

Silent Nucleotide Polymorphisms and a Phylogeny for *Mycobacterium tuberculosis*

Lucy Baker,* Tim Brown,* Martin C. Maiden,† and Francis Drobniowski*

Much remains unknown of the phylogeny and evolution of *Mycobacterium tuberculosis*, an organism that kills 2 million people annually. Using a population-based approach that analyzes multiple loci around the chromosome, we demonstrate that neutral genetic variation in genes associated with antimicrobial drug resistance has sufficient variation to construct a robust phylogenetic tree for *M. tuberculosis*. The data describe a clonal population with a minimum of four distinct *M. tuberculosis* lineages, closely related to *M. bovis*. The lineages are strongly geographically associated. Nucleotide substitutions proven to cause drug resistance are distributed throughout the tree, whereas nonsynonymous base substitutions unrelated to drug resistance have a restricted distribution. The phylogenetic structure is concordant with all the previously described genotypic and phenotypic groupings of *M. tuberculosis* strains and provides a unifying framework for both epidemiologic and evolutionary analysis of *M. tuberculosis* populations.

Mycobacterium tuberculosis has caused tuberculosis (TB) in humans for thousands of years (1,2), and the World Health Organization (WHO) estimates that one third of the global population is infected with *M. tuberculosis* (3); however, the bacterium has remained an enigma. The global resurgence of TB highlights the need for an improved understanding of its epidemiology and its evolutionary biologic features. Recent advances in molecular characterization of *M. tuberculosis* isolates, which index variation in insertion sequences (4) and repetitive genomic elements (5,6), have elucidated clusters of identical and closely related strain families (7–9). These findings have provided insights into regional (10) and national (11) epidemiologic features. However, these techniques may be less suited to global population and evolutionary analyses, and integrating information obtained from different

approaches is complex (12). Genomic comparisons have identified genetic variation for population screening; however, these analyses are limited to those sites that vary between the compared genomes and are potentially misleading (13–15). Nucleotide sequences provide robust, portable, and comparable data for studying population variation. The mutational processes that generate this variation are understood, and sequence data have been successfully used in the study of bacterial epidemiology, population biology, and evolution (16). The complete genome sequences (15–18) provide access to all regions of the chromosome and facilitate such studies. However, high-throughput gene sequencing of structural genes (19) and host immune system protein targets (20) in *M. tuberculosis* isolates indicated low levels of sequence diversity. Although extensive genomic sequencing was performed in both studies, comparable sequence data were obtained on a limited number of highly selected isolates.

We used an unbiased population approach to analyze genetically silent nucleotide sequence variation for seven unlinked loci distributed around the chromosome. The loci chosen were genes associated with antimicrobial drug resistance that have been reported to possess >95% of all sequence variation observed in 26 structural genes studied (19), which includes >90% of synonymous nucleotide substitutions, i.e., nucleotide substitutions that do not affect the translated amino acid. In a population sample of 316 U.K. clinical isolates, silent single nucleotide polymorphisms (sSNPs) resolve an unambiguous phylogeny and provide a unifying framework for epidemiologic, population, and evolutionary analyses.

Methods

Bacterial Strains

The 316 *M. tuberculosis* clinical isolates were identified in England and Wales from January 1, 1998, through

*Health Protection Agency, London, United Kingdom; and †University of Oxford, Oxford, United Kingdom

December 31, 1998, and included all the viable clinical isolates (n = 216) resistant to one or both of the firstline antituberculous drugs (rifampicin and isoniazid) and 100 randomly chosen fully susceptible isolates. One *M. tuberculosis* H37Rv isolate, four *M. bovis* isolates, two *M. africanum* type I isolates, and one *M. microti* isolate were included for comparison. *M. tuberculosis* complex isolates were identified with a combination of microscopic and macroscopic appearance, growth characteristics, biochemical analysis, and DNA hybridization (21). Clinical and epidemiologic information were obtained from laboratory records at the HPA Mycobacterium Reference Unit, London, UK, the U.K. Mycobacterial Resistance Network database (MYCOBNET), and the 1998 national TB survey (22). Duplicate isolates were excluded. Drug susceptibility was determined by the resistance ratio method (2). Strains were characterized by IS6110 restriction fragment length polymorphism (RFLP) and spoligotyping (4,5)

Amplifying and Sequencing Target Gene Loci

The nucleotide sequences were obtained for the following seven gene loci: *rpoB*, *katG*, *oxyR*, *ahpC*, *pncA*, *rpsL*, and *gyrA*. These gene loci are associated with drug resistance, but without antimicrobial drug selection pressure, they would be regarded as housekeeping genes. Primers and amplification conditions for polymerase chain reaction (PCR) are shown in Table 1. Products were purified by precipitation with 20% polyethylene glycol–2.5 mol/L NaCl and sequenced from both DNA strands by using internal nested primers (Table 2) and BigDye Ready Reaction Mix (ABI, Warrington, UK) according to the manufacturer's instructions. Unincorporated dye terminators were removed by precipitation with 96% ethanol–0.115 mol/L sodium acetate, pH 4.6. The reaction products were separated and detected with an ABI prism 3700.

Genetic Analysis

Sequences were assembled with the STADEN suite of computer programs (23). The sequences were compared, and isolates with identical sequences were assigned the same allele number. For each gene, the DNA sequence was translated in frame, and each nucleotide polymorphism characterizing the allele was classified as synonymous or genetically silent, nonsynonymous, or intergenic.

For each isolate, the concatenated sequences from the coding region of all seven gene loci were reduced to a 36-nt sequence motif, constituting a synonymous sequence profile (SSP), and distinct SSPs were assigned a synonymous sequence type (SST). Phylogenetic analysis of the SST motifs was performed with the MEGA (24) and PHYLIP software packages (25): 225 isolates were sequenced at all loci. The sequencing data from the initial

225 isolates demonstrated key polymorphisms that were lineage defining. The remaining 94 isolates were assigned a lineage based on polymorphisms at *katG*⁸⁷, *katG*⁶⁰⁹, *katG*¹³⁸⁸, *oxyR*³⁷, *oxyR*²⁸⁵, and *ahpC*⁻⁴⁶, and by the spoligotype deletion pattern. A lineage was only ascribed if all data points agreed (Table 3). The relationship between lineages, phenotypic and genotypic drug resistance, and country of birth was analyzed with chi-square and Fisher exact test.

