy number and profile, spoligotype, and principal genetic group, was genotyped for the 148 sSNPs. Of the 148 sSNPs, 35 were polymorphic only in strain H37Rv or CDC1551. The topology of the phylogenetic tree generated from the data set that included all 148 sSNPs was virtually identical to the tree based on the 113 sSNPs. The effect of including the 35 sSNPs present only in strains H37Rv or CDC1551 was to accentuate the branch distance between these 2 isolates and related genotypes (data not shown). The phylogenetic tree generated on the basis of the remaining 113 sSNPs in the core group of 112 strains contained eight major clusters of related genotypes arbitrarily designated I–VIII and marked by high bootstrap values (Figure 2). Clusters I and II contain all isolates of principal genetic group 1; clusters III–VI are composed of isolates of principal genetic group 2; and clusters VII and VIII contain all isolates of principal genetic group 3. Hence, principal genetic groups and major sSNP genotypic clusters correlated strongly. The sSNP analysis also demonstrated that principal genetic group 3 organisms (clusters VII and VIII) arose from a principal group 2 precursor cell and identified the ancestor as a cluster VI-related organism.

Analysis of sSNPs identified by comparing the genome sequence of strains H37Rv and CDC1551 with a principal genetic group 1 *M. tuberculosis* strain and an *M. bovis* strain: Previous studies (Sreevatsan *et al.* 1997a; Musser *et al.* 2000; Soini *et al.* 2000, 2001) of genetic variation among *M. tuberculosis* strains indicated that isolates of principal genetic group 1 were more polymorphic than suggested by analysis of the 113 sSNPs. For example, principal genetic group 1 organisms had the greatest breadth of IS*6110* copy number (Sreevatsan *et al.* 1997a) and substantial variation in spoligotype pattern (Soini *et al.* 2000, 2001). One hypothesis to explain the relative lack of sSNPs identified among principal genetic group 1 organisms is that the strategy used to identify sSNPs minimized the likelihood of detecting nucleotide sites that are polymorphic among group 1 organisms. This would be the case if principal genetic group 1 organisms accumulated many sSNPs after diverging from the last common ancestor giving rise to strain CDC1551 (principal genetic group 2) and strain H37Rv (principal genetic group 3). In essence, the relative lack of sSNP diversity among principal group 1 organisms could be an artifact caused by identifying SNPs solely on the basis of comparing genomes from principal genetic group 2 (CDC1551) and group 3 (H37Rv) organisms.

Principal genetic group 1 organisms have been responsible for many large tuberculosis outbreaks in the United States and for a very large proportion of cases in the former Soviet Union and many Asian countries


VIII
VII
VI
V
IV
III
II
I
Group 3
Group 2
Group 1
M. bovis
10 sSNPs

(Bifani *et al.* 1996, 1999; Sreevatsan *et al.* 1997a; Anh *et al.* 2000; Beggs *et al.* 2000; Park *et al.* 2000; Soini *et al.* 2000; Chan *et al.* 2001). Resolution of clonal relationships among principal genetic group 1 organisms is important because of their medical importance worldwide. To test the hypothesis that principal genetic group 1 organisms carry additional sSNP diversity, the two completed *M. tuberculosis* genome sequences (H37Rv and CDC1551) were aligned and compared, in turn, with partially completed genome sequences of strain 210 (principal genetic group 1; http://www.tigrblast.tigr.org) and of *M. bovis* strain AF2122/97 (http://www.sanger.ac.uk/Projects/M_bovis). Analysis of a random sample of 104 sequence-verified sSNPs among 91 strains (all group 1 organisms among the core group of 112 strains and representative isolates of each major group 2 and group 3 lineage and a selection of *M. bovis* isolates) confirmed the hypothesis. In addition, several distinct subpopulations of principal genetic group 1 organisms were resolved (Figure 3).

Phylogenetically informative sSNPs for high-throughput strain genotyping and population-based studies of tuberculosis biology: Population-based studies of *M. tuberculosis* epidemiology have been hindered because at least 50% of strains have unique or low-copy-number IS*6110* RFLP profiles (Small *et al.* 1993; Bishai *et al.* 1998; Lockman *et al.* 2001) and hence their genetic relationships to other strains are not known. This problem has greatly limited the ability to study correlations between strain genotypes and patient phenotypes, because data for at least 50% of patients are uninformative.

To test the hypothesis that this barrier to population-based studies can be removed by analysis of sSNPs, we identified 27 highly informative sSNPs and characterized 90 isolates from Houston. These 27 sSNPs were chosen such that they proportionally represent each of the eight genetic clusters and all major subclusters. Moreover, analysis of the 112 core strains for these 27 sSNPs produced a phylogenetic tree with the same overall topology as the comprehensive sSNP data set (data not shown). The 90 isolates represent a 6-month strain sample that included all organisms with low copy numbers of IS*6110* (five or fewer copies) and all strains with unique IS*6110* RFLP profiles. All 90 isolates could be assigned readily to one of the eight major phylogenetic clusters (data not shown). Hence, sSNP genotyping successfully resolved the phylogenetic position of these organisms, thereby facilitating population-based studies of tuberculosis.

