

Automated High-Throughput Genotyping for Study of Global Epidemiology of *Mycobacterium tuberculosis* Based on Mycobacterial Interspersed Repetitive Units

PHILIP SUPPLY,^{1*} SARAH LESJEAN,¹ EVGUENI SAVINE,¹ KRISTIN KREMER,²
DICK VAN SOOLINGEN,² AND CAMILLE LOCHT¹

Laboratoire des Mécanismes Moléculaires de la Pathogénèse Bactérienne, INSERM U447, Institut Pasteur de Lille, F-59019 Lille Cedex, France,¹ and Diagnostic Laboratory for Infectious Diseases and Perinatal Screening, National Institute for Public Health and the Environment, 3720 Bilthoven, The Netherlands²

Received 1 March 2001/Returned for modification 23 June 2001/Accepted 3 July 2001

Large-scale genotyping of *Mycobacterium tuberculosis* is especially challenging, as the current typing methods are labor-intensive and the results are difficult to compare among laboratories. Here, automated typing based on variable-number tandem repeats (VNTRs) of genetic elements named mycobacterial interspersed repetitive units (MIRUs) in 12 mammalian minisatellite-like loci of *M. tuberculosis* is presented. This system combines analysis of multiplex PCRs on a fluorescence-based DNA analyzer with computerized automation of the genotyping. Analysis of a blinded reference set of 90 strains from 38 countries (K. Kremer et al., J. Clin. Microbiol. 37:2607–2618, 1999) demonstrated that it is 100% reproducible, sensitive, and specific for *M. tuberculosis* complex isolates, a performance that has not been achieved by any other typing method tested in the same conditions. MIRU-VNTRs can be used for analysis of the global genetic diversity of *M. tuberculosis* complex strains at different levels of evolutionary divergence. To fully exploit the portability of this typing system, a website was set up for the analysis of *M. tuberculosis* MIRU-VNTR genotypes via the Internet. This opens the way for global epidemiological surveillance of tuberculosis and should lead to novel insights into the evolutionary and population genetics of this major pathogen.

Infectious diseases, in particular tuberculosis, represent perhaps more than ever a global threat for human health. *Mycobacterium tuberculosis* causes annually about 10 million new cases of active tuberculosis and 3 million deaths. Moreover, it is estimated that over one-third of the world population have latent tuberculosis infection, which represents a huge reservoir for the disease. Exponential development of travel extends this threat worldwide (8, 10). Because of these factors, efficient disease control can be achieved only by international epidemiological surveillance systems able to accurately monitor epidemic trends at a global level.

For this purpose, large-scale genotyping of *M. tuberculosis* is a highly appropriate tool. Genotyping allows one to identify outbreaks that remained undetected by traditional epidemiological surveillance alone and may thus be used as a basis, and not only as a help, for conventional epidemiological surveillance (1). Moreover, it can give unique insights into the international dissemination dynamics of *M. tuberculosis* by the comparison of isolates from widespread geographic areas and allows one to analyze evolutionary changes of pathogen populations. Several molecular-typing studies recently demonstrated the existence of clonally related *M. tuberculosis* families, which disseminated locally or globally (e.g., references 1, 2, 12, and 22 to 24). However, in many of these studies relatively limited numbers of strains were analyzed and they pro-

vided data that cannot be cross-compared due to the inherent drawbacks of the current “gold standard” typing method, IS6110 restriction fragment length polymorphism (RFLP). This method is labor-intensive; other drawbacks include the difficulty of reproducing results and comparing them among different laboratories. This renders the comparison of data from multicenter studies problematic, especially when large numbers of fingerprints are involved (12). Unfortunately, other methods, such as multilocus sequence typing, appropriate for large multicenter studies of many other bacteria (13), are not applicable to *M. tuberculosis*, since the gene sequence polymorphism among different strains of this species is very limited (12, 17).

Here, we propose a solution to these problems by the use of high-throughput PCR analysis of *M. tuberculosis* genomic loci containing variable-number tandem repeat (VNTR) sequences. Using this approach, strains can be typed by a numerical code corresponding to the numbers of VNTRs in 12 different loci containing novel genetic elements named mycobacterial interspersed repetitive units (MIRUs) (18, 19). These genetic markers provide nonambiguous data which are highly portable between different laboratories (7, 14, 19) and which are therefore highly suitable for global epidemiological studies. Analysis of *M. tuberculosis* isolates collected in France indicated that they have a discrimination power close to that of IS6110 RFLP and that their stability is adequate to track outbreak episodes. In addition, unlike IS6110 RFLP, they are well suited for population genetics analyses, for they unambiguously reveal the variability of independent genetic loci, a requisite for linkage disequilibrium analysis (14). In this study, we report the incorporation into the MIRU-VNTR approach of

* Corresponding author. Mailing address: Laboratoire des Mécanismes Moléculaires de la Pathogénèse Bactérienne, INSERM U447, Institut Pasteur de Lille, 1, rue du Prof. Calmette, F-59019 Lille Cedex, France. Phone: (33) 320.87.11.54. Fax: (33) 320.87.11.58. E-mail: Philip.Supply@pasteur-lille.fr.

TABLE 1. Conditions for multiplex PCRs of the 12 MIRU-VNTR loci

Polymorphism ^a	Multiplex PCR mixture	MIRU locus	MIRU-VNTR length (bp) ^b	Primer sequence (5'-3') (labeling)
High	A	4	77	GCGCGAGAGCCCCGAAGTGC (FAM) GCGCAGCAGAAACGTCAGC
		26	51	TAGGTCTACCGTCGAAATCTGTGAC CATAGGCGACCAGGCGAATAG (HEX)
		40	54	GGGTTGCTGGATGACAACGTGT (NED) GGGTGATCTCGGGAAATCAGATA
High	B	10	53	GTTCTTGACCAACTGCAGTCGTCC GCCACCTTGGTGATCAGCTACCT (FAM)
		16	53	TCGGTGATCGGGTCCAGTCCAAGTA CCCGTCGTGCAGCCCTGGTAC (HEX)
		31	53	ACTGATTGGCTTCATACGGCTTTA GTGCCGACGTGGTCTTGAT (NED)
Low	C	2	53	TGGACTTGCAGCAATGGACCAACT TACTCGGACGCCGGCTCAAAAT (FAM)
		23	53	CTGTCGATGGCCGCAACAAAACG (HEX) AGCTCAACGGGTTTCGCCCTTTTGTC
		39	53	CGCATCGACAAACTGGAGCCAAAC CGGAAACGTCTACGCCCCACACAT (NED)
Low	D	20	77	TCGGAGAGATGCCCTTCGAGTTAG (FAM) GGAGACCGCGACCAGGTACTTGTA
		24	54	CGACCAAGATGTGCAGGAATACAT GGGCGAGTTGAGTCCACAGAA (HEX)
		27	53	TCGAAAGCCTCTGCGTGCCAGTAA GCGATGTGAGCGTGCCACTCAA (NED)

^a Based on the allelic diversity of the MIRU loci observed in *M. tuberculosis* (14).

