

Comparison of Methods Based on Different Molecular Epidemiological Markers for Typing of *Mycobacterium tuberculosis* Complex Strains: Interlaboratory Study of Discriminatory Power and Reproducibility

K. KREMER,^{1*} D. VAN SOOLINGEN,¹ R. FROTHINGHAM,² W. H. HAAS,³ P. W. M. HERMANS,⁴
C. MARTÍN,⁵ P. PALITTAPONGARNPIM,⁶ B. B. PLIKAYTIS,⁷ L. W. RILEY,⁸
M. A. YAKRUS,⁹ J. M. MUSSER,¹⁰ AND J. D. A. VAN EMBDEN¹¹

*Diagnostic Laboratory for Infectious Diseases and Perinatal Screening*¹ and *Research Laboratory for Infectious Diseases*,¹¹
*National Institute of Public Health and the Environment, 3720 BA Bilthoven, and Laboratory of Pediatrics, Erasmus University Rotterdam, 3000 DR Rotterdam,*⁴ *The Netherlands; Molecular Mycobacteriology Laboratory, University of Heidelberg, 69120 Heidelberg, Germany*³; *Durham VA Medical Center, Durham, North Carolina 27705*²;
*National Center for Infectious Diseases*⁷ and *Diagnostics and Molecular Epidemiology Section,*⁹
*Centers for Disease Control and Prevention, Atlanta, Georgia 30333; School of Public Health, University of California, Berkeley, Berkeley, California 94720*⁸; *Institute for the Study of Human Bacterial Pathogenesis, Department of Pathology, Baylor College of Medicine, Houston, Texas 77030*¹⁰; *Department of Microbiology, University of Zaragoza, 50009 Zaragoza, Spain*⁵; and *Department of Microbiology, Mahidol University, Bangkok 10400, Thailand*⁶

Received 29 January 1999/Returned for modification 1 April 1999/Accepted 13 May 1999

In this study, the currently known typing methods for *Mycobacterium tuberculosis* isolates were evaluated with regard to reproducibility, discrimination, and specificity. Therefore, 90 *M. tuberculosis* complex strains, originating from 38 countries, were tested in five restriction fragment length polymorphism (RFLP) typing methods and in seven PCR-based assays. In all methods, one or more repetitive DNA elements were targeted. The strain typing and the DNA fingerprint analysis were performed in the laboratory most experienced in the respective method. To examine intralaboratory reproducibility, blinded duplicate samples were included. The specificities of the various methods were tested by inclusion of 10 non-*M. tuberculosis* complex strains. All five RFLP typing methods were highly reproducible. The reliability of the PCR-based methods was highest for the mixed-linker PCR, followed by variable numbers of tandem repeat (VNTR) typing and spoligotyping. In contrast, the double repetitive element PCR (DRE-PCR), IS6110 inverse PCR, IS6110 amplifingerprinting, and arbitrarily primed PCR (APPCR) typing were found to be poorly reproducible. The 90 strains were best discriminated by IS6110 RFLP typing, yielding 84 different banding patterns, followed by mixed-linker PCR (81 patterns), APPCR (71 patterns), RFLP using the polymorphic GC-rich sequence as a probe (70 patterns), DRE-PCR (63 patterns), spoligotyping (61 patterns), and VNTR typing (56 patterns). We conclude that for epidemiological investigations, strain differentiation by IS6110 RFLP or mixed-linker PCR are the methods of choice. A strong association was found between the results of different genetic markers, indicating a clonal population structure of *M. tuberculosis* strains. Several separate genotype families within the *M. tuberculosis* complex could be recognized on the basis of the genetic markers used.

Mycobacterium tuberculosis, *Mycobacterium africanum*, *Mycobacterium bovis*, *Mycobacterium microti*, and *Mycobacterium canettii* (55) are genetically closely related subspecies of the *M. tuberculosis* complex. The high degree of evolutionary conservation of *M. tuberculosis* complex strains is exemplified by their high degree of interstrain DNA homology (26), the conservation of their 16S rRNA gene sequence (30) and their 16S to 23S ribosomal RNA (rDNA) intergenic spacer sequences (15), their limited diversity as measured by multilocus enzyme electrophoresis (MLEE) (13) and genomic restriction fragment analysis (5, 11), and their virtual lack of antigenic variation. Consistently, Kapur et al. described a virtual lack of DNA sequence diversity in eight loci for structural genes among

seven *M. tuberculosis* strains of wide geographical origin (28). Sreevatsan et al. have found only two amino acid substitutions which were not linked to antibiotic resistance among two megabases of DNA (45).

Despite this genetic homogeneity in the *M. tuberculosis* complex, a high degree of DNA polymorphism is associated with repetitive DNA such as insertion elements (IS) and short repetitive DNA sequences. Until recently, four ISs, IS6110 (47), -1081 (6), -1547 (12), and the IS-like element (35), had been identified in *M. tuberculosis* complex strains. Due to its apparent mobility and its common presence in, on average, large numbers of copies, IS6110 is widely used as a genetic marker to differentiate clinical *M. tuberculosis* isolates for epidemiological investigations. This includes investigations of the transmission of tuberculosis in hospitals (2, 10), residential facilities for human immunodeficiency virus-infected people (7), prisons (19), and in larger populations (1, 44, 61). In contrast to IS6110, IS1081 is almost invariably present in five to seven copies per genome, and this IS has been found to be associated

* Corresponding author. Mailing address: National Institute of Public Health and the Environment (RIVM), Laboratory for Infectious Diseases and Perinatal Screening, Mycobacteria Department (pb22), P.O. Box 1, 3720 BA Bilthoven, The Netherlands. Phone: 31-30-2742261. Fax: 31-30-2744418. E-mail: kristin.kremer@rivm.nl.

with limited DNA polymorphism (6, 54). IS1547 and the IS-like element are present in one or two copies per genome (4, 12, 35). The discrimination level of IS1547-associated restriction fragment length polymorphisms (RFLP) was found to be low in comparison to IS6110-associated RFLP (12). The establishment of the complete genomic DNA sequence of *M. tuberculosis* H37Rv disclosed the presence of 25 unknown ISs which are present in one to three copies, two prophages, and a novel type of repetitive sequence, the REP13E12 family, which is present in seven copies (4, 17). The presence of these new ISs was investigated in five of the six species of the *M. tuberculosis* complex, including 32 *M. tuberculosis* strains, and only IS1532, -1533, -1534, and -1561' were absent from some of the strains tested (4, 17).

Five types of short repetitive DNA associated with some degree of genetic diversity have been identified in *M. tuberculosis* complex. Three of these, the polymorphic GC-rich tandem repeat sequence (PGRS) (41, 49), a repeat of the triplet GTG (60), and the major polymorphic tandem repeat (MPTR) (25), are present at multiple chromosomal loci. Multiple repeats of PGRSs and MPTRs are often part of the so-called Pro-Glu and Pro-Pro-Glu multigene families, respectively, and it has been postulated that the DNA polymorphism associated with these repetitive elements may result in antigenic variation (4). Furthermore, six exact tandem repeat (ETR) loci have been identified (16, 18). In contrast to the polymorphic MPTR and PGRS, these ETR loci contain tandem repeats of identical DNA sequences. Each locus has a unique repeated sequence, ranging in size from 53 to 79 bp (16). Recently, variable numbers of tandem repeats (VNTR) typing was introduced to detect the polymorphism of the MPTR and ETR loci in *M. tuberculosis* complex strains (16). The short repetitive direct repeat (DR) elements, which are present at a single genomic locus in *M. tuberculosis* complex strains, are 36 bp in length (24). In the DR locus, the DRs are interspersed by unique DNA spacer sequences of 35 to 41 bp in length. Clinical isolates differ in the presence of spacers, and this polymorphism can be visualized by spoligotyping (27).

The PGRS, the DR, and the GTG repeated sequences have mainly been used for subtyping strains for which differentiation by IS6110 fingerprinting appeared insufficient. This is the case when *M. tuberculosis* complex strains contain no or few IS6110 copies, such as in a significant part of the *M. tuberculosis* isolates from Asia (42, 52) and *M. bovis* strains from cattle (49, 51).

A variety of methods have been used to visualize the DNA polymorphism in *M. tuberculosis* complex strains. These methods include restriction fragment length polymorphism (RFLP), DNA hybridization, PCR, and combinations of these methods (14, 16, 20, 27, 33, 37, 38, 40, 48, 50–52, 54, 59, 60).

No study has been undertaken so far to compare the various genetic typing methods with regard to their reproducibility, discriminative power, and specificity. In this interlaboratory study, we compared the most frequently used typing methods by testing a blinded set of mainly *M. tuberculosis* complex strains, including duplicate samples. We compared these typing methods with methods that are generally used in population genetics, such as MLEE, restriction fragment length end labeling (RFEL) analysis, and DNA sequencing.

MATERIALS AND METHODS

Study design. Ninety *M. tuberculosis* complex and 10 non-*M. tuberculosis* complex strains were used in this study (Table 1). The 90 *M. tuberculosis* complex strains comprised 70 *M. tuberculosis* strains isolated in 34 countries, 2 *M. africanum* strains, 12 *M. bovis* strains originating in 3 countries, 3 *M. bovis* BCG vaccine strains, 2 *M. microti* strains (57), and 1 *M. canettii* strain (55). The set of 10 non-*M. tuberculosis* complex strains contained two strains of each of the following species: *Mycobacterium avium*, *Mycobacterium kansasii*, *Mycobacterium gor-*

donae, *Mycobacterium smegmatis*, and *Mycobacterium xenopi*. All strains were grown on Löwenstein-Jensen medium supplemented with pyruvate. DNA was isolated as described previously (53). Thirty-one duplicate DNA samples were made by dividing the DNA of 31 *M. tuberculosis* complex strains into two tubes. A set of 131 DNA samples was compiled, containing the 90 *M. tuberculosis* complex strains, the 31 duplicates of these, and the 10 non-*M. tuberculosis* complex strains. These 131 DNA samples were numbered, the numbers were randomized by computer, and the tubes were renumbered with these randomly generated numbers. This complete set of 131 DNA samples will be referred to as set A. The set of the 90 *M. tuberculosis* complex strains alone will be referred to as set B (Table 1). A selection of four *M. tuberculosis* strains, four *M. bovis* strains, and one strain each of *M. africanum*, *M. canettii*, *M. microti*, *M. bovis* BCG, *M. avium*, *M. kansasii*, *M. gordonae*, *M. smegmatis*, and *M. xenopi* will be referred to as set C. The *M. tuberculosis* and *M. bovis* strains were a subset of set B and were selected on the basis of nonrelatedness with regard to their IS6110 RFLP patterns.