Phylogenetic clusters and IS*6110* RFLP profiles: Because of the considerable controversy regarding the amount of phylogenetic signal present in IS*6110* profiles (Fomukong *et al.* 1997; Sreevatsan *et al.* 1997a; Mazars *et al.* 2001), we next examined the distribution of IS*6110* RFLP profiles among the strains assigned to each of the eight major phylogenetic clusters of *M. tuberculosis*. This insertion sequence is present in virtually all *M. tuberculosis* isolates and is highly polymorphic in copy number and position in the chromosome (van Embden *et al.* 1993; Fomukong *et al.* 1997; Sreevatsan *et al.* 1997a; Mazars *et al.* 2001). IS*6110* profiling has been widely used in epidemiologic studies to infer genetic relationships among strains (Small *et al.* 1993; van Embden *et al.* 1993; Bifani *et al.* 1996, 1999; Bishai *et al.* 1998; Yaganehdoost *et al.* 1999; Lockman *et al.* 2001). Several investigators have assigned *M. tuberculosis* s.s. to two primary phylogenetic lineages on the basis of low and high copy numbers of IS*6110* (Fomukong *et al.* 1997; Mazars *et al.* 2001), but the distribution of nsSNPs in *katG* and *gyrA* is at variance with this idea (Sreevatsan *et al.* 1997a).

Two important findings were revealed by our sSNP analysis. First, we found that there is no simple relationship between IS*6110* copy number and phylogenetic lineage. For example, isolates with relatively few copies of IS*6110* were present in very distinct phylogenetic lineages (Figure 4). Importantly, all isolates of principal genetic group 1 were genetically allied in two clusters by the sSNP analysis; however, these organisms contained from 0 to >20 copies of IS*6110*. Hence, IS*6110* copy number alone is not phylogenetically informative.

Second, sSNP analysis permitted unambiguous identification of phylogenetic relationships among strains with similar or identical IS*6110* profiles from diverse geographic areas. This could not be accomplished previously because of the possibility that similarity of the IS*6110* RFLP profile among strains with no known direct epidemiologic link was caused by convergence to the same RFLP profile, rather than by identity by descent. We identified one subcluster containing strains with the W IS*6110* profile (organisms responsible for many hundreds of cases of tuberculosis in New York City; Bifani *et al.* 1996) and strains of the closely related IS*6110* profiles 003 and 033, which together caused 167 cases of tuberculosis in Harris County, Texas, between

Figure 2.—Phylogenetic tree showing estimates of genetic relationships among 112 *M. tuberculosis* s.s. isolates, based on analysis of 113 sSNPs. The tree was generated using the neighbor-joining method with 1000 bootstrap replicates and distance calculation using the number of different sSNP loci (http://www.megasoftware.net/). The eight major clusters were arbitrarily designated I–VIII. The group designations refer to the three principal genetic groups of *M. tuberculosis* described previously on the basis of SNPs located at codon 463 in the *katG* gene and codon 95 in the *gyrA* gene (Sreevatsan *et al.* 1997a). Each strain is listed with a principal genetic group number (G1, G2, or G3); the strain number; and the number of IS*6110* hybridizing bands (one- or two-digit number).