Comparison with Outgroups

An in silico analysis of the seven gene loci was undertaken for two mycobacterial outgroups, *M. leprae* (26) and *M. marinum* with BLAST (Sanger Institute, Cambridge, UK; available from http://www.sanger.ac.uk/projects/M_marinum/). The complete gene sequences for each of the seven loci in *M. tuberculosis*, *M. bovis*, *M. leprae*, and *M. marinum* were aligned in frame by using Clustal-W. Two approaches were used. First, the aligned sequences for the coding regions of the seven gene loci were concatenated to produce a single sequence of 8.212 Kbp for each isolate. The concatenated sequences for fully susceptible examples of the *M. tuberculosis* SSTs were aligned to this. Second, SSPs were constructed for *M. leprae* and *M. marinum* by using the relevant aligned nucleotide for each of the previously identified variable synonymous sites in *M. tuberculosis* and *M. bovis*. For each approach, a phylogeny was constructed with the neighbor-joining tree method, and the results were compared.

TbD1-PCR Analysis

Three isolates per SST were selected for analysis. Each possessed, when possible, a different IS6110 RFLP or spoligotype pattern. This analysis was performed with the published method (13).

Results

Observed Genetic Diversity

The complete gene was sequenced for all but one locus, which provided 8,318 Kbp of nucleotide sequence data for each isolate. Across the seven loci, 115 variable sites were identified, of which 101 were within the coding region of the selected loci, and 36 were associated with genetically neutral base substitutions. The number of alleles per locus varied from 6 (*oxyR*) to 40 (*rpoB*). The proportion of variable sites present at each locus was low, 0.68% (*rpoB*) to 2.68% (*pncA*). Nonsynonymous base substitutions were more frequent than synonymous substitutions at almost all loci. The ratio of nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site (d_N/d_S ratio) varied from 0.109 to 0.848 in sensitive *M. tuberculosis* isolates and from 0.301 to 1.952

Table 1. Amplification primers

Gene/locus	Forward primer	Reverse primer	Product (bp)	Reaction conditions ^a	Cycles
<i>gyrA</i>	gyrA-ext F 5'-ACAGACACGACGTTGCCGCC-3'	gyrA-ext R 5'-GTCGATTTCCCTCAGCATCTCC-3'	435	95°C for 15 min ^b 95°C for 15 s 68°C for 30 s 72°C for 1 min 72°C for 5 min	} 30
<i>inhA</i>	mabA-ext F 5'-TCGTAGGGCGTCAATACACC-3'	mabA-ext R 5'-TCATTCGACCGAATTTGTTG-3'	605	94°C for 5 min 94°C for 30 s 60°C for 30 s 72°C for 30 s 72°C for 5 min	} 30
<i>katG</i>	katG-ext3F 5'-CGACGAAATGGGACAACAGT-3'	katG-ext3R 5'-TGCATGAGCATTATCCCGTA-3'	1,507	94°C for 5 min 94°C for 30 s 60°C for 30 s 72°C for 1 min 72°C for 7 min	} 30
	katG -ext5F 5'-TCGACTGTGCTGTTGGCGAGG-3'	katG-ext5R 5'-CTTCGCCGACGAGGTCGTGG-3'	1,531	95°C for 15 min ^b 95°C for 30 s 68°C for 30 s 72°C for 1 min 72°C for 7 min	
<i>oxyR-ahpC</i>	oxyR-ext F 5'-TCGAGCTGCGACGGTGCTGG-3'	oxyR-extR 5'-CTGCGGGTGATTGAGCTCAGG-3'	1,437	95°C for 15 min ^b 95°C for 30 s 72°C for 30 s 72°C for 1 min 72°C for 7 min	} 30
<i>pncA</i>	pncA-ext F 5''-AACCAAGGACTTCCACATCG-3'	pncA-extAR 5'-CAGAACTGCAGCATCATCG-3'	1,324	95°C for 15 min ^b 95°C for 30 s 64°C for 30 s 72°C for 1 min 72°C for 7 min	} 30
<i>rpoB</i>	rpoB-46F 5'-GGCCGTGGGCACCGCTCC-3'	rpoB 1868R 5'-CCAGCGGGGCTCGCTACG-3'	1,822	95°C for 15 min ^b 95°C for 15 s 65°C for 30 s 72°C for 3 min 72°C for 10 min	} 30
	rpoB 1711F 5'-GTGCCCTCGTCTGAGGTGGAC-3'	rpoB 3602R 5'-AAGACCGATGCGGAGTTCATCG-3'	1,891	95°C for 15 min ^b 95°C for 15 s 65°C for 30 s 72°C for 3 min 72°C for 10 min	
<i>rpsL</i>	rpsL-extF 5'-GGCCGACAAACAGAACGT-3'	rpsL-extR 5'-GTTCACTAACTGGGTGAC-3'	494	94°C for 5 min 94°C for 30 s 56°C for 30 s 72°C for 30 s 72°C for 5 min	} 30

^aA final PCR reaction volume of 25 µL was used that contained 2.5 µL of 10 x ammonium sulfate reaction buffer (Bioline, London, UK), 1.5 mmol/L magnesium chloride (Bioline); 200 µmol/L each of dATP, dTTP, dGTP, and dCTP; 300 nmol/L of each primer pair, 0.8 units of Taq DNA polymerase (Bioline), 1 µL (≈10 ng) of template DNA and sterile distilled water. Amplification was carried out in 0.2 ml thin-wall polymerase chain reaction tubes in a DNA Thermal Cycler 9600 (Applied Biosystems, Warrington, UK). Products were purified by precipitation with 20% polyethylene glycol detected with an ABI Prism 3700 or an ABI Prism 377 automated DNA sequencer (ABI, Warrington, UK).

^bReaction performed with Hotstar Taq and reaction buffer (Qiagen, Crawley, UK).

overall, which implied that resistance to antituberculous medication is indeed the selective force at most loci. Five variable sites were unique to *M. bovis*, of which four were associated with synonymous polymorphisms. A further variable site with a previously reported synonymous polymorphism (*katG*^{C609T}) (27) was identified in both *M. bovis* and *M. microti*, present in all four *M. bovis* isolates sequenced and the published *M. bovis* genome sequence (18).