Laboratories specialized in the various techniques were supplied with aliquots of dried DNA samples from set A. The laboratories were asked to subject these blinded DNA samples to the typing method in which they specialized and to analyze the results following standard protocols. The results were returned to the organizing laboratory (National Institute of Public Health and the Environment [RIVM], Bilthoven, The Netherlands), where the sample numbers were decoded. The reproducibility, discriminative power, and specificity of a typing method were determined by the organizing laboratory on the basis of the conclusions drawn by the laboratory that had performed the typing. The organizing laboratory evaluated the results of all typing methods.

RFLP typing. DNA cleavage, separation of DNA restriction fragments by electrophoresis, Southern blot hybridization, and chemiluminescence detection were performed at the RIVM as described previously (53). *Pvu*II-digested DNA was separated on a 0.8% agarose gel and hybridized with IS6110 (48, 53) and -1081 (6, 54), the genes *mtp40* (9), *mpb64* (33), and *katG* (50, 62), and with the 16S rDNA gene (3, 50). The 396-bp probe to detect the *mtp40* sequence was prepared by PCR amplification of genomic DNA of *M. tuberculosis* reference strain M14323 (48) by using primers PT1 and PT2 as described by Del Portillo et al. (9). The described *mtp40* sequence is a part of *mpcA*, a phospholipase C gene, which cross-hybridizes weakly with *mpcB* (59). The whole RFLP pattern, visualized by using the *mtp40* probe, was used for analysis.

*Alu*I-digested DNA was separated on both 0.6 and 1.5% agarose gels, blotted, and thereafter hybridized with the PGRS (41, 52) and DR probes (28), respectively. (GTG)₅ polymorphism was visualized on *Hinf*I-digested DNA as described by Wiid et al. (60).

PCR-based typing. Four PCR-based typing methods detecting IS6110 as a target sequence were performed: mixed-linker PCR (University of Heidelberg, Heidelberg, Germany) (20), IS6110 inverse PCR (University of Zaragoza, Zaragoza, Spain) (37), IS6110 amplifingerprinting (Centers for Disease Control and Prevention [CDC] Atlanta, Ga.) (40), and double repetitive element PCR (DRE-PCR) (Cornell University Medical College, New York, N.Y.) (14). In addition to IS6110, IS6110 amplifingerprinting targets MPTR, and DRE-PCR targets PGRS. Spoligotyping, a method detecting 43 known spacer sequences which intersperse the DRs in the genomic DR region of *M. tuberculosis* complex strains, was performed as described by Kamerbeek et al. (27, 56) at the RIVM. Arbitrarily primed PCR (APPCR) was done by using four different primer sets: DKU44, DKU43 plus DKU44, DKU43 plus DKU49, and DKU44 plus DKU49 (38) at the Mahidol University, Bangkok, Thailand.

VNTR typing was performed using five loci, ETR-A through ETR-E, as described by Frothingham and Meeker-O'Connell (16), at the Durham VA Medical Center, Durham, N.C. The VNTR loci were amplified by PCR by using specific primers complementary to the flanking regions. The number of tandem repeat units was determined by estimating the size of the PCR product on agarose gels. The results were expressed as a five-digit allele profile in which each digit represented the number of copies at a particular locus.

Generic typing methods. RFEL analysis was performed as described by Van Steenberg et al. (58) and adapted by Hermans et al. (23) at the Erasmus University, Rotterdam, The Netherlands. Briefly, the purified mycobacterial DNA was digested by the restriction enzyme *Eco*RI. Purified restriction fragments were [α -³²P]dATP end labeled at 72°C by using DNA polymerase (Goldstar; Eurogentec, Seraing, Belgium). The radiolabeled fragments were denatured and separated electrophoretically on a 6% polyacrylamide sequencing gel containing 8 M urea. After transfer onto filter paper and vacuum drying (Bio-Rad Laboratories, Inc., Veenendaal, The Netherlands), the gel was exposed to an X-ray film. DNA restriction fragments ranging from 160 to 400 bp were used for RFEL analysis.

The lysate preparation and protein electrophoresis for MLEE were performed according to the method described by Selander et al. (CDC) (43). Eighteen enzymes encoded by chromosomal genes were assayed: malate dehydrogenase, isocitrate dehydrogenase, 6-phosphogluconate dehydrogenase NAD, 6-phosphogluconate dehydrogenase NADP, glucose 6-phosphate dehydrogenase, benzyl alcohol dehydrogenase, alanine dehydrogenase, diaphorase, indophenol oxidase, nucleoside phosphorylase, glutamate oxaloacetic transaminase, adenylate kinase, phosphoglucose mutase, esterase, leucine aminopeptidase, fumarase, aconitase, and phosphoglucose isomerase. The enzymes were stained with the buffer systems described by Selander et al. (43), except for diaphorase, which was stained

TABLE 1. Mycobacterial strains used in this study

Species	Strain ^a	Source or characteristic	Country of isolation	Origin		
<i>M. tuberculosis</i>	53 (61), 58	Human	Argentina	A. Alito		
	60, 86		Bolivia	M. Camacho		
	40, 120		Burundi	F. A. Portaels		
	16, 82		Canada	H. Hoepfner		
	72, 121		Central African Republic	F. A. Portaels		
	41 (80), 46		Chile	M. Velasco		
	43 , 103 (108)		China	W. Guozi		
	67		Comoro Islands	F. A. Portaels		
	56		Curacao	R. J. Gijssbertha		
	50 (84), 123		Czech Republic	M. Kubin		
	10 (59), 98 (124)		Ecuador	J. Mendibele		
	27 (104)		Ethiopia	P. W. M. Hermans		
	11 (52), 17		Greenland	Z. Yang		
	23 (64), 118		Honduras	G. Dubon		
	19, 29, 36, 74 (102)		India	This laboratory		
	15 (55), 31		Iran	A. Moghaddam		
	63 (99)		Italy	A. Gori		
	34, 45 (113)		Malaysia	Z. Zainuddin		
	7 (131), 20 (22)		Mongolia	D. Enkhasaikan		
	88		Russia	V. Golyshevskaya		
	4 (125), 35 (79)		Rwanda	F. A. Portaels		
	14, 30		South Africa	W. Sturm		
	90, 111		South Korea	A. H. J. Kolk		
	66 (89), 95		Spain	S. Samper		
	13 (70), 28 (114)		Sri Lanka	J. Perrera		
	38 (75), 42 (78)		Tahiti	B. Giquel		
	49, 93		Tanzania	Z. Yang		
	44, 54		Thailand	A. H. J. Kolk		
	1, 33, 51, 65 (122), 96 (105), 112		The Netherlands	This laboratory		
	12, 77		Tunisia	F. Messadi-Akrout		
	37 (94), 97		Uganda	T. Aisu		
	18 (107), 87 (91)		United States	P. M. Small		
	8		Vietnam	A. H. J. Kolk		
	21, 26		Zimbabwe	J. ten Have		
	32 (68)		H37Rv	United States	This laboratory	
	109		H37Ra	United States	This laboratory	
	<i>M. africanum</i>		92 , 100	Human	The Netherlands	This laboratory
			<i>M. bovis</i>	Cattle	Argentina	A. Alito
	<i>M. bovis</i>		73, 130	Human	The Netherlands	This laboratory
			6, 47 , 85 (128)		The Netherlands	This laboratory
			24		Oryx	Saudi Arabia
69		The Netherlands	J. Haagsma			
76, 81 , 101		Seal	Argentina		A. Alito	
<i>M. bovis BCG</i>	71	Vaccine strain	Japan	D. Fomukong		
	2 (48)		The Netherlands	This laboratory		
	83		Russia	This laboratory		
<i>M. microti</i>	25, 62	Vole	United Kingdom	P. Draper		
<i>M. canettii</i>	116 (129)	Human	Somalia	This laboratory		
<i>M. avium</i>	3, 57	Human	The Netherlands	This laboratory		
<i>M. gordonae</i>	119	ATCC 14470		This laboratory		
	127	Human	The Netherlands	This laboratory		
<i>M. kansasii</i>	5, 115	Human	The Netherlands	This laboratory		
<i>M. smegmatis</i>	106	ATCC 10143		This laboratory		
	110	ATCC 607		This laboratory		
<i>M. xenopi</i>	9, 39	Human	The Netherlands	This laboratory		

^a Set A comprised all strains, including duplicate DNA samples of those strains which are indicated between parentheses after the original DNA sample; set B comprised all 90 *M. tuberculosis* complex strains, without the duplicate strains; set C was composed of the numbers indicated in bold.

by using the method of Harris and Hopkinson (21). To run the starch gels, a Tris-citrate buffer system, pH 8.0, was used for all the enzymes (43). Electromorphs (allozymes) of each enzyme were equated with alleles at the corresponding structural gene loci, and an absence of enzyme activity was attributed to a null allele. Distinct combinations of alleles (multilocus genotypes) were designated as electrophoretic types (ETs) (43).

The *oxyR*, *ideR*, *rs*, *aroA*, and 16-kDa-antigen genes were sequenced as described previously by Sreevatsan et al. (45) at the Baylor College of Medicine, Houston, Tex. The following nucleotides were analyzed: *oxyR* (GenBank U18263), bp 1 to 528; *ideR* (GenBank U14191), bp 102 to 730; *rs* (*M. tubercu-*

losis; GenBank X52917), bp 1 to 1046; *rs* (*M. xenopi*; GenBank X52929), bp 1 to 976; *rs* (*M. avium*; GenBank X52918), bp 1 to 996; *rs* (*M. kansasii*; GenBank X15916), bp 1 to 968; *rs* (*M. gordonae*; GenBank X52923), bp 1 to 969; *aroA* (GenBank M62708), bp 100 to 1062; and 16-kDa-antigen gene (GenBank S79757), bp 25 to 680.