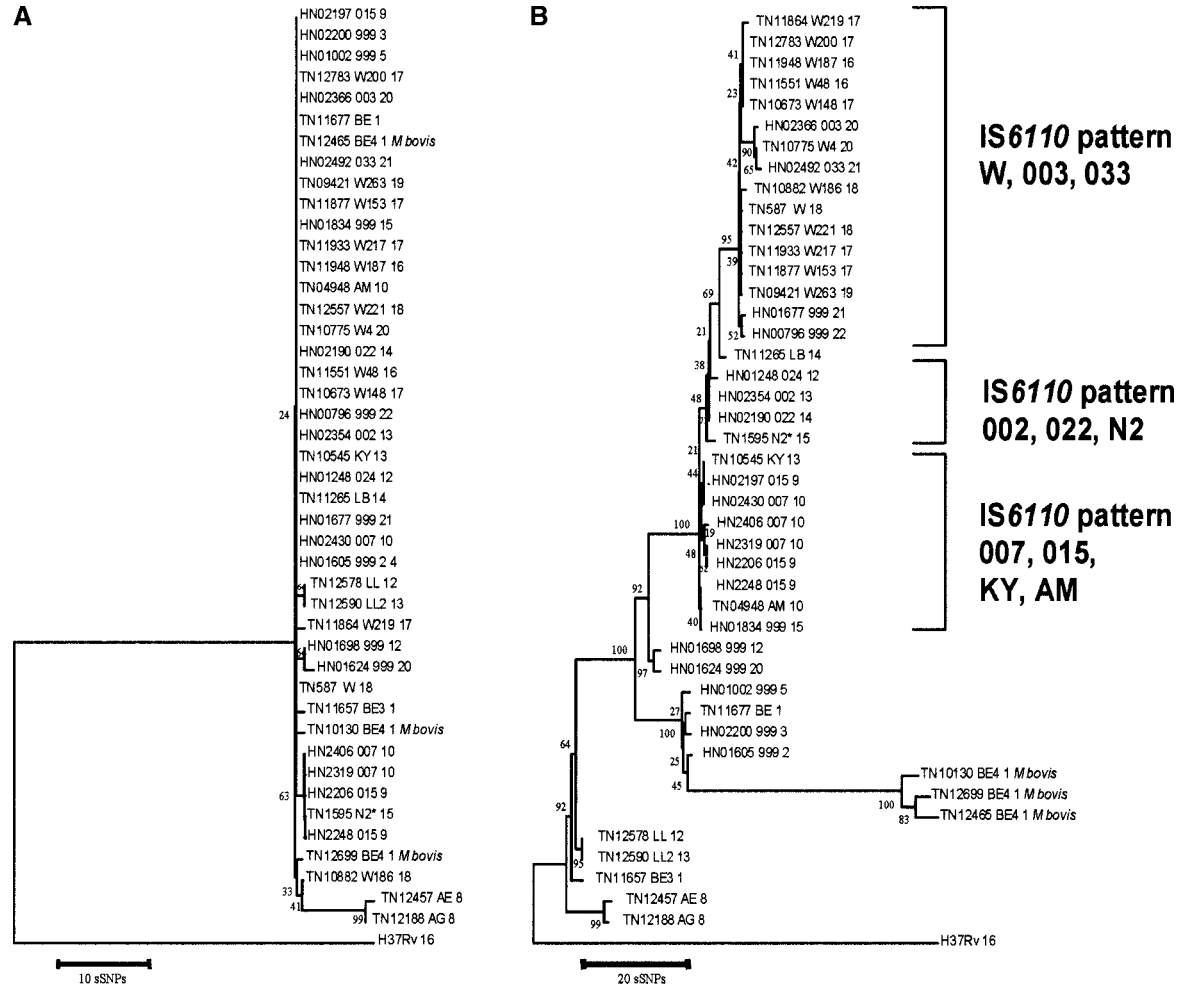


FIGURE 3.—Comparison of two phylogenetic trees showing the relationships among principal genetic group 1 organisms. The phylogenetic trees were constructed using (A) 113 sSNPs obtained from the comparison of genomes of principal genetic group 2 (strain CDC1551) and group 3 (strain H37Rv) organisms and (B) these 113 sSNPs plus 104 new sSNPs identified by comparing the above-described genomes (CDC1551 and H37Rv) with the partial genome sequences of *M. tuberculosis* group 1 strain 210 and *M. bovis* strain AF2122/97. Both trees were generated using the neighbor-joining method with 1000 bootstrap replicates and distance was calculated using the number of sSNPs (http://www.megasoftware.net/). The IS*6110* RFLP profile and copy number are indicated next to each strain. The figure shows that the lack of sSNP diversity among principal group 1 organisms (A) is caused by identifying SNPs on the basis of comparing genomes of principal genetic group 2 and 3 organisms. Inclusion of 104 additional sSNPs found by comparing *M. tuberculosis* groups 2 and 3 with *M. tuberculosis* group 1 and *M. bovis* genomes reveals genetic subpopulations of principal genetic group 1 with similar or identical IS*6110* RFLP profiles (W, 003, 033; 002, 022, N2; 007, 015, KY, AM).

1995 and 2001 (Figure 3). Similarly, another subcluster contained strains with related IS*6110* RFLP profiles designated 002, 022, and N2 that have caused abundant disease in Harris County and New York City ($n = 262$ cases). In another example, strains with IS*6110* profiles designated 007 and 015 caused 169 cases of tuberculosis in Harris County between 1995 and 2001. sSNP genotyping demonstrated that these organisms were clonally related to strains from Siberia and New York City, with IS*6110* patterns designated KY and AM, respectively (Figure 3).

Utility of sSNP genotyping for epidemiologic study purposes: We next assessed the utility of sSNP genotyping for epidemiologic purposes. Strains grouped together on the basis of IS*6110* RFLP profile and spoligotype, and known to be related by conventional epidemiologic investigation strategies such as contact tracing (YAGANEHDOOST *et al.* 1999), were studied for sSNP genotype. Strains assigned to 11 documented case clusters (11 distinct IS*6110* RFLP profiles) were analyzed. Each IS*6110* print group contained strains collected over at least 3 years. All isolates with identical IS*6110* pattern and spoligotype had the same sSNP genotype (data not shown). To further investigate the epidemiologic utility of sSNP analysis, we examined the sSNP profile of five pairs of strains cultured from patients with reactivation tuberculosis. Paired isolates cultured from the same patient had sSNP genotypes identical to one another, as