Synonymous Sequence Types and Lineages

By disregarding nonsynonymous polymorphisms, i.e., those producing an amino acid change likely in response to diversifying or stabilizing selection, a subset of 37 neutral sSNPs at 36 sites was generated; one site possessed two different synonymous substitutions. These substitutions occurred in 35 unique combinations, which we term synonymous sequence types (SSTs); each was assigned an arbitrary number (Figure 2). The variation in the SSTs conformed to the clonal model for bacterial population

Table 2. Sequencing primers^a

Locus/PCR product	Forward primers	Reverse primers
<i>gyrA</i>	<i>gyrA</i> -1F 5'-CAGCTACATCGACTATG-3'	<i>gyrA</i> -1R 5'-GGGCTTCGGTGTACCTCAT-3'
<i>inhA</i> promoter	<i>mabA</i> -1F 5'-AGAAAGGGATCCGTCATGGT-3'	<i>mabA</i> -1R 5'-GTCACATTCGACGCCAAAC-3'
<i>katG</i> 3F-3R	<i>katG</i> -1F 5'-ACGCGGGGTCTGACAAAT-3'	<i>katG</i> -1R 5'-GACAAGGCGAACCTGCTTAC-3'
	<i>katG</i> -2F 5'-GTAAGCAGGTTGCGCTTGT-3'	<i>katG</i> -2R 5'-TCGGGATTGACTGCTCACA-3'
	<i>katG</i> -3F 5'-ATCTCTCCAGGGTGCGAAT-3'	<i>katG</i> -3R 5'-GAGTGGGAGCTGACAGAG-3'
<i>katG</i> 5F-5R	<i>katG</i> -4F 5'-AGAGGTCAGTGGCCAGCAT-3'	<i>katG</i> -4R 5'-AGATGGGGCTGATCTACGTG-3'
	<i>katG</i> -5F 5'-GCTGTTTCGACGTCGTTTCAT-3'	<i>katG</i> -5R 5'-ACTACGGGCCGCTGTTTATC-3'
	<i>katG</i> -6R 5'-ACACTTCGCGATCACATCC-3'	<i>katG</i> -6R 5'-ACACTTCGCGATCACATCC-3'
<i>oxyR-ahpC</i>	<i>oxyR</i> -1 5'-CTGGCCAGGTAAGACGACC-3'	<i>oxyR</i> -2 5'-CAGACGCTCGATGCTGCC-3'
	<i>oxyR</i> -7 5'-TCATATCGAGAATGCTTGCGG-3'	<i>oxyR</i> -4 5'-TGCTTGCGCTCCACCTTGG-3'
	<i>oxyR</i> -6 5'-TGATGTCTTTGGCGTACTCGG-3'	<i>oxyR</i> -6 5'-CAATGACGAGTTCGAGGACC-3'
<i>pncA</i>	<i>pncA</i> -P1 5'-GCTGGTCATGTTGCGATCG-3'	<i>pncA</i> -R 5'-CGATGAAGGTGTCGTAGAAGC-3'
	<i>pncA</i> -F 5'-AACCAAGGACTTCCACATCG-3'	<i>pncA</i> -2F 5'-ATACCGACCACATCGACCTC-3'
<i>rpoB</i> -46-1868	<i>rpoB</i> -41F 5'-GTGGGCACCGCTCCTCTAAGG-3'	<i>rpoB</i> 509R 5'-TGACCACCACACGCTCGGTCC-3'
	<i>rpoB</i> 331F 5'-CGTTTCGACGATGTCAAGGCA-3'	<i>rpoB</i> 975R 5'-GTCGACGACGTGATGGGCTCG-3'
	<i>rpoB</i> 783F 5'-CTGGAGAAGGACAACACCGTCC-3'	<i>rpoB</i> 2 5'-GCACGTCGCGGACCTCCAGCC-3'
	<i>rpoB</i> 1 5'-GGTCGGCATGTCGCGGATGGA-3'	<i>rpoB</i> 1845R 5'-CGCTACGGACCAGCGGCACC-3'
<i>rpoB</i> 1711-3602	<i>rpoB</i> 1725F 5'-GGTGGACTACATGGACGTCTC-3'	<i>rpoB</i> 2313R 5'-GTCGGAGATGTTCCGGGATGTCG-3'
	<i>rpoB</i> 2134F 5'-GAGATGGCGTGGGCAAGAAC-3'	<i>rpoB</i> 2770R 5'-TCTGGCCGATGTTCCATCCGTCG-3'
	<i>rpoB</i> 2600F 5'-AGCTGGTGCCTGTGTATGTGG-3'	<i>rpoB</i> 3213R 5'-GGCCTGCATGCCCCAGCACTCC-3'
	<i>rpoB</i> 3013F 5'-CCGTTCCGTACCCGGTCCAGC-3'	<i>rpoB</i> 3581R 5'-GAAGAAGTTGACGTCGAGCAC-3'
	<i>rpsL</i>	<i>rpsL</i> F 5'-ACGTGAAAGCGCCAAAGATAGA -3'

^aAll sequencing reactions were performed in 96-well plates (Abgene, UK) in a DNA Thermal Cycler 9600 (Applied Biosystems, Warrington, UK) by using the following thermocycling conditions: 30 cycles of denaturation at 96°C for 10 s, annealing at 50°C for 5 s, and extension at 60°C for 2 min.

structure (28), and the maximum parsimony method generated a phylogenetic tree with no homoplasies, i.e., the lack of independent occurrence of a polymorphism in more than one branch of the tree (Figure 1A, 2). Each branch corresponds to a unique combination of sSNPs. The phylogeny was robust, whether constructed with or without outgroup SSTs generated from the *M. leprae* and *M. marinum* genome sequences. The analysis identified four prominent *M. tuberculosis* lineages (numbered I to IV); the *M. bovis* isolates formed an additional lineage. The lineages are defined by distinct combinations of sSNPs (Figure 2). Virtually all of the nodes in the tree are occupied; internal nodes tend to be represented more frequently in the isolate population. Although a number of evolutionary scenarios are possible, the most likely explanation for this observation is that the sSNP variation arose recently.

Within the *M. tuberculosis* complex, the sSNPs clearly distinguish *M. tuberculosis* from *M. bovis* and *M. microti*, with the *M. microti* SST forming a node on the *M. bovis* lineage. The *M. africanum* type 1 isolates sequenced could not be distinguished from *M. tuberculosis* because they share SST-1.