^b Loci 10, 23, and 24 contain 55-, 50-, and 52-bp MIRU-VNTRs respectively, in first position of the repeat array whereas locus 4 in clinical isolates contains an additional invariable MIRU of 53 bp in the terminal position of the repeat array (19).

technologies used for high-throughput human VNTR-based genotyping. We analyzed the relevance and reliability of this high-capacity system for the analysis of the global genetic diversity of *M. tuberculosis* at different levels of evolutionary divergence by testing a blinded reference set of 90 strains of the *M. tuberculosis* complex from 38 countries.

MATERIALS AND METHODS

Strains and genomic DNA. The automated MIRU-VNTR typing system was developed using heat-treated *M. tuberculosis* colonies and purified DNA from isolates from Paris hospitals described previously (9, 21) and from transmissions, relapses, and laboratory cross-contaminations from the collection of the Centre de Référence des Mycobactéries (Institut Pasteur, Paris, France) described previously (14). The mycobacterial genomic DNA samples used in the blinded typing experiment were described previously (12). Briefly, genomic DNA was isolated from 90 *M. tuberculosis* complex strains originating from 38 countries throughout the world and from 10 non-*M. tuberculosis* complex strains. A set of 31 duplicate *M. tuberculosis* complex samples was also included. These samples were provided as blinded dried aliquots by the National Institute of Public Health and the Environment (Bilthoven, The Netherlands) and decoded after analysis.

Multiplex PCRs. The multiplex PCR mixtures were prepared as follows, using 96-well plates and the HotStartTaq DNA polymerase kit (Qiagen, Hilden, Germany). Five nanograms of DNA was added to a final volume of 50 μ l containing 0.2 μ l of DNA polymerase (1 U), 10 μ l of Q solution (from the above-mentioned kit), 0.2 mM (each) dATP, dCTP, dGTP, and dTTP (Pharmacia, Uppsala, Sweden), 5 μ l of PCR buffer, 0.4 μ M (each) primer, and 3.0, 2.0, 2.5, and 1.5 mM MgCl₂ for mixtures A to D, respectively (Table 1). The oligonucleotides used in the PCR corresponded to the flanking regions of the polymorphic MIRU-VNTR loci identified in the *M. tuberculosis* H37Rv genome (19). For each multiplex mixture, one primer of each oligonucleotide pair was tagged with a different fluorescent dye.

The thermocycler programs for the four multiplex reactions were identical. The PCRs were carried out using a PCR Express cyclor (Hybaid, Ashford, Great Britain) starting with a denaturing step of 15 min at 95°C, followed by 40 cycles

of 1 min at 94°C, 1 min at 59°C, and 1 min 30 s at 72°C. The reactions were terminated by incubation for 10 min at 72°C. Negative controls consisted of the PCR performed on reaction mixtures lacking mycobacterial DNA.

Automated MIRU-VNTR analysis. For each multiplex PCR, 1 μ l of 20-fold-diluted PCR products was added to a loading buffer containing 1.25 μ l of deionized formamide (Amresco, Solon, Ohio), 0.5 μ l of blue dextran-EDTA (PE Applied Biosystems, Courtaboeuf, France), and 0.25 μ l of Genescan 2500 Rox-labeled size standard (PE Applied Biosystems) using a Multimek automated 96-channel pipettor (Beckman, Gagny, France). Before being loaded, the samples were denatured at 95°C for 2 min and then kept on ice. The samples were subjected to electrophoresis using a 96-well ABI 377 automatic sequencer, Long Ranger Singel packs (Bio-Whittaker Molecular Applications, Rockland, Maine), and a 36-cm-long plate. After a 2-min prerun step, each lane was loaded with a 1- μ l portion of sample using an eight-channel syringe. One run of 2 min 30 s was made between the loading of even- and odd-numbered lanes. The samples were run for 5.5 h at 3,000 V, 2,400 scans per h, and 51°C using the D filter. The sizes of the PCR fragments were estimated using the GeneScan and Genotyper software packages (PE Applied Biosystem). The sizes of the labeled amplicons were analyzed using the local Southern size calling method, with an analysis range from 1,500 to 12,000 scans and a detection threshold set to a peak height of 50. The peak sizes from the 94- to 1,115-bp fragments of the internal size standards of the different lanes were matched using a sample run at the center of the gel as a reference. Custom macro programs (available upon request) associated with Genotyper were used for automated conversion of these data into MIRU-VNTR alleles. The assignment of the various alleles was based on the corresponding expected sizes of the PCR products for the 12 MIRU-VNTR loci, calculated from the data of Supply et al. (19). Size tolerances of ± 4 and ± 6 bp were fixed for PCR fragments below and above 500 bp, respectively, with an offset of -1 bp. The genotype data were then exported (via the Microsoft Excell 98 software) to File Maker Pro, version 4.0 (Clarix Corporation, Santa Clara, Calif.), database management software, which was used for automated retrieval of duplicate samples and repeated genotypes.

MIRU-VNTR genetic distance analysis. Neighbor-joining analysis was performed with PAUP (D. Swofford, PAUP—phylogenetic analysis using parsimony (and other methods), 4.0 beta version, Sinauer Associates, Inc., Sunderland, Mass.). A consensus tree based on 10,000 bootstrap replicates was generated

using the majority rule formula. The differences in MIRU-VNTR numbers were used to estimate the genetic distance. The data were treated as ordered character states, based on the assumption of a stepwise mutation mechanism for the MIRU-VNTR loci. This assumption is supported by the analysis of MIRU-VNTR changes in locus 4 in the *Mycobacterium bovis* BCG genealogy (19).

RESULTS

Automated fluorescence-based MIRU-VNTR genotyping.

We have recently identified 12 *M. tuberculosis* minisatellite-like loci that contain MIRU-VNTRs. In these loci, the sizes of the repeat units range from 51 to 77 bp (19). The numbers of variable units in each locus are calculated by measuring the sizes of the fragments amplified with specific primers hybridizing to the MIRU flanking regions (14, 19). First, we developed four different multiplex PCR assays in order to analyze four groups of three loci simultaneously. The groups were composed based on the compatibilities of the PCRs and on the ranking of polymorphisms of the different loci, as measured by the numbers and frequencies of their respective alleles (14), such that the loci with high to low polymorphism were grouped together (mixture A to mixture D; Table 1). For each multiplex PCR assay, one primer per oligonucleotide pair was tagged with a specific fluorescent dye (Fig. 1). The PCR cycles for the four multiplex reactions were identical, which allows them to be run simultaneously on a thermocycler.