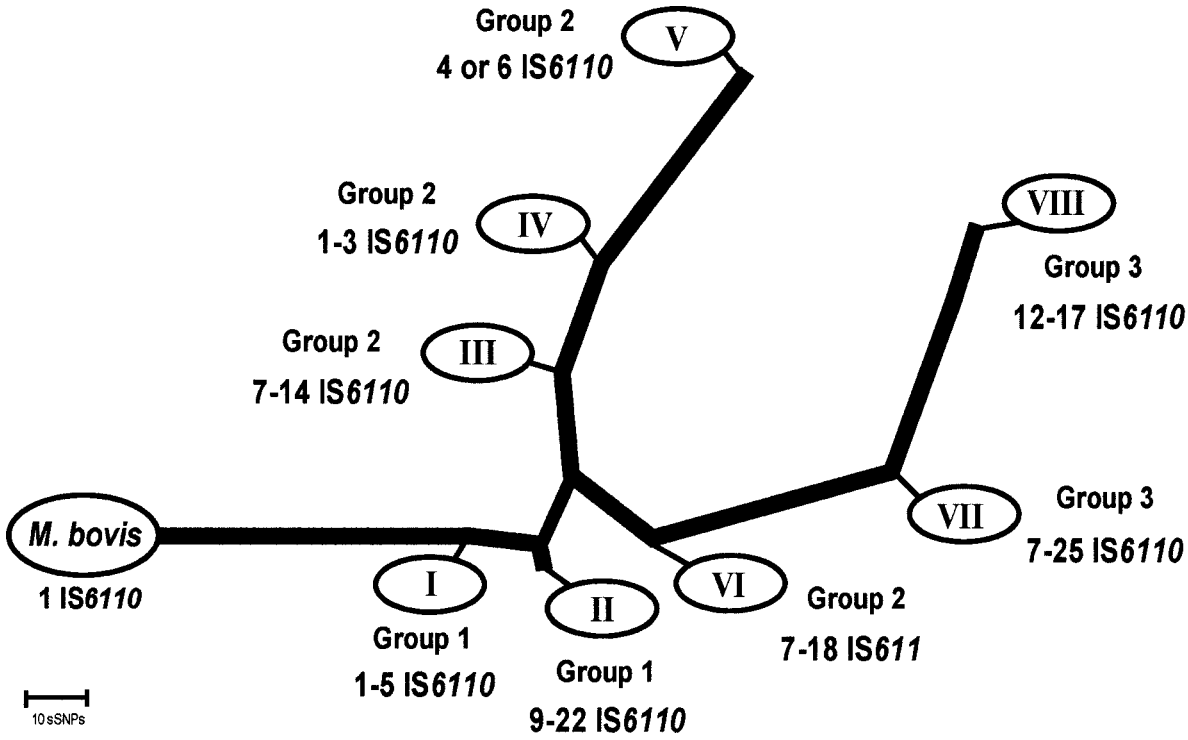


FIGURE 4.—Relationships among IS*6110* copy number and eight major genetic clusters of *M. tuberculosis* s.s. identified by analysis of 113 sSNPs. The tree is modified from the one shown in Figure 2. A sample of 112 core isolates of *M. tuberculosis* s.s. from global sources representing the span of species diversity was studied for 148 sSNPs randomly distributed around the chromosome. Thirty-five of the sSNPs were polymorphic in only strain H37Rv or CDC1551 ("monomorphic sSNPs") and were not considered further. The phylogenetic tree is based on the remaining 113 sSNPs and was generated using the neighbor-joining method with 1000 bootstrap replicates and distance calculation using the number of different sSNP loci (http://www.megasoftware.net/). The eight major clusters were arbitrarily designated I–VIII. The group designations refer to the three principal genetic groups of *M. tuberculosis* described previously on the basis of SNPs located at codon 463 in the *katG* gene and codon 95 in the *gyrA* gene (SREEVATSAN *et al.* 1997a). The number of IS*6110* hybridizing copies identified in strains assigned to each of the eight major genetic clusters of *M. tuberculosis* is shown.

did six isolates of *M. bovis* Bacillus Calmette-Guérin with very different laboratory passage histories. Together, these results confirm that sSNP profiles are stable genetic markers and demonstrate that sSNP genotyping is useful for epidemiologic strain tracing.

sSNP genotyping of *M. tuberculosis* complex organisms: The five sibling species of the *M. tuberculosis* complex are closely related (BAESS 1979; NOLTE and METCHOCK 1995), but overall estimates of phylogenetic relationships among them have not been clarified because appropriate molecular methods have not been available. To address this problem, we determined the sSNP genotypes of an additional 128 *M. tuberculosis* complex strains recovered from many host species and geographic localities, including organisms classified by conventional diagnostic methods (NOLTE and METCHOCK 1995) such as *M. africanum* ($n = 8$ isolates), *M. microti* ($n = 4$), *M. canettii* ($n = 1$), *M. bovis* ($n = 103$), and *M. tuberculosis* s.s. ($n = 12$; see Table 2 in the supplementary material at http://www.genetics.org/supplemental). sSNP analysis demonstrated that isolates of three of these non-*M. tuberculosis* s.s. sibling species are genetically allied with isolates of principal genetic group 1 *M. tuberculosis* (Figure 5). Many isolates of *M. africanum* are very closely related to isolates of *M. tuberculosis* s.s. with few copies of IS*6110*. Isolates of *M. microti* are allied with isolates of *M. tuberculosis* recovered from infected seals from several continents (COUSINS *et al.* 1993; LIEBANA *et al.* 1996; ZUMARRAGA *et al.* 1999) and with several isolates of *M. africanum* from humans (Figure 5). The results are consistent with limited data suggesting that tuberculosis in seals is caused by genetically distinct isolates of *M. tuberculosis* (COUSINS *et al.* 1993; LIEBANA *et al.* 1996; ZUMARRAGA *et al.* 1999). The *M. microti*-*M. africanum*-seal bacillus cluster of isolates comprises a very distinct sSNP group (bootstrap value = 99) that forms a phylogenetic bridge between *M. tuberculosis* s.s. (human specialist) and *M. bovis* (infecting predominantly nonhuman host species).

One genetic group of *M. bovis* warrants special comment. Four isolates recovered from cattle in Malawi are closely allied to, but distinct from, the cluster containing virtually all *M. bovis* isolates from several continents and many host species (Figure 5). The Malawi isolates have a distinct spoligotype and IS*6110* profile compared to other *M. bovis* isolates (KREMER *et al.* 1999; data not