SST Phylogeny and Population Subdivisions

A variety of approaches have been used previously to subdivide *M. tuberculosis* strains into definable groups. These include assignments based on two nonsynonymous polymorphisms in *katG* and *gyrA* (19); the presence or

absence of a TB-specific genomic region of difference, *TbD1* (13); variation within the genomic direct repeat region demonstrated by spoligotyping; and strain families defined by highly conserved DNA fingerprint patterns obtained by RFLP of the insertion element *IS6110*. Each technique defines a limited number of distinct subdivisions, which although different, overlap when techniques are compared. The sSNP phylogenetic tree was congruent with all of the previously described subdivisions (Figure 1).

katG and *gyrA* Polymorphisms

M. tuberculosis isolates can be divided into three groups (1–3) based on two apparently unselected nonsynonymous SNPs, *katG*^{G1388T}, and *gyrA*^{G284C} (19). Group 1 is defined by the combination of *katG*^{1388T} and *gyrA*^{284C}, group 2 by *katG*^{1388G} and *gyrA*^{284C}, and group 3 by *katG*^{1388G} and *gyrA*^{284G}. We characterized the *katG-gyrA* polymorphisms in all isolates. The *katG*^{G1388T} polymorphism cosegregated in all cases with the synonymous base substitution *rpoB*^{T3243C} found in lineages I, III, IV, and *M. bovis*, whereas the *gyrA*^{G284C} polymorphism subdivided SST-2 in lineage II. Group 1 isolates can therefore be subdivided into three prominent *M. tuberculosis* lineages (I, III, and IV) made up of 21 SSTs and *M. bovis*. Group 2 and 3 *M. tuberculosis* isolates are subdivisions of lineage II. Group 2 isolates can be further subdivided into 10 SSTs, whereas group 3 isolates are confined to a subdivision of SST 2 (Figure 1B). When the SST of the clinical isolate

CDC1551 was identified *in silico* by analysis of the relevant sequences (www.tigr.org), it shared the same sequence type, SST 2, as the other isolate for which a complete genome is available, H37Rv. Although these two isolates are distinguished by numerous other synonymous and nonsynonymous polymorphisms (15), these organisms are closely related, which has implications for genetic variation based on comparing the two complete genome sequences (14,29).

TbD1 Region of Difference

The presence or absence of DNA regions, identified by genomic comparisons of *M. tuberculosis* H37Rv and *M. bovis* BCG, can be used to distinguish the closely related members of the *M. tuberculosis* complex. However, only two groups of *M. tuberculosis* isolates have been described with this approach, defined by the presence or absence of the TB-specific region, TbD1 (13). TbD1 PCR analysis was performed on three epidemiologically unrelated isolates from each SST, when available. We found that the TbD1 region was present in all 13 SST types constituting lineage IV and all *M. bovis*, *M. microti*, and *M. africanum* type I isolates, but the region was absent from all other lineages and SSTs (Figure 1C). This finding implies that the TbD1 deletion occurred before both the *katG*^{G 1388 T} and the *rpoB*^{T 3243 C} mutations. Although SST-1 appears the least differentiated SST with the maximum parsimony method and sSNP data alone, when taken together with the TbD1 data, the ancestors of lineage IV are likely to have diverged from *M. africanum* type I, *M. microti*, and *M. bovis* before differentiation of the other *M. tuberculosis* lineages.

DNA Fingerprinting Techniques

Molecular epidemiologic analyses of TB populations use a combination of typing techniques, most commonly IS6110 RFLP analysis and spoligotyping. Spoligotyping has been used to distinguish members of the *M. tuberculosis* complex (5,30), and together with IS6110 RFLP, has been used to describe various *M. tuberculosis* strain families including Beijing (7), Haarlem (8), Africa (8), Delhi (9), East Africa-India (EA-I), and Latin America-Mediterranean (12). Both techniques were used to type all 316 isolates, producing 234 IS6110 RFLP patterns and 263 spoligotyping patterns; 157 isolates were assigned to strain families.

All isolates within each family were confined to a single lineage (Figure 1D). Furthermore, each lineage was defined by a distinct pattern of spacer deletions (Figure 1E). The signature spoligotype spacer deletion pattern for lineage II (lack of probe hybridization at spacers 33–36) concurs with that previously noted in group 2 and group 3 *M. tuberculosis* isolates (13,31). Lineage I was

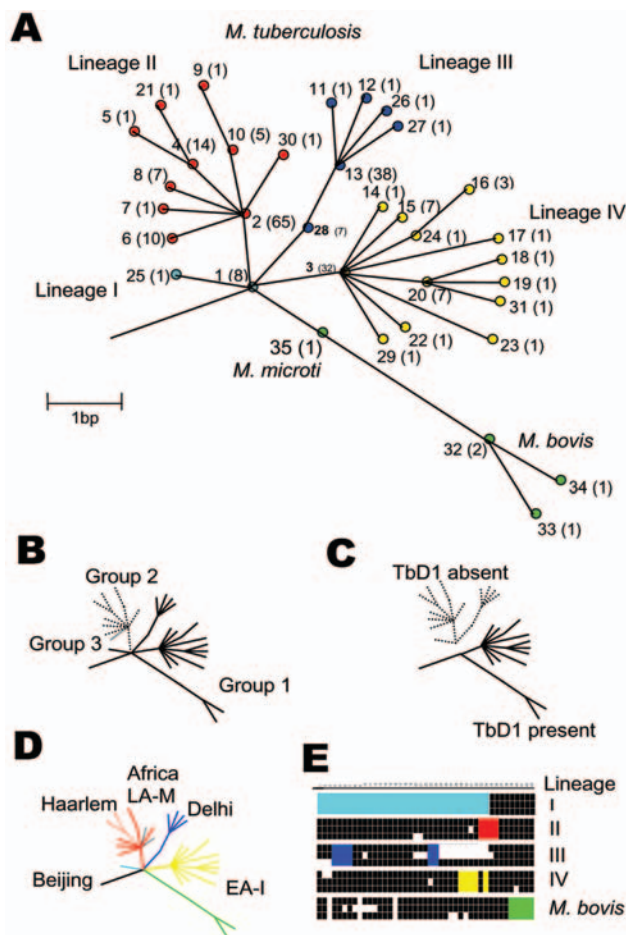


Figure 1. Unifying phylogeny for *Mycobacterium tuberculosis*. A) Maximum parsimony tree of *M. tuberculosis* and *M. bovis* based on 37 silent single-nucleotide polymorphisms in 225 isolates. Synonymous sequence types (SST) are marked 1–35. The frequency of each SST is marked in parentheses. The nodes of the major lineages are highlighted: lineage I (cyan), lineage II (red), lineage III (blue), lineage IV (yellow), and *M. bovis* (green). The colors correspond to those in Figure 2. Note both *M. africanum* Type I isolates sequenced were SST 1. B) Schematic representation of the genetic groups 1, 2, and 3 defined by the *katG-gyrA* scheme. C) Schematic representation of the presence or absence of the tuberculosis specific region of difference, TbD1. D) Schematic representation of the strain families Beijing, Haarlem, Africa, Delhi, East Africa-India (EA-I), and Latin America-Mediterranean (LA-M), previously described by IS6110 restriction fragment length polymorphism typing and spoligotyping, demonstrating concordance with the phylogenetic tree. E) Spoligotyping patterns for representative isolates of each lineage demonstrating lack of probe hybridization at spacers 1–34 in lineage I, 33–36 in lineage II, 4–7 and 23–24 in lineage III, 29–32 and 34 in lineage IV, and 39–43 in *M. bovis*.