The PCR fragments labeled with the three different fluorescent dyes from each multiplex PCR were combined with an internal size standard, labeled by a fourth dye, and then analyzed in single lanes on an automated DNA sequencer. The electrophoresis and analysis parameters were adapted in order to size amplicons from 200 to 1,000 bp, which covers the size ranges of all MIRU-VNTR alleles obtained for *M. tuberculosis* isolates studied to date (14, 19). The accuracy of the size determination was controlled by analyzing labeled PCR fragments amplified in multiplex PCR from *M. tuberculosis* H37Rv and other isolates, for which the exact sizes were determined by sequence analysis (19). Sizing was found to be reproducible, with within-run and between-run average precisions of ± 0.5 and ± 0.6 bp, respectively. Mean errors \pm standard deviations for the sizing accuracy were 1.1 ± 1.0 bp for fragments below 500 bp and 0.8 ± 4.6 bp for fragments from 500 to up to 971 bp. This ensures the unambiguous identification of MIRU-VNTR alleles, which differ in size by 51 to 77 bp.

Stutter peaks are common during the genotyping of short tandem repeat sequences and mostly reflect artifactual strand slippage of the polymerase during PCR. Such stutter peaks were also quite frequently observed for the PCRs of the various MIRU-VNTR loci. However, they could be easily diagnosed because, as for short tandem repeat sequences, they appeared as a ladder of much-lower-intensity peaks, corresponding to sizes of PCR fragments that lack one or more repeats (Fig. 1). We noticed two exceptions: *M. tuberculosis* isolates with as many as 10 and 11 MIRU-VNTRs in locus 23 (isolates 67 and 93; Table 2). In these cases, the stutter peaks from the multiplex reaction had heights comparable to those of the fragments corresponding to these alleles. When these samples were reanalyzed by PCR with only the primer pair corresponding to this locus, single high-intensity peaks were obtained, allowing us to correctly assign the alleles.

Genotyping of a blinded set of *M. tuberculosis* complex isolates from 38 countries. A synthetic interlaboratory study re-

cently compared all currently known typing methods for *M. tuberculosis* complex isolates with respect to their reproducibility, sensitivity, specificity, and discrimination (12). The methods were tested by using a blinded set of 131 samples, including 90 isolates from 38 countries, as well as duplicate samples and non-*M. tuberculosis* complex strains. We analyzed the same isolate set similarly blinded to evaluate the performances of the automated MIRU-VNTR system.

As shown in Table 2, the MIRU-VNTR typing method was 100% sensitive, as all *M. tuberculosis* complex strains were fully typeable, with the 12 alleles amplified in all cases, and 100% reproducible, as all blinded duplicate samples analyzed were correctly assigned. The method was also 100% specific, as none of the non-*M. tuberculosis* complex strains gave the patterns expected for the *M. tuberculosis* complex. For all non-*M. tuberculosis* complex strains, either the PCRs yielded no amplification product or the amplified fragments had sizes that did not correspond to any of the *M. tuberculosis* complex MIRU-VNTR alleles (not shown). Thus, the MIRU-VNTR system can be used to identify the *M. tuberculosis* complex strains.

The MIRU-VNTR typing detected 78 different patterns among the 90 *M. tuberculosis* complex isolates (Table 2 and Fig. 2). Eight new alleles not found in isolates analyzed before were detected (14, 19), which raises the potential number of different MIRU-VNTR combinations to more than 2×10^8 . In contrast to what is found for IS6110 RFLP typing, nearly all low-IS6110-copy-number ($n \leq 6$) isolates were discriminated by MIRU-VNTR typing (20 distinct patterns out of 21 isolates). The only exceptions were the Russian and French *M. bovis* BCG vaccine sister strains (see below). In addition, most of the high-copy-number isolates ($n > 6$) were also discriminated (58 patterns out of 69 isolates), except for the cases discussed below.

The genetic relationships among the 90 isolates were analyzed on the basis of the MIRU-VNTR genotypes using the neighbor-joining distance algorithm (Fig. 2). Beijing, Africa, and Haarlem genotype families, including *M. tuberculosis* isolates originating from various parts of the world, were identified in this collection (12). The isolates within these families possess distinct, but at least 50% similar, IS6110 RFLP patterns. Many of them also appear highly similar and often identical based on other typing markers. This is especially the case for the Beijing family. Most of the members previously assigned to these families were readily distinguished from each other by MIRU-VNTRs. Seven distinct patterns were obtained for the 8 Beijing isolates, seven were obtained for the 8 Africa isolates, and nine were obtained for the 13 Haarlem isolates. In addition, most of them (all of the Beijing family) were recognized as closely related by analysis of MIRU-VNTR genetic distances. Exceptions were isolates 8, 58, 87, and 123 (Haarlem family) and 72 and 40 (Africa family). Furthermore, an additional isolate from South Korea (111) was included in the Beijing group by MIRU-VNTR analysis, which is consistent with its IS6110 RFLP and spoligotype, which display motifs similar to those typical of the Beijing group (12). Two additional isolates (29 and 77) were included in two clusters with identical MIRU-VNTRs containing five Haarlem strains and one Haarlem strain, respectively. This clustering was also consistent with their respective IS6110 RFLPs and spoligotypes, which also display motifs typical of the previously defined