Computer-assisted analysis. The computer-assisted analysis of the banding patterns obtained by IS6110, IS1081, PGRS, (GTG)₅, DR, *mup40*, *mpb64*, *katG*, and 16S rDNA gene RFLP; spoligotyping; RFEL analysis; mixed-linker PCR; and IS6110 inverse PCR was done by using the Windows version of the GelCompar software (version 4.0; Applied Maths, Kortrijk, Belgium) as previously de-

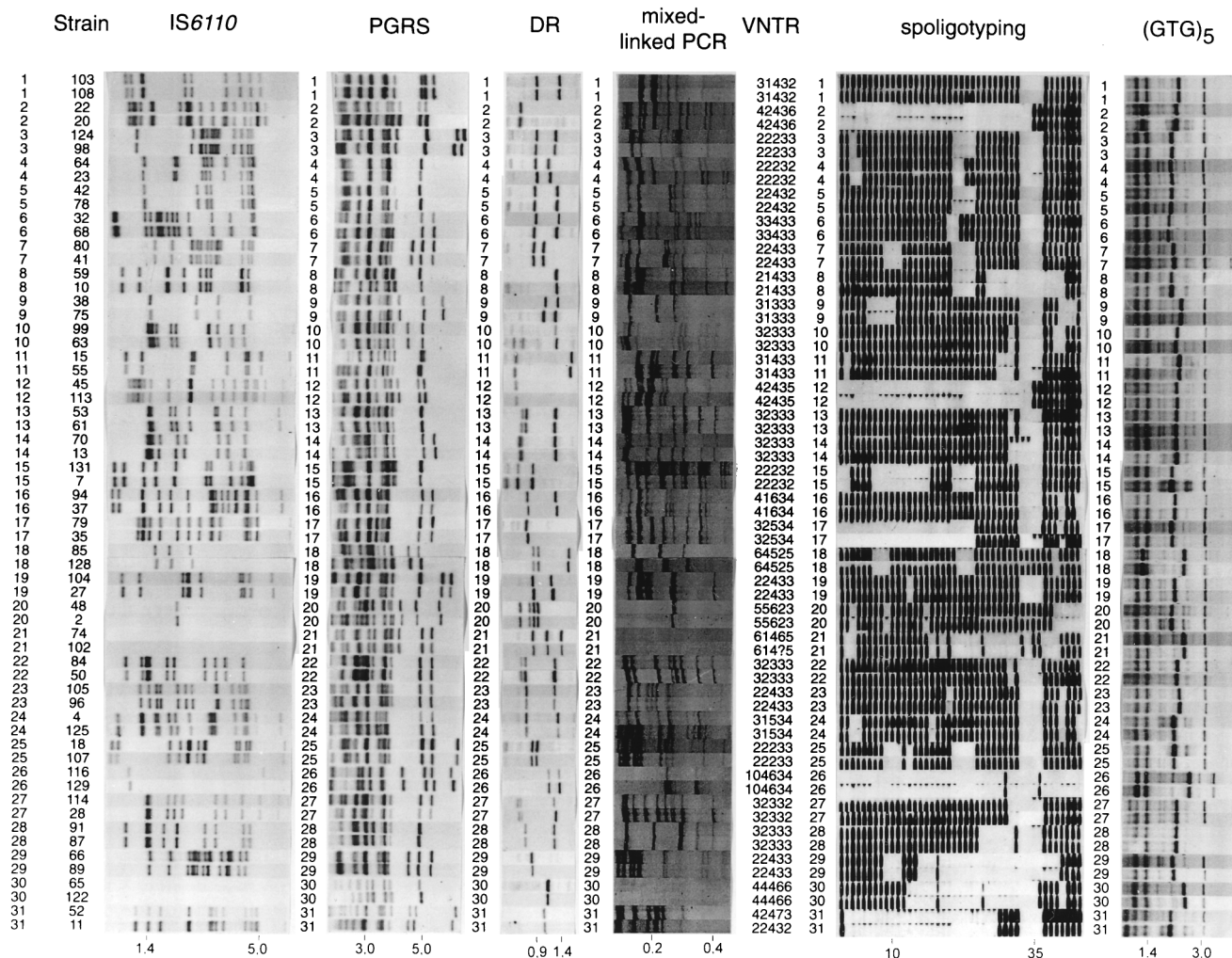


FIG. 1. Patterns of the 31 *M. tuberculosis* complex duplicate samples obtained by various typing methods. The methods are ordered from most to least reproducible (left to right). The IS6110 RFLP patterns are shown as representatives of the *Pvu*II-based RFLP methods. Numbers in vertical direction indicate duplicate samples. Numbers at the bottom indicate sizes of standard DNA fragments in kilobases, except for spoligotyping where the numbers indicate the spacer oligonucleotides.

scribed (22, 56). The results of IS6110, IS1081, PGRS, (GTG)₅, DR, *mtp40*, *mpb64*, *katG*, and 16S rDNA gene RFLP and RFEL analyses were analyzed by computer by a single person at the organizing laboratory. The remaining methods were analyzed by computer at the laboratory that performed the respective typing. Normalization of RFLP patterns of IS6110, IS1081, PGRS, (GTG)₅, and DR was accomplished by using internal markers as described by Van Embden et al. (48). Spoligotype, RFEL, *mtp40*, *mpb64*, *katG*, and 16S rDNA gene banding patterns were normalized by using *M. tuberculosis*-specific bands present in at least five lanes per gel. Mixed-linker PCR and IS6110 inverse PCR patterns were normalized by using a 100-bp ladder and lambda-*Hind*III/*phi*X174-*Hae*III as external markers (at least three external markers per gel), respectively. IS6110 amplicon patterns were analyzed with the BioImage whole-band analyzer (Bio-Image, Ann Arbor, Mich.) by using an upper and a lower DNA standard that cross-hybridized with the IS6110 probe as external markers (39). VNTR typing patterns were analyzed by the Excel program (Microsoft Co., Redmond, Wash.). DRE-PCR and APPCR patterns were analyzed visually.

In general, bands were automatically assigned by the computer and corrected manually after checking the original autoradiogram or photograph by eye. For instance, if a dark intensity band was recognized by the computer as one band and two bands were visible by eye on an autoradiogram of a shorter exposure, then one additional band was added manually to the computerized banding pattern. For DR and (GTG)₅ fingerprinting, only bands of high intensity were used for analysis, because the low-intensity bands of the reference strains were not reproducible. Comparison of the banding patterns was done by using the Dice coefficient to calculate similarities. The unweighted pair group method using arithmetic averages was used for clustering. The position tolerance (i.e., the tolerance in band position that is allowed between patterns deemed identical by the computer) differed slightly for each method and was determined by matching the reference strains that were present on each gel. For most typing methods, a band

position tolerance of 1% was sufficient for 100% matching of the banding patterns of the reference strains. For mixed-linker PCR, IS6110 inverse PCR, and (GTG)₅ and IS1081 RFLPs, however, a larger tolerance setting of 1.2% was required.

Statistical analysis. Associations were tested for statistical significance by using the χ^2 test. *P* values lower than 0.05 were considered statistically significant.

RESULTS

Reproducibility of typing methods. Eight different laboratories that specialized in the various methods typed the blinded set of 131 mycobacterial DNA samples (set A) by 12 typing methods, and the organizing laboratory evaluated the results obtained. Set A comprised 90 different *M. tuberculosis* complex strains, 10 non-*M. tuberculosis* complex strains, and 31 duplicate samples. Figure 1 shows the banding patterns of the duplicate samples obtained in the typing methods used. All RFLP typing methods, except (GTG)₅ RFLP typing, were 100% reproducible (Table 2). Because the *Pvu*II-digested DNA was hybridized with both IS6110 and -1081 and both methods were 100% reproducible, the IS1081 RFLP patterns are not shown in Figure 1. (GTG)₅ fingerprints usually contained many weak bands which were difficult to score. Therefore, only high-intensity bands were used for the analysis. Two pairs of duplicate strains were interpreted differently; therefore, the reproduc-