defined by the signature spoligotype of the Beijing family (absence of spacers 1–34), and lineage IV by the signature spoligotype of the EA-I family. The remaining strain

Gene	oxyR				ahpC				katG						pncA				rpoB							
Nucleotide position	37	204	263	285	18	480	87	399	555	609	891	942	1563	1599	1821	6	195	201	300	318	327	357	702	1701	2646	3243
Codon number					6	160	29	133	185	203	297	314	521	533	607	2	65	67	100	106	109	119	194	567	882	1081
Amino acid					Ile	Glu	Pro	Asn	Phe	Thr	Gly	Thr	Leu	Pro	Glu	Arg	Ser	Ser	Thr	Ser	Asp	Val	Val	Glu	Gly	Ala
Lineage I	SST 1	C	A	C	G	T	G	C	C	C	C	C	C	G	G	G	C	G	C	G	C	C	C	G	T	C
Lineage II	25																									
Lineage II	2																									T
Lineage II	4																									T
Lineage II	5																									T
Lineage II	21																									G
Lineage II	10																									T
Lineage II	9																									T
Lineage II	6																									T
Lineage II	6																									T
Lineage III	28																									G
Lineage III	13																									G
Lineage III	26																									G
Lineage III	27																									G
Lineage IV	3																									T
Lineage IV	20																									T
Lineage IV	18																									T
Lineage IV	19																									T
Lineage IV	31																									T
Lineage IV	24																									T
Lineage IV	16																									T
Lineage IV	15																									T
Lineage IV	17																									T
<i>M. bovis</i>	32																									A
<i>M. bovis</i>	33																									A
<i>M. bovis</i>	34																									A
<i>M. microti</i>	35																									T

Figure 2. Abbreviated figure demonstrating the relationship between the silent single nucleotide polymorphisms, the synonymous sequence type, and the major lineages (abbreviated to 26 SSTs and 26 variable sites). •, represents invariant base with respect to SST 1. For full reproduction of this figure, please see <http://www.cdc.gov/ncidod/eid/vol10no9/04-0046-G2.htm>.

families were confined to lineage subbranches, including the Haarlem family, confined to SSTs 4, 5, and 6 of lineage II, and the Delhi family, confined to lineage III, almost exclusively within SSTs 11, 12, 13, 26, and 27. The relationship between SST, lineage, spoligotype pattern, and IS6110 RFLP pattern are shown in more detail in online Figure available at <http://www.cdc.gov/ncidod/eid/vol10no9/04-0046-G2.htm>.

Seventy-one isolates shared identical RFLP and spoligotype patterns with one or more isolate, grouped in 23 clusters, of which 10 possessed RFLP patterns containing five or fewer IS6110 copies. Cluster sizes ranged from two to seven isolates, with a median cluster size of two. Isolate clusters were present in all lineages, but isolates with low copy number were confined to lineage II and IV. Three of the low copy number clusters, each characterized by a single IS6110 band and distinct spoligotype pattern, were subdivided by SST, whereas among high copy number clusters, all isolates within a cluster possessed the same SST. To prevent introducing a selection bias, no correction was made for strain transmission.

Polymorphisms and Antimicrobial Drug Resistance to Lineage

Having demonstrated a robust phylogeny for *M. tuberculosis* on the basis of neutral genetic variation, we annotated the tree with the nsSNPs. Most nsSNPs were rare within the isolate collection examined, many of which were represented uniquely. However, a number of nsSNPs known to confer drug resistance occurred frequently, including *rpoB*^{C 1367 G}, *rpoB*^{C 1367 T}, *rpoB*^{C 1351 T} (32,33), *katG*^{A 944 C} (34,35), *rpsL*^{A 128 G} (36), and *inhA*^{C-15 T} promoter mutation (35). These polymorphisms were distributed throughout the phylogeny, which implies that they arose independently on many occasions, presumably in response to the positive selection imposed by antimicrobial drug use. In contrast, an intergenic SNP, *oxyR-ahpC*^{G -46 A} (37), associated with, but not proven to cause, isoniazid resistance, occurred exclusively in lineage III and was present in all isolates within the lineage, which implies that this SNP may have arisen under neutral selection.

Although mutations conferring drug resistance were present in all lineages, the proportion of resistant and susceptible isolates varied between lineages. When antimicro-

Table 3. Characteristic features of the four major lineages of *Mycobacterium tuberculosis* with respect to *M. bovis*

Species/lineage	TbD1 region of difference	Silent single nucleotide polymorphism						Nonsynonymous base substitutions		Spoligotype signature spacer deletion
		<i>rpoB</i>	<i>rpoB</i>	<i>katG</i>	<i>katG</i>	<i>oxyR</i>	<i>oxyR</i>	<i>katG</i>	<i>ahpC</i>	
		3243 T	2646 G	87 A	609 T	285 A	37 T	1388 G	-46 A	
<i>M. tuberculosis</i> Lineage I	-	-	-	-	-	-	-	-	-	1-34
<i>M. tuberculosis</i> Lineage II	-	+	-	-	-	-	-	+	-	33-36
<i>M. tuberculosis</i> Lineage III	-	-	+	-	-	-	-	-	+	4-8, 23-24
<i>M. tuberculosis</i> Lineage IV	+	-	-	-	-	-	+	-	-	29-32 and 34
<i>M. bovis</i>	+	-	-	+	+	+	-	-	-	39-43

bial drug susceptibility data were used, lineage III was positively associated with isoniazid resistance (51/62, $p = 0.002$), Lineages I and III were positively associated with streptomycin resistance (12/20, $p = 0.0004$ and 24/62, $p = 0.004$ respectively), and lineage IV was positively associated with fully susceptible isolates (30/62, $p = 0.002$). Rifampicin resistance was not associated with any lineage.