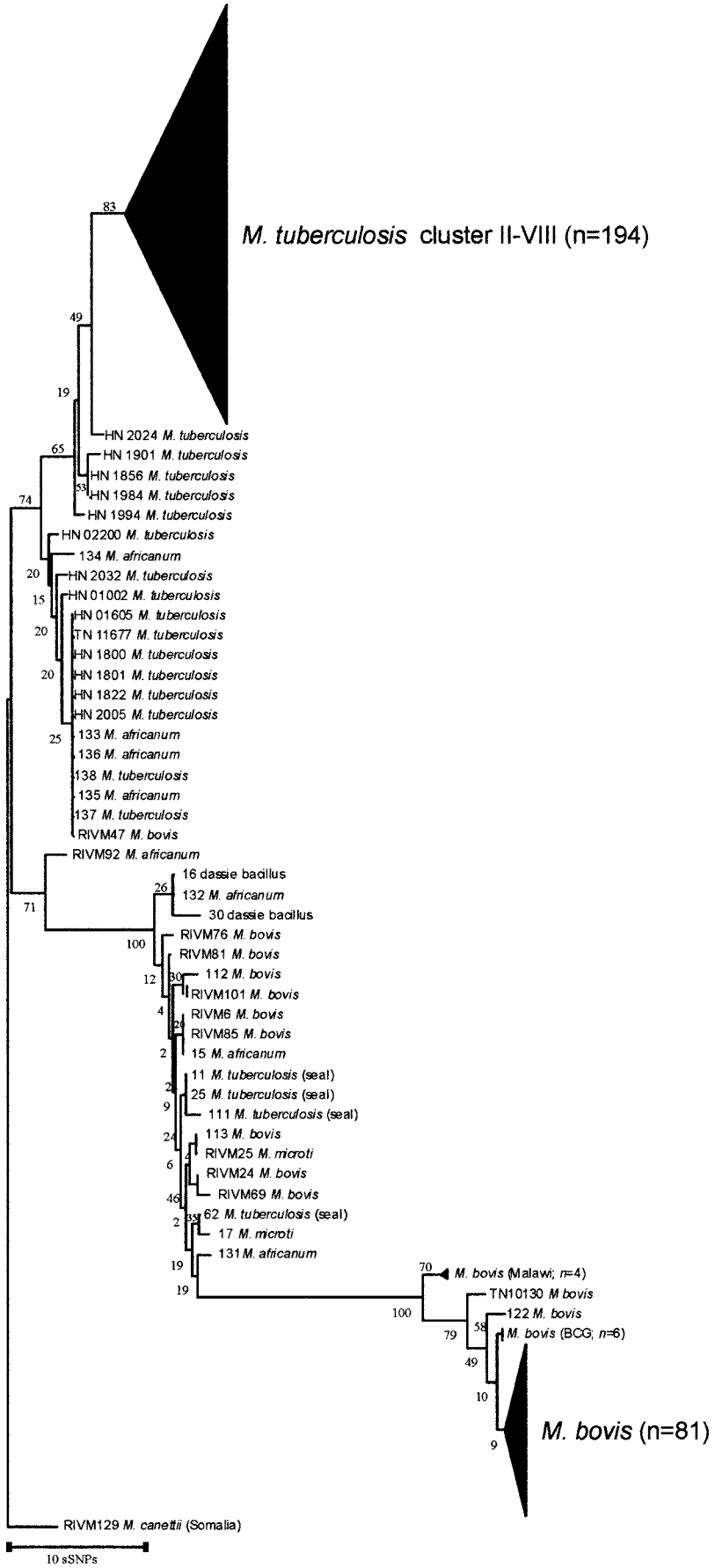


FIGURE 5.—Phylogenetic tree showing estimates of genetic relationships among 330 *M. tuberculosis* complex isolates, based on study of 90 sSNPs. The phylogenetic tree, generated using the neighbor-joining method with 1000 bootstrap replicates and distance calculated using the number of different sSNP loci (http://www.megasoftware.net/), was rooted with *M. canettii*, the most divergent member of the *M. tuberculosis* complex (VAN SOOLINGEN *et al.* 1997; KREMER *et al.* 1999; VAN EMBDEN *et al.* 2000; BROSCH *et al.* 2002). To more clearly show the evolutionary relationships occurring among members of the *M. tuberculosis* complex, certain branches of the phylogenetic tree were compressed: *M. tuberculosis* isolates assigned to cluster II–VIII (Figure 2) and *M. bovis* isolates showing little or no sSNPs divergence. *M. tuberculosis* and *M. bovis* are well differentiated from one another, and strains of *M. africanum*, *M. microti*, and *M. tuberculosis* isolated from seals are clearly located between these two members of the complex.

shown). Importantly, these isolates have the allele of the pyrazinamidase gene (*pncA*) commonly present in *M. tuberculosis* s.s. (Sreevatsan *et al.* 1997b) and the 1973-bp variant of *Rv3135* found in *M. bovis* isolates and many principal genetic group 1 organisms (Musser *et al.* 2000). The data are consistent with the hypothesis that these Malawi isolates represent an immediate relative of the great majority of extant *M. bovis* isolates causing disease in developed countries.

Inasmuch as sSNP genotyping provided new information about overall evolutionary relationships among four of the five *M. tuberculosis* complex organisms, we extended the investigation to *M. canettii*, the most divergent member of the complex (van Soolingen *et al.* 1997; Kremer *et al.* 1999; van Embden *et al.* 2000). The rooted tree revealed a deep branching of *M. canettii* (or a related last common ancestor) into two distinct lineages, one leading to *M. tuberculosis* s.s. and the second resulting in the *M. microti-M. africanum*-seal bacillus cluster (Figure 5). This latter cluster then gave rise to the large majority of organisms responsible for extant tuberculosis in animals in developed countries. In summary, sSNP genotyping resolved phylogenetic relationships among all five members of the *M. tuberculosis* complex.