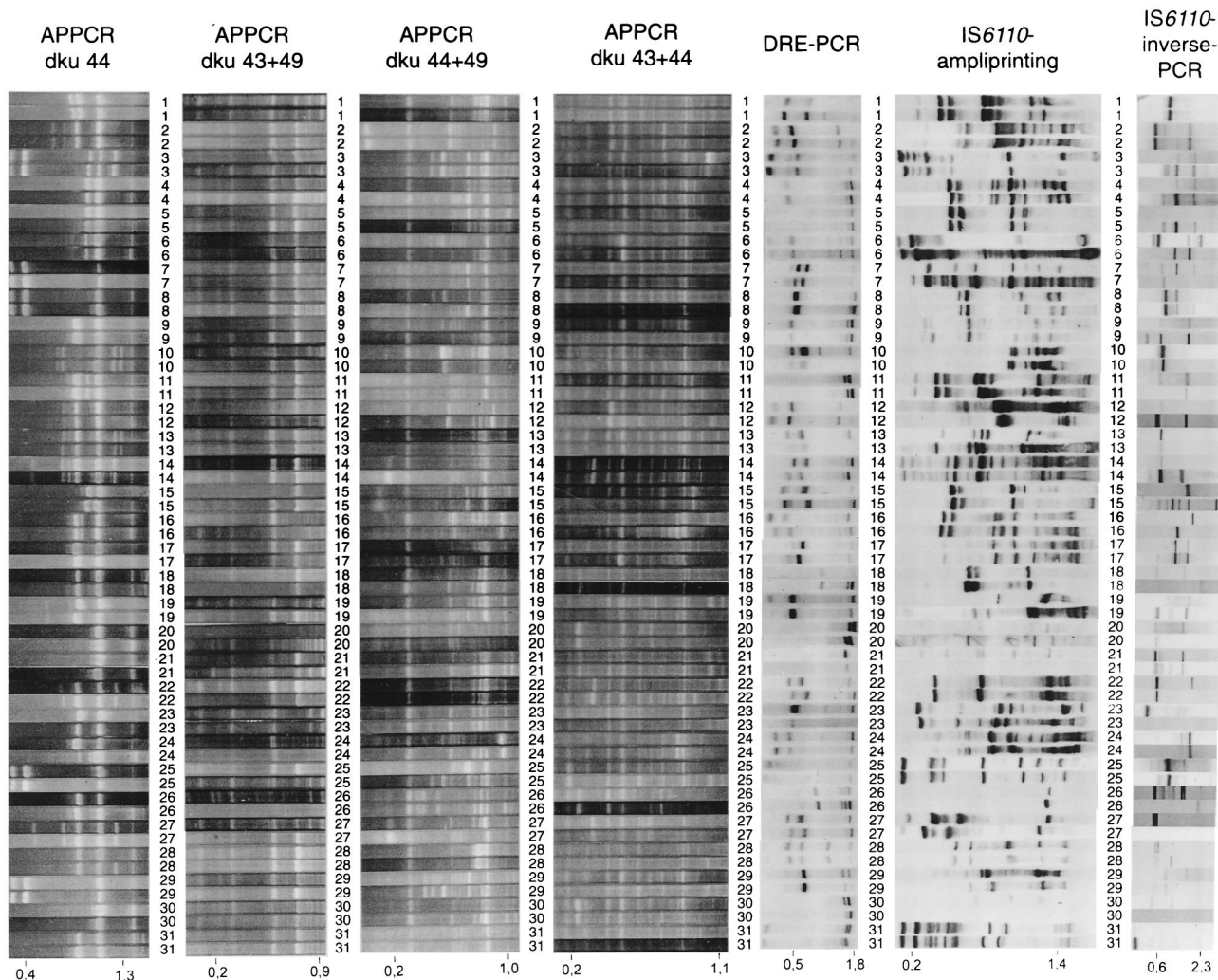


FIG. 1—Continued.

ibility of (GTG)₅ RFLP typing amounted to 94% (Fig. 1, duplicates 27 and 28). PGRS fingerprints were also difficult to analyze by computer alone, due to the high density of bands and variability in band intensities. Therefore, these banding patterns were analyzed by eye, after a preliminary computer-based ordering on the basis of similarity. The PGRS RFLP patterns of the duplicate samples were interpreted as identical in all cases (100%).

Of the PCR-based methods, only the mixed-linker PCR was 100% reproducible. The reproducibility of the other PCR-based methods ranged from 97 to 6% (Table 2). Only one pair of duplicate samples showed discordant results after VNTR typing (97% reproducible). The profiles observed for that pair of duplicate samples were 22432 and 42473, thus differing in four of the five alleles (duplicate 31). Furthermore, a single allele was not typeable in one strain because of a negative PCR amplification of the ETR-D locus. The laboratory did not score this result as discordant because they considered missing data to be the most parsimonious result (Fig. 1, duplicate 21).

In spoligotyping, 29 out of the 31 duplicate samples were scored identically (94%). One discordant result was due to a computer error that was not corrected manually; the computer found a band for spacer 37 due to the strong reaction of that spacer in the adjacent strain (Fig. 1, duplicate 29). The other

mismatch occurred because the hybridization signal of two spacers was weak, and therefore these spacers were interpreted differently in the computer analysis (Fig. 1, spacers 18 and 26 of duplicate 30).

APPCR typing was done by using four primer pairs. The reproducibilities of the primer combinations differed significantly (Fig. 1). The most reproducible results were found with primer DKU44, followed by primer pairs DKU43 plus DKU49, DKU44 plus DKU49, and DKU43 plus DKU44. The use of primer DKU44 resulted in nonidentical patterns among five duplicates (duplicates 23, 24, 25, 26, and 28). Use of the combination of the four primer pairs resulted in an overall reproducibility of 71%.

Similar to APPCR, the reproducibilities of DRE-PCR (58%) and IS6110 amplifrinting (39%) were poor due to difficulties in the scoring of low-intensity bands and differences in the intensities of the whole patterns (Fig. 1, duplicates 1, 2, 10, 12, 13, 18, and 23 for DRE-PCR and 6, 7, 9, 11, 19, and 23 for IS6110 amplifrinting). Some discrepancies in these data were due to the application of duplicate fingerprints on separate gels. The laboratories occasionally scored patterns of duplicate DNA samples differently, although later evaluation by the organizing laboratory confirmed these samples to be identical.

The IS6110 inverse PCR method (reproducibility, 6%) re-

TABLE 2. Number of types and reproducibility obtained by using different typing methods and housekeeping genes for differentiation of 90 *M. tuberculosis* complex strains and 10 non-*M. tuberculosis* complex strains

DNA target	Method used	Reproducibility (%) ^a	No. of types obtained	Avg no. of bands detected	No. of nontypeable strains	No. of non- <i>M. tuberculosis</i> complex strains
IS6110	RFLP (<i>Pvu</i> II)	100	84	10	2	0
IS6110	Mixed-linker PCR	100	81	8	2	3
IS6110	IS6110 inverse PCR	6	ND ^b	4	1	9
IS6110/MPTR	IS6110 ampliprinting	39	ND	7	5	5
IS6110/PGRS	DRE-PCR	58	63	5	0	4
PGRS	RFLP (<i>Alu</i> I)	100	70	9	0	8
DR locus	Spoligotyping	94	61	28	0	0
DR locus	RFLP (<i>Alu</i> I)	100	48	2	0	1
ETRs A-E	VNTR typing	97	56	5	0	ND
(GTG) ₅	RFLP (<i>Hinf</i> I)	94	30	4	0	10
Total genome	APPCR	71	71	8 ^c	0	10
IS1081	RFLP (<i>Pvu</i> II)	100	12	6	0	1
<i>mtp40</i>	RFLP (<i>Pvu</i> II)	100	12	3	0	0
<i>mpb64</i>	RFLP (<i>Pvu</i> II)	100	8	1	1	0
<i>katG</i>	RFLP (<i>Pvu</i> II)	100	5	2	0	0
16S rRNA gene	RFLP (<i>Pvu</i> II)	100	5	1	0	10

^a Fraction of duplicates showing identical types (31).

^b ND, not done.

^c The numbers of bands of four primer combinations were combined.

sulted in highly dissimilar patterns among the duplicate samples (e.g., duplicates 6, 7, 15, 16, and 26) (Fig. 1).

The most highly reproducible techniques were the RFLP-based methods, mixed-linker PCR, VNTR typing, and spoligotyping. The reproducibility of these methods, ranging from 100 to 94%, did not differ significantly ($P > 0.1$). The next most reproducible method, APPCR (reproducibility, 71%), was significantly less reproducible ($P < 0.05$).

Degree of differentiation of the various typing methods. The 90 *M. tuberculosis* complex strains of set A were subjected to 12 typing methods, in which eight repetitive DNA sequences were targeted. The number of different types obtained in each method is depicted in Table 2. The highest degree of discrimination was obtained with IS6110 RFLP and mixed-linker PCR, yielding 84 and 81 types, respectively. APPCR, PGRS RFLP, DRE-PCR, spoligotyping, VNTR typing, and DR RFLP resulted in 71, 70, 63, 61, 56, and 48 types, respectively. It should be noted, however, that the reproducibilities of APPCR and DRE-PCR were significantly less than those of the other typing methods. Due to the poor reproducibility of IS6110 ampliprinting and IS6110 inverse PCR, the number of types obtained with these methods is not given. (GTG)₅ RFLP differentiated 30 types. As expected, the level of discrimination obtained by using IS1081 was very limited (52, 54). In addition to probes frequently used as genetic markers for strain differentiation, we also hybridized the genomic *Pvu*II blots with DNA from the housekeeping genes *mtp40*, *mpb64*, *katG*, and 16S rRNA. The number of RFLP types obtained with these probes ranged from five for the 16S rRNA gene probe to 12 when the *mtp40* sequence was used as a probe (Table 2) (Fig. 2).

Although IS6110 RFLP and mixed-linker PCR exhibited the highest differentiation levels, these methods were, for apparent reasons, less useful for typing strains containing few IS6110 copies, confirming previous observations (42, 51, 52). These IS6110-based methods both grouped two and five *M. bovis* strains in two clusters. VNTR typing of these strains yielded six types. In PGRS RFLP, DR RFLP, and spoligotyping, these strains were differentiated second best, yielding five types, followed by (GTG)₅ RFLP (4 types), IS1081 RFLP (3 types), and *mpb64* RFLP (2 types).

In IS6110 RFLP typing, IS6110 ampliprinting, and mixed-

linker PCR, two strains yielded negative results because there was no IS6110 element present in either genome. These strains were therefore nontypeable. In IS6110 ampliprinting, two *M. bovis* and one other *M. tuberculosis* strain were also nontypeable. IS6110 inverse PCR failed to recognize only one strain (*M. bovis* BCG Japan). This method yielded positive results for the two strains lacking IS6110. All the other typing methods yielded signals for all strains (Table 2) (Fig. 2). The Dutch *M. bovis* BCG (strain P3) lacked the *mpb64* gene, consistent with the findings of Li et al. (33).

Figure 2 shows the fingerprints obtained in RFLP typing on the basis of IS6110, PGRS, DR, (GTG)₅, IS1081, *mtp40*, *mpb64*, *katG*, and the 16S rDNA gene and with spoligotyping and VNTR typing. Because IS6110 was found to be the most discriminating genetic marker and because the banding patterns obtained with this element contained, on average, more distinct bands than patterns obtained by using the other genetic markers, we ordered the 90 *M. tuberculosis* strains on the basis of similarity by using IS6110 RFLP. The mixed-linker PCR patterns showed a grouping similar to that of the IS6110 RFLP patterns (data not shown). Figure 2 shows that grouping *M. tuberculosis* strains on the basis of IS6110 RFLP results showed a similar grouping of patterns based on other typing results. Strains within three of the major IS6110 RFLP groupings showed very limited variation as measured by the other genetic markers, thus hampering strain differentiation by these markers.