Genotypic analysis of phenotypically resistant isolates showed that among isoniazid-resistant isolates, the resistance-conferring mutation *katG*^{A944C} was positively associated with lineage III (odds ratio [OR] 2.44, $p = 0.016$) and the *inhA*^{C-15T} promoter mutation positively associated with lineage IV (OR 3.28, $p = 0.006$). Among streptomycin-resistant isolates, the resistance-conferring mutation *rpsL*^{A128G} was positively associated with lineage I (OR 7.83, $p = 0.012$) and negatively associated with lineage II (OR 0.32, $p = 0.036$). No genotype-lineage association was identified for rifampicin resistance. Different lineages have significant differences in their antimicrobial drug susceptibility to certain antimicrobial agents, perhaps demonstrating the effect of genomic environment on the probability of mutational events conferring resistance to these antimicrobial drugs.

Lineage to Country of Birth

Drug resistance in countries with low TB incidence has been associated with foreign-born migrants (11); 44% of TB cases in England and Wales occur in the indigenous population (22). Information about country of birth was available for 225 (71%) of the patients; they represented

45 countries. Foreign-born patients had a median residency of 4 years in the United Kingdom. There was no significant difference between patients infected with susceptible or drug-resistant *M. tuberculosis* with respect to patient country or continent of birth. However, highly significant associations existed between continent of birth and lineage (Table 4). Lineages I, II, and III were significantly associated with southeastern Asia, Europe, and the Indian subcontinent, respectively. Lineage IV, in contrast, was globally distributed but had a negative association with Europe. This finding provides strong evidence for geographic structuring in *M. tuberculosis* populations.

Discussion

Synonymous nucleotide polymorphisms reflect neutral genomic variation, which remains informative, even in genes that have recently experienced positive selection attributable to introducing antimicrobial agents. By sequencing widely at multiple gene loci around the chromosome in a population sample of *M. tuberculosis* isolates, selection bias is avoided and all neutral variation within the sequenced regions will be identified. The indexed variation is highly unlikely to arise by convergence, which provides a robust base for constructing a phylogenetic tree. Our data support the belief that *M. tuberculosis* is a strictly clonal organism, with no evidence of lateral gene transfer.

Individual sSNPs are confined to clonally related organisms and accumulated by subsequent generations. Each of the lineages defined here can be defined on the basis of

Table 4. Relationship between *Mycobacterium tuberculosis* lineage and continent of birth of patient^a

Lineage	n	Europe				Africa				Indian subcontinent				Southeast Asia			
		Not		OR	p value	Not		OR	p value	Not		OR	p value	Not		OR	p value
		Eur	Eur			Afr	Afr			ISC	ISC			SEA	SEA		
I	17	5	12	0.86	0.997	1	16	0.17	0.079	1	16	0.12	0.034	10	7	28.8	<0.00001
II	118	59	59	6.4	<0.00001	33	85	1.34	0.473	19	99	0.2	<0.00001	4	114	0.21	0.006
III	50	3	47	0.1	0.00001	9	41	0.58	0.245	37	13	11.31	<0.00001	0	50	0.0	0.009
IV	40	5	35	0.25	0.005	15	25	2.04	0.08	15	25	1.36	0.514	5	35	1.66	0.166
Total	225	72	153			58	167			72	153			19	206		

^aOR, odds ratio; Eur, Europe; Afr, Africa; ISC, Indian subcontinent; SEA, southeast Asia.

single sSNPs, yet the resultant maximum parsimony tree provides a robust and unifying phylogeny for *M. tuberculosis*. The documented population diversity is relatively recent, demonstrated by SSTs that represent almost all of the phylogenetic nodes. In contrast, *M. tuberculosis* and *M. bovis* separated from a common ancestor more distantly, which reinforces the evidence that *M. tuberculosis* could not have arisen from *M. bovis*, as previously thought.

Although we have sequenced a small number of *M. tuberculosis* complex isolates, our data support the evolutionary scenario described by Brosch et al. (13). *M. microti* is a subdivision of the *M. bovis* lineage, diverging after the separation of *M. bovis* from its common ancestor with *M. tuberculosis*. *M. africanum* type I, which cannot be distinguished from *M. tuberculosis* on the basis of neutral variation in the genes sequenced in this study, can be distinguished from *M. tuberculosis* SST-1 by the presence of the TbD1 region. Taken together, the sequence data support divergence of *M. africanum* type I from a common ancestor with *M. tuberculosis* before the subsequent divergence of *M. microti* and *M. bovis*. Analyzing silent nucleotide polymorphisms in *gyrB*, which has been used to distinguish members of the *M. tuberculosis* complex (38,39), would provide further neutral sequence variation to support the evolutionary scenario.

M. tuberculosis isolates in England and Wales represent four clearly defined lineages. Although the isolates were all obtained from patients residing in the United Kingdom, the patients represented 45 countries of birth, from four continents. The population is not globally representative; for example, few patients originated from the Americas. Nevertheless, the strong geographic structuring of the *M. tuberculosis* population is striking. *M. tuberculosis* is an obligate human pathogen, with a delay between initial infection and the development of clinical disease (often up to 5 years) and long periods of latency between disease control and subsequent clinical reactivation. The evolution and global dissemination of *M. tuberculosis* are by definition associated with the activities of its human host. Although foreign-born patients may have been infected with *M. tuberculosis* in the United Kingdom, the short median residency in the United Kingdom and the lack of strain clustering support the hypothesis that these are imported strains reflecting *M. tuberculosis* populations in the patient's country of birth. Clonal expansion of geographically restricted, genetically distinct lineages presumably reflects the previously geographically limited human population movements, with higher rates of transmission within, rather than between, geographic regions. No single *M. tuberculosis* lineage dominates in African-born patients. As in human populations, Africa appears to be a melting pot for genetic diversity. This fact may reflect the dissemination of *M. tuberculosis* by ancient human migra-

tion and trade routes but could be further elucidated by analysis of unselected isolates obtained in Africa.

Unlike lineages I, II and III, lineage IV is globally distributed, with no discernible geographic association. Not only is it the only lineage possessing the TbD1 region of difference, in common with *M. bovis*, *M. microti*, and *M. africanum* type I (13), but a large proportion of the isolates possess only a single IS6110 copy, and isolates from the lineage are negatively associated with antimicrobial drug resistance. These data suggest that *M. tuberculosis* isolates from lineage IV are more closely related to the common ancestor of the *M. tuberculosis* complex, unexposed to antimicrobial selection pressure, and provide evidence to support the hypothesis that *M. tuberculosis* isolates possessing a single IS6110 copy may be ancestral (40,41). In contrast, isolates possessing a high number of IS6110 copies are present in all four *M. tuberculosis* lineages, which reflects independent IS6110 transposition events in different parts of the phylogeny.