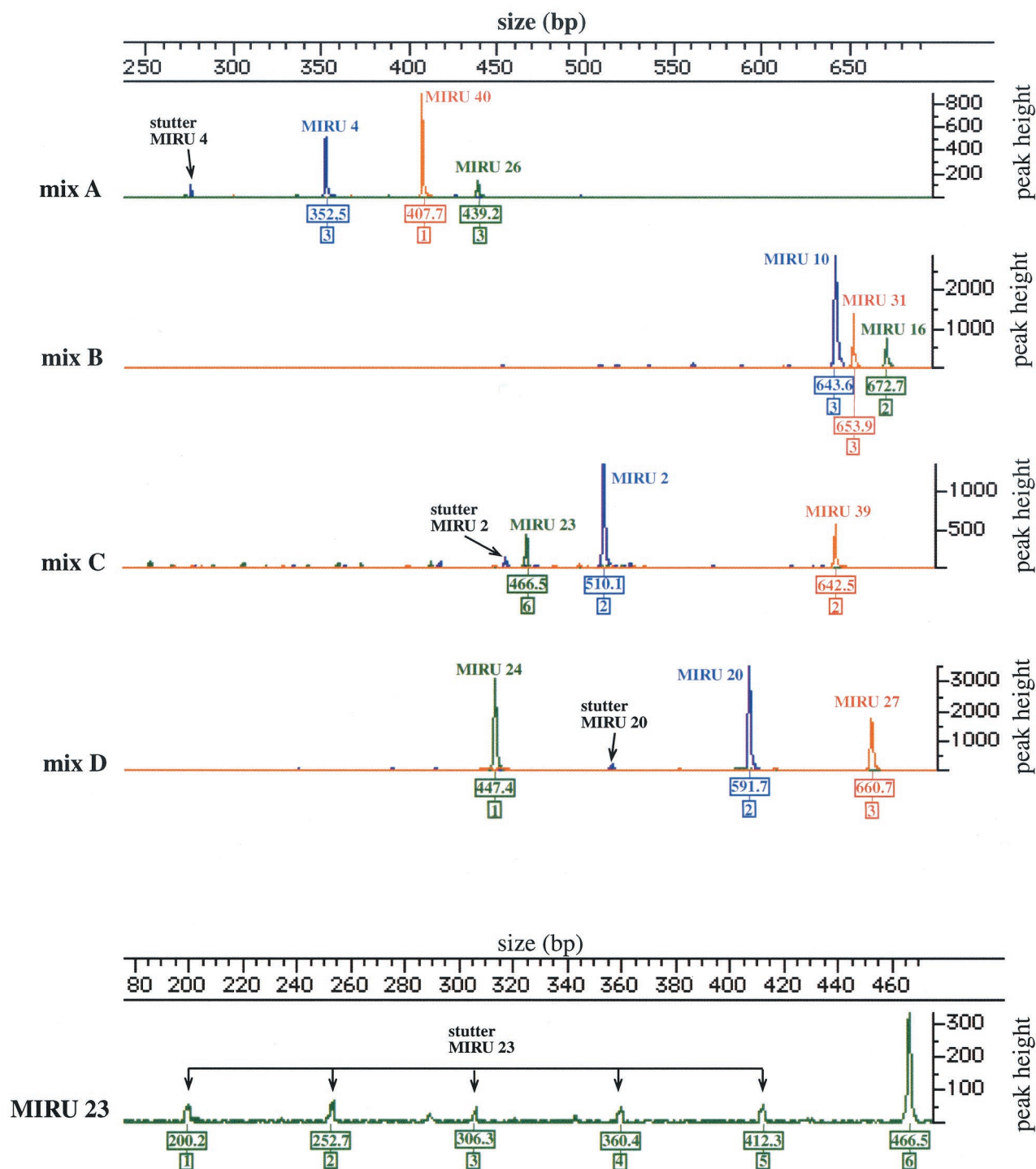


FIG. 1. Multiplex PCR analysis of the 12 MIRU-VNTR loci (top) of *M. tuberculosis* H37Rv with a detailed view of locus 23 (bottom). Four multiplex PCRs were performed on genomic DNA from *M. tuberculosis* H37Rv, and the products were loaded onto four lanes of an automated sequencer. The GeneScan software was used for the collection of the raw sizing data. The results from the analysis of these data with the Genotyper software are shown. Boxed numbers below the signal peaks, calculated sizes (in base pairs) of the labeled PCR products and the corresponding identified MIRU-VNTR alleles. Positions of stutter peaks, easily diagnosed as low-intensity PCR artifacts, are shown.

Haarlem strains (12). However, one or two *M. tuberculosis* isolates with more distantly related or unrelated IS6110 RFLPs were included in the three families by MIRU-VNTR genetic distance analysis (isolates 1 and 56 for the Haarlem family, 96 for the Africa family, and 72 for the Beijing family).

The remaining clinical isolates with identical MIRU-VNTR genotypes included a group of two *M. tuberculosis* strains from Ecuador (isolate 10) and Spain (isolate 95) and a group of two high-copy-number *M. bovis* strains from Saudi Arabia (isolate 24) and The Netherlands (isolate 69). These isolates also have