A group of eight *M. tuberculosis* strains, originating from four Southeast Asian countries and South Africa (Fig. 2, strains 43, 45, 54, 30, 14, 44, 20, and 34) were identical when evaluated with six genetic markers. The VNTR types of these eight strains were all identical, except for a single strain that differed in one allele (ETR-E). In PGRS RFLP typing, five similar patterns were differentiated (similarity, >49%), and in both IS1081 and (GTG)₅ RFLP, two types were differentiated. The IS6110 fingerprints of these eight strains were clearly distinct, but these patterns shared at least 67% of their IS6110-containing restriction fragments (Fig. 2). This group of strains, "the Beijing family," has previously been recognized as the most prevalent group in Southeast Asia (56).

A second homogenous group was composed of 13 strains

which shared at least 54% similarity on the basis of their IS6110 RFLP patterns. VNTR typing and RFLP typing using IS1081, *mtp40*, *mpb64*, *katG*, and the 16S rDNA gene as probes resulted in patterns of these 13 strains which were virtually identical, with only one or two strains showing a different pattern. Five spoligotypes, still sharing at least 94% similarity, were detected among these strains. Typing with (GTG)₅, PGRS, and DR RFLPs showed more variation: 55, 52, and 38% similarities, respectively (Fig. 2, strains 8, 60, 58, 13, 51, 123, 33, 53, 86, 28, 50, 63, and 87). These isolates originated from eight different countries in the Americas, Asia, and Europe, and therefore this clone of *M. tuberculosis* strains seems to be widespread globally. We designated this *M. tuberculosis* genotype the "Haarlem family," because the first recognized strain was isolated from a patient living in Haarlem, The Netherlands.

The third group of strains ($n = 8$) also shared a high degree of similarity among the banding patterns generated with most genetic markers. Because all eight strains of this group originated exclusively from central Africa, we designated this group "the Africa family" (Fig. 2, strains 37, 97, 4, 40, 35, 72, 120, and 121). Although the grouping by VNTR typing and spoligotyping was less obvious for these strains than for the Beijing and the Haarlem families, the similarities of the IS6110, PGRS, DR, and (GTG)₅ banding patterns were 52, 52, 51, and 77%, respectively. Virtually no polymorphism was observed by IS1081, *mtp40*, *mpb64*, *katG*, and 16S rDNA gene RFLPs.

In addition to the clinical isolates, the *M. tuberculosis* laboratory strains H37Rv and the isogenic avirulent derivative H37Ra were included in set A. These two strains have been cultured separately for about 7 decades (46). As shown in Fig. 2 (strains 32 and 109), the IS6110 banding patterns were distinct; only 11 of the 14 bands matched. Surprisingly, no difference was found between H37Rv and H37Ra by any of the other genetic markers.

Specificity of typing methods for *M. tuberculosis* complex strains. All typing methods were performed on the mycobacterial strains of set A, which also included 10 non-*M. tuberculosis* complex strains of five distinct mycobacterial species. The sequences IS6110, *mtp40*, *mpb64*, and *katG* did not cross-hybridize with non-*M. tuberculosis* complex strains in RFLP (Table 2). Exclusively *M. tuberculosis* complex strains were also detected by spoligotyping. The PCR-based methods targeting IS6110, however, were less specific for the *M. tuberculosis* complex. With these methods, positive results were obtained for three to nine of the non-*M. tuberculosis* complex strains (Table 2). In DR RFLP, a weak reaction with an *M. gordonae* strain was observed. (GTG)₅ RFLP and APPCR yielded high-intensity patterns for the non-*M. tuberculosis* complex strains, but these were distinct from the patterns obtained for the *M. tuberculosis* complex strains.

Generic typing methods. We analyzed a set of 17 strains (set C) composed of 12 *M. tuberculosis* complex strains and five non-*M. tuberculosis* complex species with MLEE, RFEL analysis, and multilocus sequencing.

By MLEE, 14 of the 18 enzymes assayed for the 12 *M. tuberculosis* complex isolates were monomorphic. Malate dehydrogenase and alanine dehydrogenase showed two alleles, glutamate oxaloacetic transaminase showed three alleles, and isocitrate dehydrogenase showed four alleles. In total, eight distinctive allele profiles or ETs were distinguished (data not shown). The mean genetic diversity per locus amounted to 0.12, and the genetic distance between any of the 12 *M. tuberculosis* complex strains was less than 0.14. These results reinforce the extreme genetic homogeneity of *M. tuberculosis* complex strains, in spite of the selection of the *M. tuberculosis* strains of set C on the basis of their display of the most diverse

IS6110 fingerprints and their diverse geographic origins. The genetic distance between the non-*M. tuberculosis* complex strains (*M. smegmatis*, *M. kansasii*, *M. gordonae*, *M. xenopi*, and *M. avium*) was at least 0.81 (data not shown). All enzymes were polymorphic, with two to five different mobilities per enzyme. The mean diversity per allele amounted 0.88 for the non-*M. tuberculosis* complex strains.

In RFEL analysis (58), end-labeled restriction fragments are fractionated on a high-resolution polyacrylamide gel, and therefore, more fragments can be analyzed in RFEL than in restriction enzyme analysis on agarose gels after ethidium bromide staining (5). On average, 50 restriction fragments were detected for each of the 12 *M. tuberculosis* complex strains, and at least 44 fragments were shared (data not shown). This high degree of similarity is comparable with the values for restriction enzyme analysis reported by Collins and de Lisle for *M. bovis* isolates (5). The similarity among the 12 *M. tuberculosis* complex strains of set C amounted to 0.94 or more, as measured by the Dice coefficient, whereas the similarity among the non-*M. tuberculosis* complex strains ranged from 0.53 to 0.75.

The strains of set C were subjected to a third generic method: multigene sequencing. The PCR amplifications of most of the non-*M. tuberculosis* complex strains were negative, illustrating the differences in DNA sequence between *M. tuberculosis* and non-*M. tuberculosis* complex strains (data not shown). All *M. tuberculosis* strains of set C were identical by their 16S rRNA gene sequences and by the DNA sequence encoding the 16-kDa antigen. One *M. tuberculosis* strain showed a point mutation in *ideR* (strain 43, 628 C→A). One *M. africanum* and one *M. bovis* strain showed a point mutation in *aroA* (strain 92, 212 A→G; strain 126, 654 G→A). The *aroA* sequence of *M. canettii* was exceptional in its display of seven base pair substitutions (strain 129, -48 T→C; 766 C→G; 835 C→G; 859 A→G; 945 G→C; 984 G→A; 977 A→G). *oxyR* showed single-base-pair substitutions among four strains (strains 47 and 74, C→T; strains 48 and 126, 285 G→A). In summary, among the 12 *M. tuberculosis* complex strains, seven multilocus sequence types were observed. These data confirm the previous observations of the high level of genetic relatedness within the *M. tuberculosis* complex and the distinctness of the complex from other mycobacterial species (13, 15, 30) and the distant relatedness of *M. canettii* to the other members of the *M. tuberculosis* complex (55).

DISCUSSION

This study shows that strain typing of *M. tuberculosis* complex strains by RFLP was reproducible for all 31 duplicate samples tested, irrespective of the DNA probe used. Recommendations for a standardized method of strain typing for *M. tuberculosis* were formulated in 1993 (48), and in virtually all published studies, investigators have complied with these recommendations (1, 44, 61). Therefore, the results of this study suggest that fingerprints generated in different laboratories can be compared, thus allowing the investigation of the prevalence of various types or genotypes in different regions and the trace of interregional transmission of *M. tuberculosis*. It should be noted, however, that in this study, duplicate samples were analyzed in a single laboratory. In a recent study on the interlaboratory reproducibility of IS6110 fingerprinting, we disclosed large differences in the quality of DNA fingerprints among laboratories due to differences in resolution, in the use of reference markers, and in computer-assisted analysis (31). One may therefore expect that the potential to compare *M. tuberculosis* RFLP patterns between different laboratories is restricted, in particular when large numbers of fingerprints are