DISCUSSION

Advantages of sSNP genotyping: sSNPs afford many advantages for analysis of phylogenetic relationships among microbial strains, especially closely related clonal organisms such as the *M. tuberculosis* complex. Most or all sSNPs are selectively neutral and hence minimally subject to convergence, a process that can obscure or distort evolutionary relationships (Kimura 1983). Binary data are obtained, which means that the information is readily amenable to storage, retrieval, analysis by personal computers equipped with simple software, and comparison between different laboratories and studies. Importantly, the considerable biomedical interest in human SNPs (Schork *et al.* 2000; Daly *et al.* 2001; Gut 2001; Johnson *et al.* 2001) means that microbial pathogen research will benefit extensively from the ongoing development and implementation of methods to index very large numbers of SNPs efficiently, inexpensively, and automatically (Kwok 2001). Because of the many advantages of using sSNP analysis for estimating genetic relationships among strains, and because of the public health and medical research need to classify strains into identical or closely similar genetic groups, it is likely that in the future *M. tuberculosis* complex strains will be routinely categorized on the basis of sSNP genotypes. This process will simplify population genetic analysis and epidemiologic investigations. Indeed, the cost of SNP genotyping is decreasing rapidly, suggesting that routine profiling of large samples of SNPs in many species of closely related pathogens will become feasible.

sSNP-based population genetic framework: One goal of bacterial population genetic research is to understand the relationships between genetic diversity, clonal lineages, and biomedically relevant phenotypes such as virulence, transmissibility, host specialization, and evolutionary success (Musser 1996). The estimates of overall chromosomal relationships for all strains provided by sSNP genotyping makes possible the mapping of traits onto the *M. tuberculosis* phylogenetic tree, and this now can be done for all isolates, heretofore impossible.

IS*6110* RFLP profiling (IS*6110* fingerprinting) is used extensively to categorize strains for epidemiologic studies and to infer relationships among strains (Small *et al.* 1993; van Embden *et al.* 1993; Bifani *et al.* 1996, 1999; Bishai *et al.* 1998; Yaganehdoost *et al.* 1999; Lockman *et al.* 2001). Strains with more than five copies of IS*6110* and identical RFLP profiles are considered to be clonally related by descent. This view is supported by our sSNP analysis, but the converse is not always true. That is, strains with very different IS*6110* copy numbers and RFLP patterns also may be clonally related by descent, as exemplified by strains typed as principal genetic group 1 (Figure 5). Importantly, sSNP genotyping clearly ruled out the idea that strains of *M. tuberculosis* s.s. with many IS*6110* copies ("high-copy strains") and few IS*6110* copies ("low-copy strains") are genetically distinct populations, thereby resolving a long-standing controversy (Fomukong *et al.* 1997; Sreevatsan *et al.* 1997a; Mazars *et al.* 2001).

sSNP analysis provided new insight into genetic relationships among *M. tuberculosis* complex organisms. We found that isolates of *M. tuberculosis* s.s. and *M. bovis* are well differentiated from one another and located in distinct branches of the sSNP-based phylogenetic tree. Moreover, rooting of the tree with *M. canettii* revealed that strains of *M. microti* and *M. tuberculosis* isolates recovered from seals with tuberculosis and some *M. africanum* isolates were placed phylogenetically between *M. tuberculosis* s.s. and most *M. bovis* isolates. Thus, the present genome-wide sSNPs study adds some important insights to evolutionary relationships recently described by other authors (Brosch *et al.* 2002; Mostowy *et al.* 2002). Taken together, these results reveal that repeatedly isolated clones of *M. bovis* causing disease in cattle and other animals in many developed countries and *M. tuberculosis* isolated worldwide are distinct, deeply branching genotypic complexes whose extant members did not emerge directly from one another in the recent past, as is generally thought to be the case (Diamond 1997; Stead 1997).

sSNP genotyping for rapid characterization of strains of pathogenic microbes: Many circumstances require rapid and unambiguous characterization of the genetic profile (chromosomal fingerprint) of microbial strains, including highly virulent pathogens. High-throughput sSNP genotyping is an attractive method for conducting analyses of this type because many hundreds to thou-

sands of SNPs can be indexed in less than a few days. It is reasonable to anticipate that in a putative outbreak situation this magnitude of sSNP genotyping would be required.

Our study was greatly assisted by the availability of genome sequence data from four strains that represent distinct phylogenetic lineages of the *M. tuberculosis* complex. Were this not the case, the insights obtained would be far more circumscribed and the conclusions less robust. For example, analysis of sSNPs identified solely by comparison of the genomes of strains H37Rv and CDC1551 failed to reveal the complexity of relationships among and between principal genetic group 1 organisms and isolates of the other *M. tuberculosis* complex members. sSNP-based studies of genetic relationships in other pathogens should take into consideration the benefits afforded by selection of sSNPs on the basis of comparison of multiple divergent chromosomes. In this regard, data obtained from the increasing number of comparative genome sequencing projects in many pathogens will benefit subsequent sSNP studies, and more so if strains from distinct phylogenetic lineages are sequenced.

In summary, sSNP genotyping provided many new insights into phylogenetic relationships among members of the *M. tuberculosis* complex, which are closely related and heretofore could not be unambiguously assigned to distinct lineages. sSNP genotyping of strains of pathogenic microbes is a powerful new strategy that is generally applicable and especially useful for indexing genetic relationships among closely related organisms such as *Yersinia pestis* (Achtman *et al.* 1999), *Bacillus anthracis* (Keim *et al.* 1997, 2000), Bordetella species (Musser *et al.* 1986), *M. leprae* (Clark-Curtiss and Walsh 1989), and *Plasmodium falciparum* (Rich *et al.* 1998; Volkman *et al.* 2001), agents that have considerable detrimental impact on human health.