Geographic structuring of a clonal population will result in genetically and phenotypically distinct *M. tuberculosis* populations, which may explain, in part, the geographically variable response to vaccination with *M. bovis* BCG, or striking differences in clinical features, such as the predominance of extrapulmonary disease in patients originally from the Indian subcontinent. This finding may also have implications for the successful development of new TB vaccines. Nucleotide substitutions arising under neutral, positive, and negative pressure will all become fixed, inherited by all clonal descendants. Analyzing mutations that confer antimicrobial drug resistance provides an insight into this evolutionary process. By definition, resistance-conferring mutations are associated with phenotypic resistance absolutely. The genes involved all encode essential metabolic functions, restricting nonsynonymous nucleotide substitutions. The data demonstrate that the most frequently reported resistance-conferring mutations are present in all lineages, which implies that they have arisen independently on multiple separate occasions; however, phenotypically antimicrobial drug resistance is significantly associated with lineage. The significantly greater proportion of phenotypically resistant isolates with the *katG*³¹⁵ mutation in lineage III and the observation that the mutation is not present in all isolates within clonally related subbranches of the tree confirms the relatively recent influence of antimicrobial drug selection pressure. This finding implies that isolates within the lineage may be biochemically more susceptible to acquiring the same resistance conferring mutation. Rifampicin resistance and multidrug-resistant TB isolates were unrelated to lineage, although the numbers were relatively small (52 phenotypically rifampicin-resistant isolates, of which 46 were multidrug resistant).

We have shown that the sequenced *M. tuberculosis* strains, CDC1551 and H37Rv, are closely related and come from the same major lineage (lineage II, SST-2). This relation has implications for sSNP analyses based on comparison of only these genomes, collapsing subbranches and skewing any resultant phylogenetic tree (14,29). Although this tendency can be reduced slightly by comparing genomes that are genetically more divergent, the lack of horizontal gene transfer in a clonal population means that variation in other branches of the phylogeny will not be revealed. In fact, only four of the sSNPs described here can be resolved by genome comparison of the four completed genome sequences *M. tuberculosis* H37Rv (17), CDC1551 (15), strain 210 (www.tigr.org), and *M. bovis* (18). None were used in the SNP analysis performed by Gutacker et al. (14). By sequencing widely at multiple gene loci around the chromosome, we have identified all the indexable genetically neutral variation (sSNPs) within the sequenced regions. Although Sreevatsan et al. used a similar approach in their study of 26 structural genes, which included regions of all seven genes sequenced in this study (19), the isolates were selected from a large collection of *M. tuberculosis* strains in part on diversity in IS6110 RFLP (introducing a bias towards high copy number strains), and the number of isolates sequenced at each locus varied, with no defined minimum dataset. We identified a similar level of genomic sequence diversity, but by using an unbiased population approach, we have shown phylogenetically significant neutral variation.

The phylogeny described here is unambiguous and can be defined with a limited number of sSNPs. These could easily be identified with rapid screening techniques. Simultaneous identification of nsSNPs associated with antimicrobial drug resistance would provide data valuable for clinical, epidemiologic, and evolutionary purposes in a single, cost-effective, and highly portable format that is amenable to electronic database comparisons (16).

Acknowledgments

We thank the research and reference staff at the HPA Mycobacterium Reference unit for assistance with IS6110 RFLP analysis and spoligotyping, the HPA Communicable Disease Surveillance Center for additional epidemiologic data, and the research staff at the Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, UK, for help with DNA sequencing and analysis.

L.V.B. is a British Lung Foundation Research Fellow. M.C.J.M. is a Wellcome Trust Senior Research Fellow. This study was funded by the British Lung Foundation.

Dr. Baker was a British Lung Foundation fellow at the U.K. HPA National Mycobacterium Reference Unit, London. She is

currently consultant respiratory physician at University Hospital, Lewisham, London, with a particular interest in respiratory infections including tuberculosis and infections affecting persons with cystic fibrosis.

References

1. Salo WL, Aufderheide AC, Buikstra J, Holcomb TA. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proc Natl Acad Sci U S A*. 1994;91:2091-4.
2. Nerlich AG, Haas CJ, Zink A, Szeimies U, Hagedorn HG. Molecular evidence for tuberculosis in an ancient Egyptian mummy. *Lancet*. 1997;350:1404.
3. Drobniewski FA, Pablos-Mendez A, Raviglione MC. *Seminars in Respiratory and Critical Care Medicine*. 1997;18:419-29.
4. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol*. 1993;31:406-9.
5. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol*. 1997;35:907-14.
6. Frothingham R, Meeker-O'Connell WA. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology*. 1998;144:1189-96.
7. van Soolingen D, Qian L, de Haas PE, Douglas JT, Traore H, Portaels F, et al. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. *J Clin Microbiol*. 1995;33:3234-8.
8. Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PW, Martin C, et al. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. *J Clin Microbiol*. 1999;37:2607-18.
9. Bhanu N, van Soolingen D, van Embden J, Dar L, Pandey R, Seth P. Predominance of a novel *Mycobacterium tuberculosis* genotype in the Delhi region of India. *Tuberculosis (Edinb)*. 2002;82:105.
10. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N Engl J Med*. 1994;330:1703-9.
11. Lambregts-van Weezenbeek CS, Jansen HM, Veen J, Nagelkerke NJ, Sebek MM, van Soolingen D. Origin and management of primary and acquired drug-resistant tuberculosis in the Netherlands: the truth behind the rates. *Int J Tuberc Lung Dis*. 1998;2:296-302.
12. Sola C, Filliol I, Legrand E, Mokrousov I, Rastogi N. *Mycobacterium tuberculosis* phylogeny reconstruction based on combined numerical analysis with IS1081, IS6110, VNTR, and DR-based spoligotyping suggests the existence of two new phylogeographical clades. *J Mol Evol*. 2001;53:680-9.
13. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A*. 2002;99:3684-9.
14. Gutacker MM, Smoot JC, Migliaccio CA, Ricklefs SM, Hua S, Cousins DV, et al. Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms. Resolution of genetic relationships among closely related microbial strains. *Genetics*. 2002;162:1533-43.
15. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, et al. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol*. 2002;184:5479-90.

16. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A*. 1998;95:3140–5.
17. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. 1998;393:537–44.
18. Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, et al. The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci U S A*. 2003;100:7877–82.
19. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, et al. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A*. 1997;94:9869–74.
20. Musser JM, Amin A, Ramaswamy S. Negligible genetic diversity of *Mycobacterium tuberculosis* host immune system protein targets: evidence of limited selective pressure. *Genetics*. 2000;155:7–16.
21. Collins C, Grange J, Yates MD. Tuberculosis, bacteriology, organisation and practice (2nd Edition). Oxford: Butterworth-Heinemann; 1997.
22. Rose AM, Watson JM, Graham C, Nunn AJ, Drobniowski F, Ormerod LP, et al. Tuberculosis at the end of the 20th century in England and Wales: results of a national survey in 1998. *Thorax*. 2001;56:173–9.
23. Staden R. The Staden sequence analysis package. *Mol Biotechnol*. 1996;5:233–41.
24. Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics*. 2001;17:1244–5.
25. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol*. 1981;17:368–76.
26. Cole S, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. *Nature*. 2001;409:1007–11.
27. Frothingham R, Strickland PL, Bretzel G, Ramaswamy S, Musser JM, Williams DL. Phenotypic and genotypic characterization of *Mycobacterium africanum* isolates from West Africa. *J Clin Microbiol*. 1999;37:1921–6.
28. Maynard-Smith J, Smith N H, O'Rourke M, Spratt BG. How clonal are bacteria? *Proc Natl Acad Sci U S A*. 1993;90:4384–8.
29. Alland D, Whittam TS, Murray MB, Cave MD, Hazbon MH, Dix K, et al. Modeling bacterial evolution with comparative-genome-based marker systems: application to *Mycobacterium tuberculosis* evolution and pathogenesis. *J Bacteriol* 2003;185:3392–9.
30. van Soolingen D, van der Zanden AG, de Haas PE, Noordhoek GT, Kiers A, Foudraïne NA, et al. Diagnosis of *Mycobacterium microti* infections among humans by using novel genetic markers. *J Clin Microbiol*. 1998;36:1840–5.
31. Soini H, Pan X, Amin A, Graviss EA, Siddiqui A, Musser JM. Characterization of *Mycobacterium tuberculosis* isolates from patients in Houston, Texas, by spoligotyping. *J Clin Microbiol*. 2000;38:669–76.
32. Telenti A, Imboden P, Marchesi F, Lowrie D, Cole S, Colston MJ, et al. Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. *Lancet*. 1993;341:647–50.
33. Miller LP, Crawford JT, Shinnick TM. The *rpoB* gene of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*. 1994;38:805–11.
34. Wengenack NL, Uhl JR, St Amand AL, Tomlinson AJ, Benson LM, Naylor S, et al. Recombinant *Mycobacterium tuberculosis* KatG(S315T) is a competent catalase-peroxidase with reduced activity toward isoniazid. *J Infect Dis*. 1997;176:722–7.
35. Musser JM, Kapur V, Williams DL, Kreiswirth BN, van Soolingen D, van Embden JD. Characterization of the catalase-peroxidase gene (*katG*) and *inhA* locus in isoniazid-resistant and -susceptible strains of *Mycobacterium tuberculosis* by automated DNA sequencing: restricted array of mutations associated with drug resistance. *J Infect Dis*. 1996;173:196–202.
36. Nair J, Rouse DA, Bai GH, Morris SL. The *rpsL* gene and streptomycin resistance in single and multiple drug-resistant strains of *Mycobacterium tuberculosis*. *Mol Microbiol*. 1993;10:521–7.
37. Sreevatsan S, Pan X, Zhang Y, Deretic V, Musser JM. Analysis of the *oxyR-ahpC* region in isoniazid-resistant and -susceptible *Mycobacterium tuberculosis* complex organisms recovered from diseased humans and animals in diverse localities. *Antimicrob Agents Chemother*. 1997;41:600–6.
38. Kasai H, Ezaki T, Harayama S. Differentiation of phylogenetically related slowly growing mycobacteria by their *gyrB* sequences. *J Clin Microbiol*. 2000;38:301–8.
39. Niemann S, Harmsen D, Rusch-Gerdes S, Richter E. Differentiation of clinical *Mycobacterium tuberculosis* complex isolates by *gyrB* DNA sequence polymorphism analysis. *J Clin Microbiol*. 2000;38:3231–4.
40. Fomukong NG, Tang TH, al Maamary S, Ibrahim WA, Ramayah S, Yates M, et al. Insertion sequence typing of *Mycobacterium tuberculosis*: characterization of a widespread subtype with a single copy of IS6110. *Tuber Lung Dis*. 1994;75:435–40.
41. Fomukong N, Beggs M, el Hajj H, Templeton G, Eisenach K, Cave MD. Differences in the prevalence of IS6110 insertion sites in *Mycobacterium tuberculosis* strains: low and high copy number of IS6110. *Tuber Lung Dis*. 1998;78:109–16.

Address for correspondence: Francis Drobniowski, HPA Mycobacterium Reference Unit, Dept of Microbiology, Guy's, King's and St Thomas' School of Medicine, King's College Hospital, London, SE22 8QF, UK; fax: +20734666477; email: francis.drobniowski@kcl.ac.uk

OPPORTUNITIES FOR PEER REVIEWERS

The editors of *Emerging Infectious Diseases* seek to increase the roster of reviewers for manuscripts submitted by authors all over the world for publication in the journal. If you are interested in reviewing articles on emerging infectious disease topics, please e-mail your name, address, curriculum vitae, and areas of expertise to eideditor@cdc.gov

At *Emerging Infectious Diseases*, we always request reviewers' consent before sending manuscripts, limit review requests to three or four per year, and allow 2–4 weeks for completion of reviews. We consider reviewers invaluable in the process of selecting and publishing high-quality scientific articles and acknowledge their contributions in the journal once a year.

Even though it brings no financial compensation, participation in the peer-review process is not without rewards. Manuscript review provides scientists at all stages of their career opportunities for professional growth by familiarizing them with research trends and the latest work in the field of infectious diseases and by improving their own skills for presenting scientific information through constructive criticism of those of their peers. To view the spectrum of articles we publish, information for authors, and our extensive style guide, visit the journal web site at www.cdc.gov/eid.

For more information on participating in the peer-review process of *Emerging Infectious Diseases*, e-mail eideditor@cdc.gov or call the journal office at 404-371-5329.