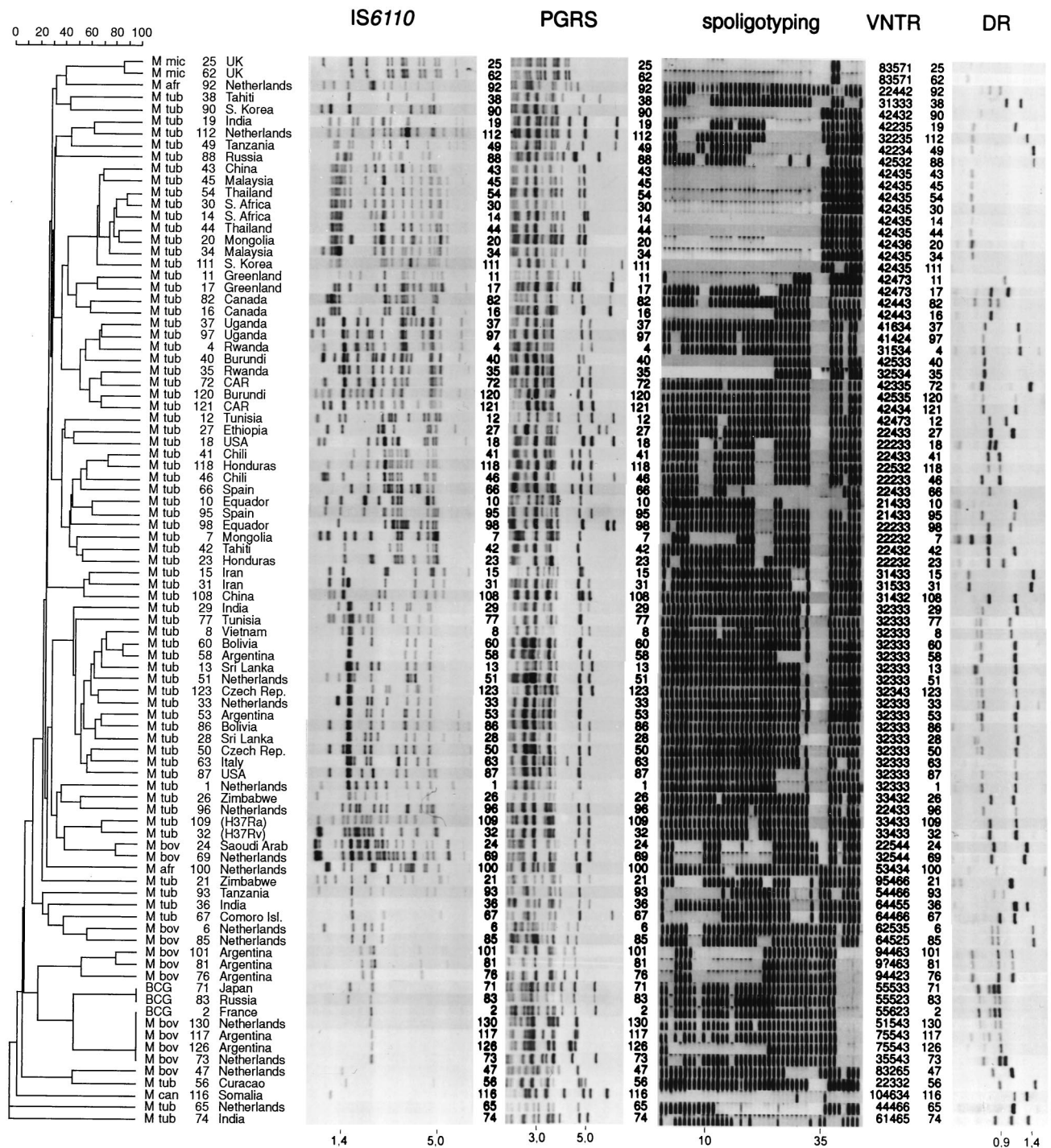


FIG. 2. Patterns of the 90 *M. tuberculosis* complex strains of set B obtained by using various genetic markers. The methods are ordered from most to least discriminative (left to right). The genetic distance of the strains as depicted in the dendrogram is based upon IS6110 RFLP analysis. The mixed-linker PCR patterns are not shown, as they show a similar grouping as IS6110 RFLP. The banding patterns of the nonreproducible typing methods are also not shown. Numbers in vertical direction indicate strain numbers. Numbers at the bottom indicate sizes of standard DNA fragments in kilobases, except for spoligotyping where the numbers indicate spacer oligonucleotides. M mic, *M. microti*; M afr, *M. africanum*; M tub, *M. tuberculosis*; M bov, *M. bovis*; BCG, *M. bovis* BCG; M can, *M. canettii*.

compared. IS6110 RFLP patterns are relatively easy to compare, because they generally show hybridizing bands of equal intensity, as most IS6110-containing restriction fragments carry an intact part of an IS6110 copy of equal size. In contrast, RFLP typing using repetitive DNA such as PGRS or (GTG)₅ as a probe results in hybridizing bands of variable intensity. In

this study, we noticed that computer analysis of such banding patterns requires additional analysis by eye.

Four of the seven PCR-based typing methods tested were unreliable, showing reproducibility values ranging from 6 to 71%. This is unfortunate, because these methods were the easiest and quickest to perform. The reasons for irreproduc-

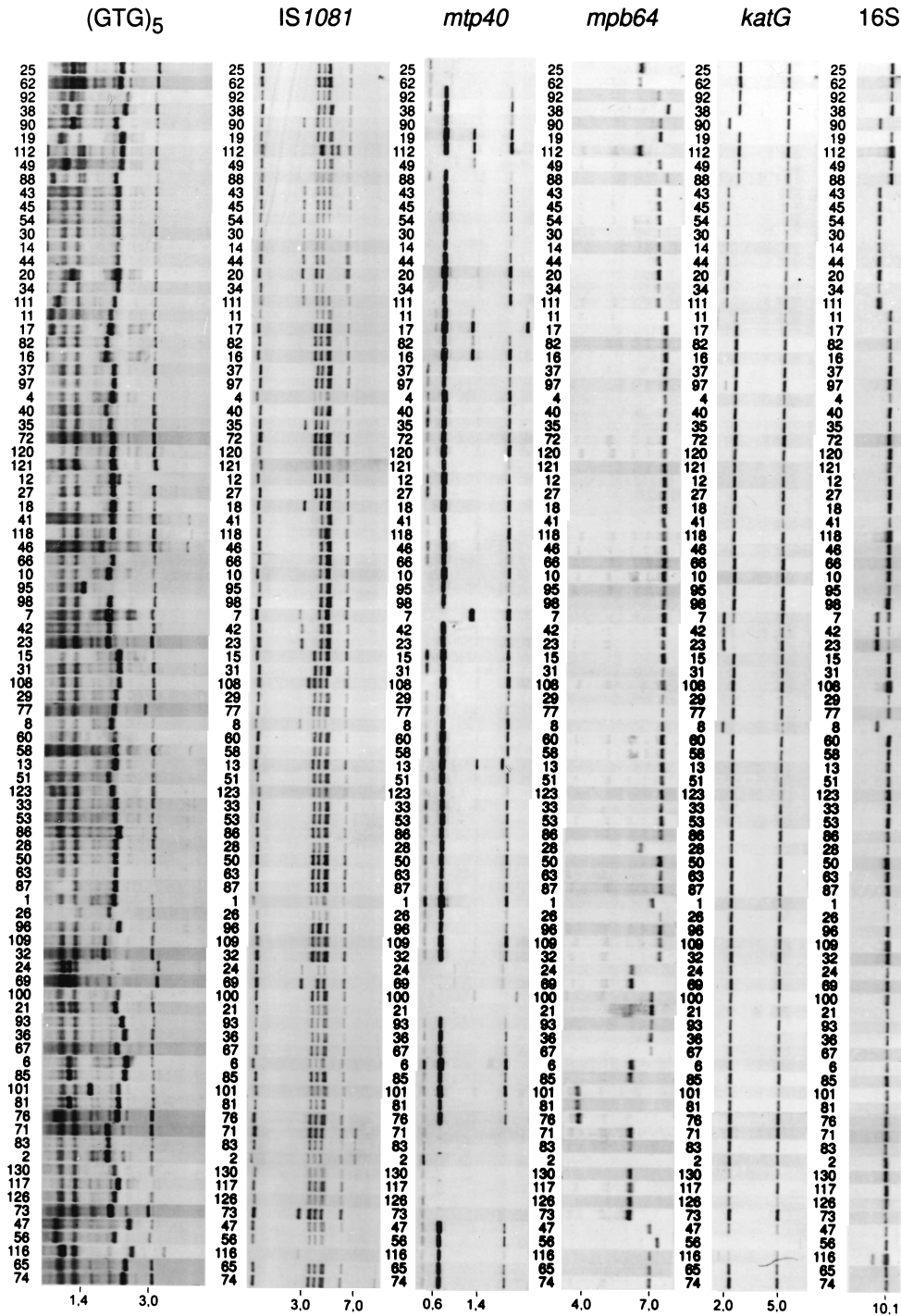


FIG. 2—Continued.

ibility were differences in the intensity of whole banding patterns which resulted in the disappearance of weaker bands and the inconsistent interpretation of bands of weak intensity. Completely different banding patterns were also occasionally generated. Some of the discrepant results were obtained because laboratories failed to recognize identical patterns. Mixed-linker PCR, VNTR typing, and spoligotyping were found to be reproducible. The errors found by the latter two methods were due to mistakes that are not inherent in these techniques.

Mislabeled or cross contamination seemed likely to be responsible for a discordant result obtained in VNTR typing, and a computer misreading caused an error in spoligotyping. Therefore, mixed-linker PCR, VNTR typing, and spoligotyping are the methods of choice for reproducible PCR-based typing of *M. tuberculosis*. Moreover, VNTR typing and spoligotyping have the advantage that their results can be fully expressed in a simple, digital format. Thus, inter- and intralaboratory comparisons should be equally reproducible.

The differences in reproducibility of mixed-linker PCR, VNTR typing, spoligotyping, and the less-reproducible PCR-based methods may be partly explained by differences in specificity of priming during PCR amplification. In the three methods that were found to be reproducible, the target for PCR priming is expected to be nonvariable. Furthermore, it has been demonstrated by sequencing or hybridization of internal sequences that the PCR products obtained by these methods were indeed the expected targets (16, 20, 27). In contrast, the PCR products generated by IS6110 inverse PCR, IS6110 ampliprinting, and DRE-PCR have not been characterized. One may expect that the poor reproducibilities of IS6110 ampliprinting and DRE-PCR are due to arbitrary priming caused by variation in the target sequences. It should be noted that we investigated the best-case scenario for reproducibility, as the reproducibility was determined by analysis of completely identical DNA samples. It is to be expected that, in practice, variations in DNA preparation (e.g., purity, size, concentration, and presence of inhibitors) may decrease the reproducibility of any of the methods.

The discriminative power of various typing methods was determined by using a set of 90 *M. tuberculosis* complex strains. On basis of the number of types obtained, the methods could be ordered as follows from most discriminating to least discriminating: IS6110 RFLP, mixed-linker PCR, APPCR, PGRS RFLP, DRE-PCR, spoligotyping, VNTR typing, DR RFLP, (GTG)₅ RFLP, and IS1081 RFLP or *mtp40* RFLP. We conclude that for epidemiological investigations, strain differentiation by IS6110 RFLP or mixed-linker PCR are the methods of choice. When less strain discrimination is required, VNTR typing and spoligotyping are reproducible alternatives. The discriminatory power of some methods might increase if the methods are improved. Because multiple tandem repeats of the PGRS are present in about 100 genes in the *M. tuberculosis* genome (4), one may expect that more types would be distinguished by PGRS RFLP typing if more PGRS-containing restriction fragments were resolved electrophoretically, a result which could be achieved by using longer agarose gels (8). Spoligotyping may differentiate better when the location of more spacers is investigated (27), and the differentiation level of VNTR typing may increase if more VNTR loci are investigated (16). The discriminative power of (GTG)₅ RFLP was previously described as being superior to IS6110 RFLP (60), but in this study we found only 30 different (GTG)₅ patterns among 90 strains, compared to 84 patterns by IS6110 RFLP. In the (GTG)₅ fingerprints, we observed many faint hybridizing bands which were not reproducible and therefore not usable for reliable strain typing.