We thank T. Bowland and A. Lee for technical assistance; N. Williams-Bouyer for assistance with large strain collections; J. Driscoll for providing spoligotype data for some of the isolates; S. D. Reid for assistance with phylogenetic analysis; and V. Deretic, N. P. Hoe, H. Ochman, S. D. Reid, and K. Virtaneva for critical review of the manuscript. This research was supported in part by National Institutes Health contract N01-AO-02738.

LITERATURE CITED

Achtman, M., K. Zurth, G. Morelli, G. Torrea, A. Guiyoule *et al.*, 1999 *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proc. Natl. Acad. Sci. USA **96:** 14043–14048.

Anh, D. D., M. W. Borgdorff, L. N. Van, N. T. Lan, T. van Gorkom *et al.*, 2000 *Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. Emerg. Infect. Dis. **6:** 302–305.

Baess, I., 1979 Deoxyribonucleic acid relatedness among species of slowly-growing mycobacteria. Acta Pathol. Microbiol. Scand. Sect. B **87:** 221–226.

Beggs, M. L., K. D. Eisenach and M. D. Cave, 2000 Mapping of IS6110 insertion sites in two epidemic strains of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **38:** 2923–2928.

Bifani, P. J., B. B. Plikaytis, V. Kapur, K. Stockbauer, X. Pan *et al.*, 1996 Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family. JAMA **275:** 452–457.

Bifani, P. J., B. Mathema, Z. Liu, S. L. Moghazeh, B. Shopsin *et al.*, 1999 Identification of a W variant outbreak of *Mycobacterium tuberculosis* via population-based molecular epidemiology. JAMA **282:** 2321–2327.

Bishai, W. R., N. M. Graham, S. Harrington, D. S. Pope, N. Hooper *et al.*, 1998 Molecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy. JAMA **280:** 1679–1684.

Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser *et al.*, 2002 A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc. Natl. Acad. Sci. USA **99:** 3684–3689.

Chan, M. Y., M. Borgdorff, C. W. Yip, P. E. de Haas, W. S. Wong *et al.*, 2001 Seventy percent of the *Mycobacterium tuberculosis* isolates in Hong Kong represent the Beijing genotype. Epidemiol. Infect. **127:** 169–171.

Clark-Curtiss, J. E., and G. P. Walsh, 1989 Conservation of genomic sequences among isolates of *Mycobacterium leprae*. J. Bacteriol. **171:** 4844–4851.

Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher *et al.*, 1998 Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393:** 537–544.

Cousins, D. V., S. N. Williams, R. Reuter, D. Forshaw, B. Chadwick *et al.*, 1993 Tuberculosis in wild seals and characterisation of the seal bacillus. Aust. Vet. J. **70:** 92–97.

Daly, M. J., J. D. Rioux, S. F. Schaffner, T. J. Hudson and E. S. Lander, 2001 High-resolution haplotype structure in the human genome. Nat. Genet. **29:** 229–232.

Diamond, J., 1997 *Guns, Germs and Steel: The Fates of Human Societies*. W. W. Norton, New York.

Feizabadi, M. M., I. D. Robertson, D. V. Cousins and D. J. Hampson, 1996 Genomic analysis of *Mycobacterium bovis* and other members of the *Mycobacterium tuberculosis* complex by isoenzyme analysis and pulsed-field gel electrophoresis. J. Clin. Microbiol. **34:** 1136–1142.

Fomukong, N., M. Beggs, H. el Hajj, G. Templeton, K. Eisenach *et al.*, 1997 Differences in the prevalence of IS6110 insertion sites in *Mycobacterium tuberculosis* strains: low and high copy number of IS*6110*. Tuber. Lung Dis. **78:** 109–116.

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:** 1189–1196.

Groenen, P. M., A. E. Bunschoten, D. van Soolingen and J. D. van Embden, 1993 Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*: application for strain differentiation by a novel typing method. Mol. Microbiol. **10:** 1057–1065.

Gut, I. G., 2001 Automation in genotyping of single nucleotide polymorphisms. Hum. Mutat. **17:** 475–492.

Johnson, G. C., L. Esposito, B. J. Barratt, A. N. Smith, J. Heward *et al.*, 2001 Haplotype tagging for the identification of common disease genes. Nat. Genet. **29:** 233–237.

Kapur, V., T. S. Whittam and J. M. Musser, 1994 Is *Mycobacterium tuberculosis* 15,000 years old? J. Infect. Dis. **170:** 1348–1349.

Keim, P., A. Kalif, J. Schupp, K. Hill, S. E. Travis *et al.*, 1997 Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. J. Bacteriol. **179:** 818–824.

Keim, P., L. B. Price, A. M. Klevytska, K. L. Smith, J. M. Schupp *et al.*, 2000 Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J. Bacteriol. **182:** 2928–2936.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.

Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans *et al.*, 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.

Kurepina, N. E., S. Sreevatsan, B. B. Plikaytis, P. J. Bifani, N. D. Connell *et al.*, 1998 Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS*6110* elements

in *Mycobacterium tuberculosis*: non-random integration in the dnaA-dnaN region. Tuber. Lung Dis. **79:** 31–42.

Kwok, P. Y., 2001 Methods for genotyping single nucleotide polymorphisms. Annu. Rev. Genomics Hum. Genet. **2:** 235–258.

Liebana, E., A. Aranaz, B. Francis and D. Cousins, 1996 Assessment of genetic markers for species differentiation within the *Mycobacterium tuberculosis* complex. J. Clin. Microbiol. **34:** 933–938.