TABLE 2. MIRU-VNTR analysis of 90 strains of the *M. tuberculosis* complex

Strain ^a	Copy no. for MIRU locus ^b												Species	Origin
	2	4	10	16	20	23	24	26	27	31	39	40		
1	2	2	5	3	1	3	1	5	3	3	1	1	<i>M. tuberculosis</i>	The Netherlands
2 ^c	2	2	2	3	2	4	2	5	3	3	2	2	<i>M. bovis BCG</i>	The Netherlands
4 ^c	2	2	3	3	1	5	1	4	3	4	2	3	<i>M. tuberculosis</i>	Rwanda
6	2	2	5	3	2	4	2	4	3	5	2	2	<i>M. bovis</i>	The Netherlands
7 ^c	1	2	4	3	2	5	1	5	3	2	2	4	<i>M. tuberculosis</i>	Mongolia
8	2	2	4	3	2	5	1	5	3	3	2	3	<i>M. tuberculosis</i>	Vietnam
10 ^c	2	2	4	2	2	6	1	5	3	3	2	1	<i>M. tuberculosis</i>	Ecuador
11 ^c	2	6	2	2	1	1	1	5	3	3	2	3	<i>M. tuberculosis</i>	Greenland
12	2	2	3	3	2	5	1	5	2	2	2	3	<i>M. tuberculosis</i>	Tunisia
13 ^c	2	2	5	3	2	5	1	5	3	3	2	3	<i>M. tuberculosis</i>	Sri Lanka
14	2	2	3	3	2	5	1	7	3	5	4	3	<i>M. tuberculosis</i>	South Africa
15 ^c	2	2	2	3	2	5	1	2	3	3	1	3	<i>M. tuberculosis</i>	Iran
16	2	3	3	3	2	5	1	5	3	3	2	5	<i>M. tuberculosis</i>	Canada
17	3	6	2	2	1	1	1	5	3	3	2	4	<i>M. tuberculosis</i>	Greenland
18 ^c	2	2	4	2	2	6	1	5	3	3	2	4	<i>M. tuberculosis</i>	United States
19	2	2	6	4	2	5	1	5	3	5	3	2	<i>M. tuberculosis</i>	India
20 ^c	2	2	3	3	2	5	1	7	3	6	3	3	<i>M. tuberculosis</i>	Mongolia
21	2	5	4	3	1	6	2	2	3	6	4	3	<i>M. tuberculosis</i>	Zimbabwe
23 ^c	1	2	4	2	2	3	1	5	3	2	2	4	<i>M. tuberculosis</i>	Honduras
24	2	3	7	4	2	4	1	4	3	4	2	2	<i>M. bovis</i>	Saudi Arabia
25	2	6	7	4	1	4	2	2	2	1	2	2	<i>M. microti</i>	United Kingdom
26	2	2	4	3	1	5	1	5	3	2	2	1	<i>M. tuberculosis</i>	Zimbabwe
27 ^c	2	2	3	3	2	5	1	4	3	3	2	4	<i>M. tuberculosis</i>	Ethiopia
28 ^c	2	2	5	3	2	5	1	5	3	2	2	3	<i>M. tuberculosis</i>	Sri Lanka
29	2	2	5	3	2	5	1	5	3	3	2	3	<i>M. tuberculosis</i>	India
30	2	2	3	3	2	5	1	8	3	5	3	3	<i>M. tuberculosis</i>	South Africa
31	2	2	2	3	2	5	1	5	3	3	2	3	<i>M. tuberculosis</i>	Iran
32 ^c	2	3	3	2	2	6	1	3	3	3	2	1	<i>M. tuberculosis</i>	H37Rv
33	2	2	5	3	2	3	1	5	3	3	2	3	<i>M. tuberculosis</i>	The Netherlands
34	2	2	3	3	2	5	1	7	3	5	3	3	<i>M. tuberculosis</i>	Malaysia
35 ^c	2	2	3	3	2	5	1	5	4	4	2	4	<i>M. tuberculosis</i>	Rwanda
36	2	4	4	3	2	6	2	2	3	5	2	2	<i>M. tuberculosis</i>	India
37 ^c	2	2	3	3	1	5	1	4	3	4	2	3	<i>M. tuberculosis</i>	Uganda
38 ^c	2	2	4	3	2	5	1	5	3	3	2	5	<i>M. tuberculosis</i>	Tahiti
40	2	2	3	3	2	5	1	3	3	3	2	3	<i>M. tuberculosis</i>	Burundi
41 ^c	2	2	3	3	2	6	1	3	3	3	2	1	<i>M. tuberculosis</i>	Chile
42 ^c	1	2	4	3	2	6	1	5	3	2	2	6	<i>M. tuberculosis</i>	Tahiti
43	2	2	3	3	2	5	1	7	3	5	3	3	<i>M. tuberculosis</i>	China
44	2	2	3	3	2	5	1	5	3	5	3	3	<i>M. tuberculosis</i>	Thailand
45 ^c	2	2	3	3	1	5	1	5	3	5	4	4	<i>M. tuberculosis</i>	Malaysia
46	2	2	4	2	2	7	1	5	3	3	2	3	<i>M. tuberculosis</i>	Chile
47	2	5	3	3	2	6	2	2	3	5	1	3	<i>M. bovis</i>	The Netherlands
49	2	2	8	3	2	5	1	1	3	4	3	4	<i>M. tuberculosis</i>	Tanzania
50 ^c	2	2	5	3	2	5	1	5	3	3	2	3	<i>M. tuberculosis</i>	Czech Republic
51	2	2	5	3	2	5	1	5	3	3	2	3	<i>M. tuberculosis</i>	The Netherlands
53 ^c	2	2	5	3	1	3	1	5	3	3	2	1	<i>M. tuberculosis</i>	Argentina
54	2	2	3	3	2	5	1	7	3	5	3	2	<i>M. tuberculosis</i>	Thailand
56	2	2	5	3	2	5	1	5	3	2	2	4	<i>M. tuberculosis</i>	Curacao
58	2	2	6	3	2	5	1	5	3	3	2	3	<i>M. tuberculosis</i>	Argentina
60	2	2	5	3	2	5	1	5	3	3	2	3	<i>M. tuberculosis</i>	Bolivia
62	2	6	7	3	1	4	2	2	2	1	2	2	<i>M. microti</i>	United Kingdom
63 ^c	2	2	5	3	1	3	1	5	3	3	2	4	<i>M. tuberculosis</i>	Italy
65 ^c	2	5	4	3	2	6	2	2	3	6	1	3	<i>M. tuberculosis</i>	The Netherlands
66 ^c	2	2	4	3	2	6	1	3	3	3	2	3	<i>M. tuberculosis</i>	Spain
67	2	5	4	3	2	A	2	2	3	6	3	3	<i>M. tuberculosis</i>	Comoro Islands

Continued on following page

TABLE 2—Continued

Strain ^a	Copy no. for MIRU locus ^b												Species	Origin
	2	4	10	16	20	23	24	26	27	31	39	40		
71	2	3	2	3	2	4	2	5	3	3	2	2	<i>M. bovis BCG</i>	Japan
72	2	2	3	3	2	5	1	4	3	5	3	3	<i>M. tuberculosis</i>	Central African Republic
73	2	3	2	2	2	4	2	4	3	3	2	2	<i>M. bovis</i>	The Netherlands
74 ^c	2	5	4	3	2	6	2	2	3	5	3	4	<i>M. tuberculosis</i>	India
76	2	1	6	4	2	4	2	2	2	3	2	2	<i>M. bovis</i>	Argentina
77	2	2	5	3	2	3	1	5	3	3	2	3	<i>M. tuberculosis</i>	Tunisia
81	2	5	6	3	2	4	2	2	2	3	2	1	<i>M. bovis</i>	Argentina
82	2	3	3	3	2	5	1	4	3	3	2	4	<i>M. tuberculosis</i>	Canada
83	2	2	2	3	2	4	2	5	3	3	2	2	<i>M. bovis BCG</i>	Russia
85 ^c	2	1	5	4	2	4	2	5	3	5	2	2	<i>M. bovis</i>	The Netherlands
86	2	2	5	3	2	5	1	5	3	3	2	3	<i>M. tuberculosis</i>	Bolivia
87 ^c	2	2	3	3	2	5	1	4	3	3	2	3	<i>M. tuberculosis</i>	United States
88	2	2	7	2	2	5	1	1	3	2	2	3	<i>M. tuberculosis</i>	Russia
90	2	2	3	3	2	5	1	7	3	2	3	3	<i>M. tuberculosis</i>	South Korea
92	2	3	4	4	2	4	2	2	4	2	2	1	<i>M. africanum</i>	The Netherlands
93	2	5	4	3	2	B	2	2	3	6	3	3	<i>M. tuberculosis</i>	Tanzania
95	2	2	4	2	2	6	1	5	3	3	2	1	<i>M. tuberculosis</i>	Spain
96 ^c	2	2	3	1	2	5	1	5	3	3	2	2	<i>M. tuberculosis</i>	The Netherlands
97	2	1	3	3	1	5	1	2	3	4	2	3	<i>M. tuberculosis</i>	Uganda
98 ^c	2	2	4	2	2	6	1	6	3	3	2	1	<i>M. tuberculosis</i>	Ecuador
100	2	2	6	2	2	5	1	7	3	4	3	3	<i>M. africanum</i>	The Netherlands
101	2	5	6	4	2	4	2	2	2	3	2	1	<i>M. bovis</i>	Argentina
103 ^c	2	2	2	3	2	5	1	5	3	2	2	3	<i>M. tuberculosis</i>	China
109	2	3	3	2	2	6	1	3	3	3	2	1	<i>M. tuberculosis</i>	H37Ra
111	2	2	3	4	2	5	1	7	3	5	3	3	<i>M. tuberculosis</i>	South Korea
112	2	2	6	4	2	5	1	7	3	5	3	2	<i>M. tuberculosis</i>	The Netherlands
116 ^c	3	2	3	2	1	2	6	3	2	4	2	8	<i>M. canettii</i>	Somalia
117	2	3	2	2	2	4	2	6	3	3	2	2	<i>M. bovis</i>	Argentina
118	2	2	4	2	2	6	1	3	4	2	2	3	<i>M. tuberculosis</i>	Honduras
120	2	2	3	3	2	5	1	5	3	5	2	2	<i>M. tuberculosis</i>	Burundi
121	2	2	3	3	2	5	1	5	3	4	2	3	<i>M. tuberculosis</i>	Central African Republic
123	2	3	5	3	2	6	1	4	3	3	2	4	<i>M. tuberculosis</i>	Czech Republic
126	2	3	2	2	2	4	2	5	3	3	2	2	<i>M. bovis</i>	Argentina
130	2	3	2	3	2	4	2	5	3	3	2	2	<i>M. bovis</i>	The Netherlands