The set of 90 strains used to investigate the potential discrimination of *M. tuberculosis* complex strains contained 18 strains with fewer than five IS6110 copies each. Such strains are frequently encountered among *M. bovis* isolated from all regions and among *M. tuberculosis* isolated in Asia. This study confirmed previous observations that such strains are poorly differentiated by IS6110-based typing methods (42, 51, 52). Therefore, typing methods using other targets, such as PGRS, (GTG)₅, or the DR locus, are more discriminative for such strains (51). The methods that allowed subtyping of the two low-copy-number IS6110 clusters of two and five *M. bovis* strains, in decreasing order of discrimination, were as follows: VNTR typing, PGRS RFLP or DR RFLP, spoligotyping, (GTG)₅ RFLP, IS1081 RFLP, and *mpb64* RFLP. The four *M. bovis* strains of bovine origin containing one or two IS6110 copies and the three vaccine strains we investigated in this study were only identical in *mtp40*, *katG*, and 16S rDNA RFLPs. The DNA polymorphism obtained with other genetic

markers was relatively high. The similarities between the patterns obtained by PGRS, DR, and spoligotyping were 22, 30, and 53%, respectively. VNTR typing recognized six different types.

Two of the 90 *M. tuberculosis* strains investigated in this study were devoid of IS6110, and as expected, these were not typeable by IS6110 RFLP and mixed-linker PCR. However, in vitro amplification by the other three PCR-based typing methods using IS6110 as a target sequence resulted in detectable amplicons, indicating nonspecific amplification of DNA. Such an IS6110 nonspecific amplification was also observed with many of the samples from non-*M. tuberculosis* complex strains, which are known to lack IS6110. A similar lack of specificity has been observed previously (29, 34, 36). These two strains without IS6110 sequences exhibited different patterns in PGRS and DR RFLPs, spoligotyping, and VNTR typing, indicating that these strains are not recently derived from a common ancestor.

In order to compare the currently used genetic markers for strain typing of *M. tuberculosis* with generic markers used in population genetics of bacteria, we subjected a small number of *M. tuberculosis* complex strains to analysis by MLEE, RFEL, and multilocus DNA sequencing. We compared the polymorphisms obtained for *M. tuberculosis* complex with those obtained with five nontuberculosis mycobacterial species. The genetic diversity among strains within the *M. tuberculosis* complex was found to be very small, which is in agreement with the observations of others (5, 13, 15, 45). Furthermore, these *M. tuberculosis* complex strains were very distinct from other mycobacterial species. In agreement with a previous study, we confirmed that *M. canettii* is more distantly related to *M. tuberculosis* than to *M. africanum*, *M. bovis*, or *M. microti* (55). Our data clearly show that the traditional methods to reveal polymorphisms among bacterial populations are not useful for strains within the *M. tuberculosis* complex, because of the unusually low structural gene variation among *M. tuberculosis* isolates, even when they originate from very diverse geographic regions.

The RFLP patterns obtained with various markers clearly showed more polymorphisms associated with repetitive DNA such as IS6110, PGRS, DR, and VNTRs than with nonrepetitive DNA such as *katG*, the 16S rDNA gene, *mpt40*, and *mpb64*. This indicates that transposition, homologous recombination, and perhaps slipped-strand mispairing during replication play an important role in the generation of DNA polymorphisms, as disclosed by markers which are frequently used for strain typing. The polymorphisms obtained with the various genetic markers showed, in general, a strong mutual association, suggesting a clonal population structure of *M. tuberculosis* complex strains. Consistently, we found that ordering of the strains by IS6110 RFLP led to similar groupings obtained by other genetic markers, such as VNTR types and polymorphisms in the DR locus. The *M. tuberculosis* complex group of bacteria consists of the following five subspecies: *M. tuberculosis*, *M. africanum*, *M. bovis*, *M. microti*, and *M. canettii*. The latter three subspecies have a species-specific spoligotype signature (27, 55, 57). In this study, however, we found four *M. bovis* strains which did not exhibit an *M. bovis*-specific spoligotype. But these strains were also more like *M. tuberculosis* by other markers. Although the number of strains is limited, our data suggest that *M. tuberculosis* complex species-specific signatures can also be recognized by VNTR typing and perhaps by PGRS typing. In addition, other groupings of genetically related strains within the *M. tuberculosis* complex may also be recognized by the genetic markers used for the epidemiology of tuberculosis.

The existence of the Beijing family of *M. tuberculosis* has

been previously described (56). This group of genetically related strains is highly prevalent in Southeast Asia, and strains belonging to this grouping did not only show exceptional IS6110 RFLP and spoligotyping patterns (56), but in this study, they also appeared to be recognizable by other typing methods. The Beijing strains in this study originated from Asia and South Africa and showed identical spoligotypes and VNTR types, distinct from all other strains. Furthermore, their IS6110 banding patterns were highly similar and very little or no polymorphism was observed with the other genetic markers. These data indicate that strains belonging to this group are genetically very closely related and distinct from other *M. tuberculosis* strains. Therefore we assume that, although strains of this group disseminated globally, they expanded clonally from a recent common ancestor. The "W" strain, which caused a large outbreak of multidrug-resistant tuberculosis in New York City and other U.S. cities, also belongs to the Beijing family (32). Two other, not previously described, genotype groups could be recognized by the shared polymorphisms in the various typing methods. One group was composed of 13 strains originating from eight countries in Asia, Europe, and the Americas. Organisms in this group showed at least 54% similarity in their IS6110 RFLP patterns. When analyzed by VNTR typing and spoligotyping, these strains were nearly identical and distinct from other strains. This group will be designated provisionally as the Haarlem family. The remaining group, composed of eight strains, also shared polymorphisms with many genetic markers. These strains originated from central Africa and therefore we designated this group provisionally as the Africa family.

Recently, Sreevatsan et al. have suggested the division of *M. tuberculosis* complex strains in three distinct genotypic groups based on the combination of catalase-peroxidase (*katG463*) and gyrase (*gyrA95*) gene sequences (45). In the evolutionary pathway they proposed, genotypic group 1 organisms include *M. africanum*, *M. bovis*, *M. tuberculosis*, and *M. microti* and groups 2 and 3 only include *M. tuberculosis*. Group 1 organisms are expected to carry a higher level of genetic diversity than those in groups 2 and 3 (45). As the W strain belongs to genotypic group 1 (45) and to the Beijing family (32), we assume that all the members of this family represent genotypic group 1. Also, strains devoid of IS6110 were of group 1. Furthermore, *M. tuberculosis* H37Ra and H37Rv belonged to genotypic group 3 (45). Based on the polymorphisms observed in this study, we found that strains devoid of IS6110 and those belonging to the Beijing family are more diverse than the two *M. tuberculosis* reference strains. These two *M. tuberculosis* reference strains were the only two strains that were identical by 10 of the 11 investigated genetic markers. The Beijing family, however, was the most conserved of the three major *M. tuberculosis* genotype families in this study. Further research should be done to elucidate the association between the polymorphisms observed in sequence analysis and those observed by other genetic markers.

This study has shown that markers used in strain typing for the epidemiology of tuberculosis may lead to the recognition of well-defined genotype families within the *M. tuberculosis* complex. Given the enormous potential of gene variation by the abundance of repetitive polymorphic DNA in about 10% of the *M. tuberculosis* Pro-Glu and Pro-Pro-Glu genes (4), one may expect that these recently evolved genotype families share many of these potentially variable genes. Strains belonging to these genotype families possibly share particular phenotypic properties, such as antigens and virulence factors, which may be expressed as distinct manifestations in the pathology and the epidemiology of tuberculosis.

ACKNOWLEDGMENTS

All collaborators who supplied us with strains are gratefully acknowledged. We thank Annelies Bunschoten for performing spoligotyping; Annette de Boer, Petra de Haas, and Herre Heersma for support in computer analysis; Saskia Jansen for (GTG)₅ RFLP typing; Tridia van der Laan for excellent laboratory assistance; Marcel Sluijter for RFEL analysis; and Percy L. Strickland and Alison J. Cobb for VNTR typing. Simone van de Pas and Frans van den Berg are acknowledged for preparation of figures.

This study was financially supported by the European Community Program for Biomedical Research, Science Technology and Development (grant BMH4-CT97-91202). Research in the laboratory of Jim Musser was supported by Public Health Services grants DA-09238 and AI-370040.