Lockman, S., J. D. Sheppard, C. R. Braden, M. J. Mwasekaga, C. L. Woodley *et al.*, 2001 Molecular and conventional epidemiology of *Mycobacterium tuberculosis* in Botswana: a population-based prospective study of 301 pulmonary tuberculosis patients. J. Clin. Microbiol. **39:** 1042–1047.

Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell *et al.*, 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.

Mathema, B., P. J. Bifani, J. Driscoll, L. Steinlein, N. Kurepina *et al.*, 2002 Identification and evolution of an IS6110 low-copy-number *Mycobacterium tuberculosis* cluster. J. Infect. Dis. **185:** 641–649.

Mazars, E., S. Lesjean, A. L. Banuls, M. Gilbert, V. Vincent *et al.*, 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.

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.

Musser, J. M., 1996 Molecular population genetic analysis of emerged bacterial pathogens: selected insights. Emerg. Infect. Dis. **2:** 1–17.

Musser, J. M., E. L. Hewlett, M. S. Peppler and R. K. Selander, 1986 Genetic diversity and relationships in populations of Bordetella spp. J. Bacteriol. **166:** 230–237.

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.

Nolte, F. S., and B. Metchock, 1995 Mycobacterium, pp. 400–437 in *Manual of Clinical Microbiology*, edited by P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover and L. H. Yolken. American Society for Microbiology Press, Washington, DC.

Ochman, H., J. G. Lawrence and E. A. Groisman, 2000 Lateral gene transfer and the nature of bacterial innovation. Nature **405:** 299–304.

Park, Y. K., G. H. Bai and S. J. Kim, 2000 Restriction fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolated from countries in the western Pacific region. J. Clin. Microbiol. **38:** 191–197.

Ramaswamy, S., and J. M. Musser, 1998 Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. Tuber. Lung Dis. **79:** 3–29.

Reid, S. D., N. P. Hoe, L. M. Smoot and J. M. Musser, 2001 Group A streptococcus: allelic variation, population genetics, and host-pathogen interactions. J. Clin. Invest. **107:** 393–399.

Rich, S. M., M. C. Licht, R. R. Hudson and F. J. Ayala, 1998 Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA **95:** 4425–4430.

Ross, B. C., K. Raios, K. Jackson and B. Dwyer, 1992 Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. J. Clin. Microbiol. **30:** 942–946.

Schork, N. J., D. Fallin and J. S. Lanchbury, 2000 Single nucleotide polymorphisms and the future of genetic epidemiology. Clin. Genet. **58:** 250–264.

Small, P. M., R. W. Shafer, P. C. Hopewell, S. P. Singh, M. J. Murphy *et al.*, 1993 Exogenous reinfection with multidrug-resistant *Mycobacterium tuberculosis* in patients with advanced HIV infection. N. Engl. J. Med. **328:** 1137–1144.

Soini, H., X. Pan, A. Amin, E. A. Graviss, A. Siddiqui *et al.*, 2000 Characterization of *Mycobacterium tuberculosis* isolates from patients in Houston, Texas, by spoligotyping. J. Clin. Microbiol. **38:** 669–676.

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.

Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth *et al.*, 1997a Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. USA **94:** 9869–9874.

Sreevatsan, S., X. Pan, Y. Zhang, B. N. Kreiswirth and J. M. Musser, 1997b Mutations associated with pyrazinamide resistance in *pncA* of *Mycobacterium tuberculosis* complex organisms. Antimicrob. Agents Chemother. **41:** 636–640.

Stead, W. W., 1997 The origin and erratic global spread of tuberculosis. How the past explains the present and is the key to the future. Clin. Chest Med. **18:** 65–77.

Supply, P., S. Lesjean, E. Savine, K. Kremer, D. van Soolingen *et al.*, 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.

Valway, S. E., M. P. Sanchez, T. F. Shinnick, I. Orme, T. Agerton *et al.*, 1998 An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. N. Engl. J. Med. **338:** 633–639.

van Embden, J. D., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach *et al.*, 1993 Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. **31:** 406–409.

van Embden, J. D., T. van Gorkom, K. Kremer, R. Jansen, B. A. van Der Zeijst *et al.*, 2000 Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. J. Bacteriol. **182:** 2393–2401.

van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore *et al.*, 1995 Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J. Clin. Microbiol. **33:** 3234–3238.

van Soolingen, D., T. Hoogenboezem, P. E. de Haas, P. W. Hermans, M. A. Koedam *et al.*, 1997 A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. Int. J. Syst. Bacteriol. **47:** 1236–1245.

Volkman, S. K., A. E. Barry, E. J. Lyons, K. M. Nielsen, S. M. Thomas *et al.*, 2001 Recent origin of *Plasmodium falciparum* from a single progenitor. Science **293:** 482–484.

World Health Organization, 1998 *World Health Report*. World Health Organization, Geneva.

Yaganehdoost, A., E. A. Graviss, M. W. Ross, G. J. Adams, S. Ramaswamy *et al.*, 1999 Complex transmission dynamics of clonally related virulent *Mycobacterium tuberculosis* associated with barhopping by predominantly human immunodeficiency virus-positive gay men. J. Infect. Dis. **180:** 1245–1251.

Zumarraga, M. J., A. Bernardelli, R. Bastida, V. Quse, J. Loureiro *et al.*, 1999 Molecular characterization of mycobacteria isolated from seals. Microbiology **145:** 2519–2526.

Communicating editor: Y.-X. Fu