^a Strain numbers are identical to those of Kremer et al. (12).

^b MIRU-VNTR alleles are based on the conventions described by Supply et al. (19), except that letters are used for copy numbers greater than 9 (A = 10, B = 11), which allows easier coding of strains by simple 12-digit profiles (J. Crawford, personal communication). Locus 4 contains an additional invariable truncated MIRU in clinical isolates, which is absent in BCG, H37Rv, and H37Ra strains. *M. bovis* strains 71 (BCG) and 130 (clinical isolate) are distinguished by this single feature.

^c Strains with duplicate DNA samples.

highly similar, although not identical, IS6110 RFLP patterns (12), suggesting that they are clonal. *M. tuberculosis* laboratory strain H37Rv and its avirulent derivative H37Ra, which have been cultivated separately for several decades, also displayed identical MIRU-VNTR patterns, although they differ by 3 of 14 IS6110 bands. Moreover, the Russian and French *M. bovis* BCG vaccine sister strains displayed identical MIRU-VNTRs, whereas the Japanese sister strain differed from these by only a single unit change in locus 4, consistent with the data of Supply et al. (19).

DISCUSSION

The 12-locus MIRU-VNTR method described here represents a powerful approach to high-resolution genotyping of *M. tuberculosis* isolates. It uses multiplex PCR with fluorescence-labeled primers to analyze four sets of three MIRU-VNTR loci

simultaneously and provides the automation that is necessary for quick and reliable production and management of the genotyping data. To our knowledge, up to now, only one comparable VNTR-based method has been reported for typing another bacterium, *Bacillus anthracis* (11). MIRU-VNTR typing was found to be 100% reproducible, sensitive, and specific for *M. tuberculosis* complex isolates, a performance that has not been achieved by any other resolution typing method, including IS6110 RFLP, tested in the same conditions (12). Moreover, the discrimination power of MIRU-VNTR typing was close to that of IS6110 RFLP (78 compared to 84 patterns, respectively), and MIRU-VNTR typing performed better than the other RFLP-based or IS6110-independent PCR-based typing methods (compare the present results with those in reference 12). In addition, the genetic-distance dendrogram built from MIRU-VNTRs correlated well with that based on IS6110 RFLP, with about 80% of the isolates similarly grouped by the

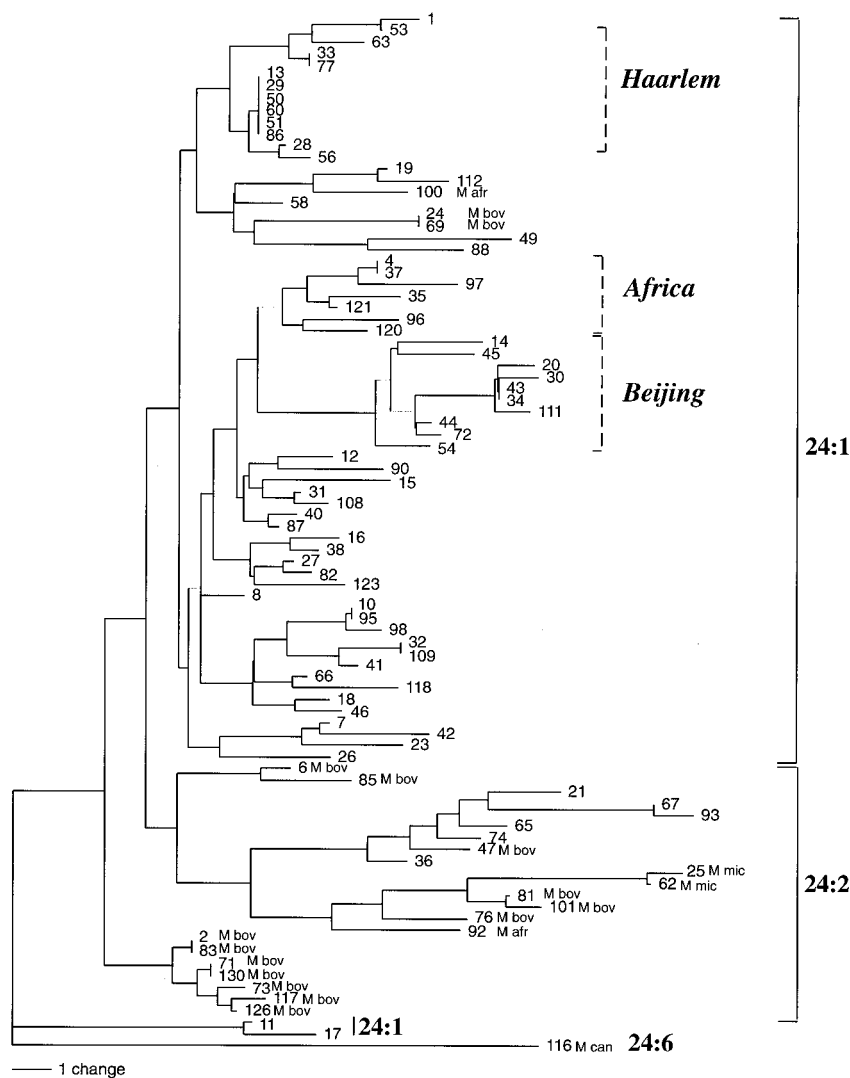


FIG. 2. Dendrogram of genetic relationships among 90 isolates of the *M. tuberculosis* complex from 38 countries based on the 12 MIRU-VNTR loci. A majority rule consensus tree was built using the neighbor-joining distance algorithm as described in Materials and Methods. The linkage distance scale is indicated at the bottom. *M. bovis* isolate 130 is discriminated from *M. bovis* isolate 71 based on the presence of an additional 53-bp invariable MIRU unit in locus 4 in the first strain (this unit is not taken into account in the MIRU-VNTR dendrogram analysis). Dotted lines, groups including mainly Haarlem, Africa, and Beijing isolates identified by Kremer et al. (12). The numbers of MIRU-VNTRs in locus 24 (24:*n*) are indicated at the right. Isolates are *M. tuberculosis*, except for those indicated as M bovis, M afr, M can, and M mic, corresponding to *M. bovis*, *Mycobacterium africanum*, *M. canettii*, and *M. microti* isolates, respectively.

two methods. These findings are similar to those of a more limited study of *M. tuberculosis* strains isolated in Paris (14). The results for 90 strains from 38 countries presented here thus extend these observations to the scale of global genetic diversity of the *M. tuberculosis* complex worldwide.

MIRU-VNTRs clearly distinguished the majority of the isolates of Beijing, Africa, and Haarlem families from each other. Most of them were also recognized as being closely related by MIRU-VNTRs (see also below). The Beijing family appeared the most homogenous family, which is consistent with the results obtained with IS6110 RFLP and other markers (12). Our analysis using the high-resolution typing system fully independent from IS6110 thus supports the assumption that these isolates have expanded clonally from common ancestors and

subsequently disseminated globally (12). In contrast, almost all of the clinical isolates with high IS6110 copy numbers that had not been assigned to the Beijing, Africa, and Haarlem families had unique MIRU-VNTRs, except for two groups of two isolates each. However, these isolates might correspond to other genotype families, as they also had high similarities in their IS6110 RFLPs. In addition, MIRU-VNTRs distinguished all strains with low IS6110 copy numbers except for two *M. bovis* BCG vaccine sister strains, in contrast to IS6110 RFLP, for which two clusters of two and five strains with identical patterns were obtained. This class of isolates includes many *M. tuberculosis* isolates from several regions, especially Asia, and about 80% *M. bovis* isolates from all regions. Their typing is often problematic and requires combinations of several tech-

niques (4). Thus, MIRU-VNTR typing is especially useful for analyzing the genetic diversity among these isolates.

The differences in genetic-relationship results between MIRU-VNTR typing and IS6110 RFLP analysis, observed for only a minority of isolates, may be due to several factors. Some of these differences correspond to strains or groups of strains which apparently share a few IS6110 bands with the other strains of the complex (for instance, *Mycobacterium microti* isolates 25 and 62). IS6110 RFLP analysis suggested the closest relationships between these strains and groups including *M. tuberculosis* isolates (12), while MIRU-VNTR analysis suggested the closest relationships to *M. bovis* strains (Fig. 2). The latter grouping is consistent with results from comparative genomics suggesting closer relationships between these strains, which are pathogenic for animals (3). This difference might be due to the fact that markers such as IS6110 RFLP may evolve too fast to be reliable for estimating higher levels of divergence (20). Moreover, the documented existence of IS6110 insertion hot spots (5, 6, 15) together with the intrinsic difficulties of interpreting RFLP patterns can conceivably mislead the analysis of some strains, especially those sharing only few bands. On the other hand, we cannot exclude the possibility that independent convergence to closely related MIRU-VNTR genotypes by unrelated isolates may occasionally occur. This might explain the grouping by MIRU-VNTR typing of a few isolates assigned to the Africa and Haarlem families, which, based on analysis with IS6110 RFLP and other markers, are classified as unrelated isolates (and vice versa). However, this convergence is necessarily limited by the huge number of possible MIRU-VNTR combinations ($>2 \times 10^8$) provided by the 12 loci, which strongly reduces the probability of obtaining totally identical genotypes only by chance. Consistently, it was noted that the few isolates with identical MIRU-VNTR types in the different genotype families, as well as H37Rv and H37Ra strains, all had only slightly distinct IS6110 RFLP patterns.

The last observation supports the assumption that the combined molecular clock of the 12 MIRU-VNTR loci runs slightly slower than that of IS6110 RFLP. It may be useful to note that, like multilocus enzyme electrophoresis for instance, the genetic variation indexed by MIRU-VNTR typing relies on independent genomic loci that can be separately targeted by specific PCR primers. This is different from IS6110 RFLP, where all genomic loci containing IS6110 are probed simultaneously, without the possibility to assess individual loci. This provides the possibility to modulate the speed of the combined molecular clock of MIRU-VNTRs by selecting one or several targeted loci to analyze different scales of evolutionary divergence within the same system. Interestingly, the inclusion of two other VNTR loci identified by Frothingham and Meeker-O'Connell (7), namely, VNTR-A and VNTR-C, allowed us to distinguish three additional patterns among strains with identical MIRU-VNTRs, including two members of the Africa family (strains 4 and 37), BCG Pasteur and BCG Russia (strains 2 and 83), and two *M. bovis* isolates with high IS6110 copy numbers (strains 24 and 69) (12). Inclusion of other polymorphic tandem repeat loci in the *M. tuberculosis* complex genomes (16; Supply et al., unpublished results) may thus further improve the performance of MIRU-VNTR typing compared to that of IS6110 RFLP for tracing recent transmission of strains circulating within defined geographic areas.

Conversely, the selective use of some less-variable MIRU-VNTR loci, such as locus 24, may be especially informative for higher levels of evolutionary divergence. Locus 24 is among the least diverse loci, with only three alleles (Table 2). This locus is polymorphic between three groups, which appear genetically distant by both MIRU-VNTR and IS6110 RFLP typing (12), while it is monomorphic within each of these groups (Fig. 2). One group includes mainly *M. tuberculosis* and *M. bovis* isolates with high IS6110 copy numbers, *M. microti* isolates together with low-IS6110-copy-number *M. bovis* and *M. tuberculosis* isolates, and a single *M. canettii* isolate. This polymorphism, only observed among genetically distant groups, suggests a very slow evolution rate of locus 24. Therefore, we speculate that the hierarchy of allelic diversity among the MIRU-VNTR loci (14) may be correlated to the speed of their respective molecular clocks, in addition perhaps to selective constraints on the numbers of alleles.

Using the automated MIRU-VNTR typing reported here, several hundred isolates can be genotyped per week on a single sequencer, starting from genomic DNA or bacterial colonies. This is considerably faster than IS6110 RFLP or manual VNTR analysis. Adequate equipment and software are widely available, while specific files and macro software for standardized allele scoring will be provided by the authors upon request. Compared to IS6110 RFLP, the portability of MIRU-VNTR typing results will greatly facilitate the comparisons of results between different laboratories worldwide. To ensure consistency of results on a daily basis within and among laboratories, genotypic tests of standardized reference strains H37Rv or H37Ra (Table 2) should be run as validating controls. To generate a database allowing laboratories to compare their data, we have set up a website for the analysis of *M. tuberculosis* complex MIRU-VNTR genotypes via the Internet (<http://www.ibl.fr/mirus/mirus.html>). We hope that the use of this automated typing system for large-scale genotyping of *M. tuberculosis* will provide a new tool for *M. tuberculosis* global epidemiological surveillance and lead to novel comprehensive insights into the population genetics and evolution of this major human pathogen.

ACKNOWLEDGMENTS

Vincent Vatin and Karl Oulmi are gratefully acknowledged for their help in software utilization.

The work was supported by INSERM, Institut Pasteur de Lille, a grant from the Ministère de l'Éducation Nationale, de la Recherche et de la Technologie, and the European Community (grant QLK2-CT-2000-00630). P.S. is a Chercheur du Centre National de Recherche Scientifique. E.S. holds a Poste Vert from the INSERM.

REFERENCES

1. Bifani, P. J., B. Mathema, Z. Liu, S. L. Moghazeh, B. Shopsis, B. Tempalski, J. Driscoll, R. Frothingham, J. M. Musser, P. Alcades, and B. N. Kreiswirth. 1999. Identification of a W variant outbreak of *Mycobacterium tuberculosis* via population-based molecular epidemiology. *JAMA* **282**:2321–2327.
2. Bifani, P. J., B. B. Plikaytis, V. Kapur, K. Stockbauer, X. Pan, M. L. Lutfeey, S. L. Moghazeh, W. Eisner, T. M. Daniel, M. H. Kaplan, J. T. Crawford, J. M. Musser, and B. N. Kreiswirth. 1996. Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family. *JAMA* **275**:452–457.
3. Brosch, R., S. V. Gordon, K. Eiglmeier, T. Garnier, F. Tekaia, E. Yeramian, and S. T. Cole. 1999. Genomics, biology, and evolution of the *Mycobacterium tuberculosis* complex, p. 19–36. In G. F. Hatful and W. R. Jacobs, Jr. (ed.), *Molecular genetics of mycobacteria*. American Society for Microbiology, Washington, D.C.
4. Collins, D. M. 1999. Molecular epidemiology: *Mycobacterium bovis*, p. 123–

135. In C. Ratledge and J. Dale (ed.), *Mycobacteria. Molecular biology and virulence*. Blackwell Science Ltd., Oxford, United Kingdom.
5. Fang, Z., and K. J. Forbes. 1997. A *Mycobacterium tuberculosis* IS6110 preferential locus (*ipl*) for insertion into the genome. *J. Clin. Microbiol.* **35**:479–481.
 6. Fang, Z., N. Morrison, B. Watt, C. Doig, and K. J. Forbes. 1998. IS6110 transposition and evolutionary scenario of the direct repeat locus in a group of closely related *Mycobacterium tuberculosis* strains. *J. Bacteriol.* **180**:2102–2109.
 7. 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.
 8. Gleissberg, V. 1999. The threat of multidrug resistance: is tuberculosis ever untreatable or uncontrollable? *Lancet* **353**:998–999.
 9. Goguet de la Salmoniere, Y. O., H. M. Li, G. Torrea, A. Bunschoten, J. van Embden, and B. Gicquel. 1997. Evaluation of spoligotyping in a study of the transmission of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **35**:2210–2214.
 10. Grange, J. M., and A. Zumla. 1999. Paradox of the global emergency of tuberculosis. *Lancet* **353**:996.
 11. Keim, P., L. B. Price, A. M. Klevytska, K. L. Smith, J. M. Schupp, R. Okinaka, P. J. Jackson, and M. E. Hugh-Jones. 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J. Bacteriol.* **182**:2928–2936.
 12. Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. van Embden. 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. *J. Clin. Microbiol.* **37**:2607–2618.
 13. Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* **95**:3140–3145.
 14. Mazars, E., S. Lesjean, A.-L. Bañuls, M. Gilbert, V. Vincent, B. Gicquel, M. Tibayrenc, C. Locht, and P. Supply. 2001. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc. Natl. Acad. Sci. USA* **98**:1901–1906.
 15. McHugh, T. D., and S. H. Gillespie. 1998. Nonrandom association of IS6110 and *Mycobacterium tuberculosis*: implications for molecular epidemiological studies. *J. Clin. Microbiol.* **36**:1410–1413.
 16. Smittipat, N., and P. Palittapongarnpim. 2000. Identification of possible loci of variable number of tandem repeats in *Mycobacterium tuberculosis*. *Tuber. Lung Dis.* **80**:69–74.
 17. Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam, and J. M. Musser. 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl. Acad. Sci. USA* **94**:9869–9874.
 18. Supply, P., J. Magdalena, S. Himpens, and C. Locht. 1997. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. *Mol. Microbiol.* **26**:991–1003.
 19. Supply, P., E. Mazars, S. Lesjean, V. Vincent, B. Gicquel, and C. Locht. 2000. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol. Microbiol.* **36**:762–771.
 20. Tibayrenc, M. 1999. Toward an integrated genetic epidemiology of parasitic protozoa and other pathogens. *Annu. Rev. Genet.* **33**:449–477.
 21. Torrea, G., C. Offredo, M. Simonet, B. Gicquel, P. Berche, and C. Pierre-Audigier. 1996. Evaluation of tuberculosis transmission in a community by 1 year of systematic typing of *Mycobacterium tuberculosis* clinical isolates. *J. Clin. Microbiol.* **34**:1043–1049.
 22. van Soolingen, D., M. W. Borgdorff, P. E. de Haas, M. M. Sebek, J. Veen, M. Dessens, K. Kremer, and J. D. van Embden. 1999. Molecular epidemiology of tuberculosis in The Netherlands: a nationwide study from 1993 through 1997. *J. Infect. Dis.* **180**:726–736.
 23. van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. van Embden. 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. *J. Clin. Microbiol.* **33**:3234–3238.
 24. Yaganehdoost, A., E. A. Graviss, M. W. Ross, G. J. Adams, S. Ramaswamy, A. Wanger, R. Frothingham, H. Soini, and J. M. Musser. 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.