REFERENCES

- Alland, D., G. E. Kakut, A. R. Moss, R. A. McAdam, J. A. Hahn, W. Bosworth, E. Drucker, and B. R. Bloom. 1994. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. *N. Engl. J. Med.* **330**:1710-1716.
- Beck-Sagué, C., S. W. Dooley, M. D. Hutton, J. Otten, A. Breeden, J. T. Crawford, A. E. Pitchenik, C. Woodley, G. Cautchen, and W. R. Jarvis. 1992. Hospital outbreak of multidrug-resistant *Mycobacterium tuberculosis* infections. Factors in transmission to staff and HIV-infected patients. *JAMA* **268**:1280-1286.
- Bergmans, A. M. C., J. Groothedde, J. F. P. Schellekens, J. D. A. van Embden, J. M. Ossewaarde, and L. M. Schouls. 1995. A comparative study on the etiology of cat scratch disease: PCR detection of *Bartonella* (formerly *Rochalimaea*) and *Aptipia felis* DNA, serology and skin testing. *J. Infect. Dis.* **171**:916-923.
- Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry III, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McClean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537-544.
- Collins, D. M., and G. W. de Lisle. 1984. DNA restriction endonuclease analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BGC. *J. Gen. Microbiol.* **130**:1019-1021.
- Collins, D. M., and D. M. Stephens. 1991. Identification of insertion sequence, IS1081, in *Mycobacterium bovis*. *FEMS Lett.* **83**:11-16.
- Coronado, V. G., C. M. Beck-Sagué, M. D. Hutton, B. J. Davis, P. Nicholas, C. Villareal, C. L. Woodley, J. O. Kilburn, J. T. Crawford, T. R. Frieden, R. L. Sinkowitz, and W. R. Jarvis. 1993. Transmission of multidrug-resistant *Mycobacterium tuberculosis* among persons with human immunodeficiency virus infection in an urban hospital: epidemiologic and restriction fragment length polymorphism analysis. *J. Infect. Dis.* **168**:1052-1055.
- Cousins, D. V., S. N. Williams, B. C. Ross, and T. M. Ellis. 1993. Use of a repetitive element isolated from *Mycobacterium tuberculosis* in hybridization studies with *Mycobacterium bovis*: a new tool for epidemiological studies of bovine tuberculosis. *Vet. Microbiol.* **37**:1-17.
- Del Portillo, P., L. A. Murillo, and M. E. Patarroyo. 1991. Amplification of a species-specific DNA fragment of *Mycobacterium tuberculosis* and its possible use in diagnosis. *J. Clin. Microbiol.* **29**:2163-2168.
- Edlin, B. R., J. I. Tokars, M. H. Grieco, J. T. Crawford, J. Williams, E. M. Sordillo, K. R. Ong, J. O. Kilburn, S. W. Dooley, K. G. Castro, W. R. Jarvis, and S. D. Holmberg. 1992. An outbreak of multidrug-resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. *N. Engl. J. Med.* **326**:1514-1521.
- Eisenach, K. D., J. T. Crawford, and J. H. Bates. 1986. Genetic relatedness among strains of the *Mycobacterium tuberculosis* complex: analysis of restriction fragment heterogeneity using cloned DNA probes. *Am. Rev. Respir. Dis.* **133**:1065-1068.
- 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.
- Feizabadi, M. M., I. D. Robertson, D. V. Cousins, and D. J. Hampson. 1996. Genomic analysis of *Mycobacterium bovis* and other members of the *Mycobacterium tuberculosis* complex by isoenzyme analysis and pulsed-field gel electrophoresis. *J. Clin. Microbiol.* **34**:1136-1142.
- Friedman, C. R., M. Y. Stoeckle, W. D. Johnson, Jr., and L. W. Riley. 1995. Double repetitive element PCR method for subtyping *Mycobacterium tuberculosis* clinical isolates. *J. Clin. Microbiol.* **33**:1064-1069.
- Frothingham, R., H. G. Hills, and K. H. Wilson. 1994. Extensive DNA sequence conservation throughout the *Mycobacterium tuberculosis* complex. *J. Clin. Microbiol.* **32**:1639-1643.
- 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.
17. Gordon, S. V., B. Heym, J. Parkhill, B. Barrell, and S. T. Cole. 1999. New insertion sequences and a novel repeated sequence in the genome of *Mycobacterium tuberculosis* H37Rv. *Microbiology* **145**:881–892.
 18. Goyal, M., D. Young, Z. Zhang, P. A. Jenkins, and R. J. Shaw. 1994. PCR amplification of variable sequence upstream of *katG* gene to subdivide strains of *Mycobacterium tuberculosis* complex. *J. Clin. Microbiol.* **32**:3070–3071.
 19. Greifinger, R., J. Grabau, A. Quinlan, A. Loeder, G. DiFerdinando, Jr., and D. L. Morse. 1992. Transmission of multidrug-resistant tuberculosis among immunocompromised persons in a correctional system—New York, 1991. *Morbidity Mortal. Weekly Rep.* **41**:507–509.
 20. Haas, W. H., W. R. Butler, C. L. Woodley, and J. T. Crawford. 1993. Mixed-linker polymerase chain reaction: a new method for rapid fingerprinting of isolates of the *Mycobacterium tuberculosis* complex. *J. Clin. Microbiol.* **31**:1293–1298.
 21. Harris, M. H., and D. A. Hopkinson. 1976. *Handbook of enzyme electrophoresis in human genetics*. Elsevier Science Publishing, Inc., New York, N.Y.
 22. Heersma, H. F., K. Kremer, and J. D. A. van Embden. 1998. Computer analysis of IS6110 RFLP patterns of *Mycobacterium tuberculosis*, p. 395–422. *In* T. Parish and N. G. Stoker (ed.), *Methods in molecular biology*, vol. 101. *Mycobacteria protocols*. Humana Press, Inc., Totowa, N.J.
 23. Hermans, P. W. M., M. Sluijter, T. Hogeboezem, H. Heersma, A. van Belkum, and R. de Groot. 1995. Comparative study of five different DNA fingerprint techniques for molecular typing of *Streptococcus pneumoniae* strains. *J. Clin. Microbiol.* **33**:1606–1612.
 24. Hermans, P. W. M., D. van Soolingen, E. M. Bik, P. E. W. de Haas, J. W. Dale, and J. D. A. van Embden. 1991. The insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect. Immun.* **59**:2695–2705.
 25. Hermans, P. W. M., D. van Soolingen, and J. D. A. van Embden. 1992. Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium goodii*. *J. Bacteriol.* **174**:4157–4165.
 26. Imeada, T. 1985. Deoxyribonucleic acid relatedness among selected strains of *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium bovis* BCG, *Mycobacterium microti*, and *Mycobacterium africanum*. *Int. J. Syst. Bacteriol.* **35**:147–150.
 27. Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, and J. van Embden. 1997. Rapid detection and simultaneous strain differentiation of *Mycobacterium tuberculosis* for diagnosis and tuberculosis control. *J. Clin. Microbiol.* **35**:907–914.
 28. Kapur, V., T. S. Whittam, and J. M. Musser. 1994. Is *Mycobacterium tuberculosis* 15,000 years old? *J. Infect. Dis.* **170**:1348–1349.
 29. Kent, L., T. D. McHugh, O. Billington, J. W. Dale, and S. H. Gillespie. 1995. Demonstration of homology between IS6110 of *Mycobacterium tuberculosis* and DNAs of other *Mycobacterium* spp. *J. Clin. Microbiol.* **33**:2290–2293.
 30. Kirschner, P., B. Springer, U. Vogel, A. Meier, A. Wrede, M. Kiekenbeck, F.-C. Bange, and E. C. Böttger. 1993. Genotypic identification of mycobacteria by nucleic acid sequence determination: report of a 2-year experience in a clinical laboratory. *J. Clin. Microbiol.* **31**:2882–2889.
 31. Kremer, K. Personal communication.
 32. Kurepina, N. E., S. Sreevatsan, B. B. Plikaytis, P. J. Bifani, N. D. Connell, R. J. Donnelly, D. van Soolingen, J. M. Musser, and B. N. Kreiswirth. 1998. Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS6110 elements in *Mycobacterium tuberculosis*: non-random integration in the dnaA-dnaN region. *Tuber. Lung Dis.* **79**:31–42.
 33. Li, H., J. C. Ulstrup, T. O. Jonassen, K. Melby, S. Nagai, and M. Harboe. 1993. Evidence for absence of the MPB64 gene in some substrains of *Mycobacterium bovis* BCG. *Infect. Immun.* **61**:1730–1734.
 34. Liébana, E., A. Aranaz, B. Francis, and D. Cousins. 1996. Assessment of genetic markers for differentiation within the *Mycobacterium tuberculosis* complex. *J. Clin. Microbiol.* **34**:933–938.
 35. Mariani, F., E. Piccolella, V. Colizzi, R. Rappuoli, and R. Gross. 1993. Characterization of an IS-like element from *Mycobacterium tuberculosis*. *J. Gen. Microbiol.* **139**:1767–1772.
 36. McHugh, T. D., L. E. Newport, and S. H. Gillespie. 1997. IS6110 homologs are present in multiple copies in mycobacteria other than tuberculosis-causing mycobacteria. *J. Clin. Microbiol.* **35**:1769–1771.
 37. Otal, I., S. Samper, M. P. Asensio, M. A. Victoria, M. C. Rubio, R. Gómez-Lus, and C. Martín. 1997. Use of a PCR method based on IS6110 polymorphism for typing *Mycobacterium tuberculosis* strains from BACTEC cultures. *J. Clin. Microbiol.* **35**:273–277.
 38. Palittapongarnpim, P., S. Chomyc, A. Fanning, and D. Kunimoto. 1993. DNA fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolates by arbitrarily primed polymerase chain reaction. *J. Infect. Dis.* **16**:975–978.
 39. Plikaytis, B. B. Personal communication.
 40. Plikaytis, B. B., J. T. Crawford, C. L. Woodley, W. R. Butler, K. D. Eisenach, M. D. Cave, and T. M. Shinnick. 1993. Rapid, amplification-based fingerprinting of *Mycobacterium tuberculosis*. *J. Gen. Microbiol.* **139**:1537–1542.
 41. Ross, B. C., K. Raios, K. Jackson, and B. Dwyer. 1992. Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. *J. Clin. Microbiol.* **30**:942–946.
 42. Sahadevan, R., S. Narayanan, C. N. Paramasivan, R. Prabhakar, and P. R. Narayanan. 1995. Restriction fragment length polymorphism typing of clinical isolates of *Mycobacterium tuberculosis* from patients with pulmonary tuberculosis in Madras, India, by use of direct-repeat probe. *J. Clin. Microbiol.* **33**:3037–3039.
 43. Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam. 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl. Environ. Microbiol.* **51**:873–884.
 44. Small, P. M., P. C. Hopewell, S. P. Singh, A. Paz, J. Parsonnet, D. C. Ruston, G. F. Schecter, C. L. Daley, and G. K. Schoolnik. 1994. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N. Engl. J. Med.* **330**:1703–1709.
 45. 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.
 46. Steenken, W., Jr., and L. U. Gardner. 1946. History of H37 strain of tubercle bacillus. *Am. Rev. Tuberc. Pulm. Dis.* **54**:62–66.
 47. Thierry, D., A. Brisson-Noël, V. Vincent-Lévy-Frébault, S. Nguyen, J. Guesdon, and B. Gicquel. 1990. Characterization of a *Mycobacterium tuberculosis* insertion sequence, IS6110, and its application in diagnosis. *J. Clin. Microbiol.* **28**:2668–2673.
 48. van Embden, J. D. A., J. T. Crawford, J. W. Dale, B. Gicquel, P. Hermans, R. McAdam, T. Shinnick, and P. M. Small. 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J. Clin. Microbiol.* **31**:406–409.
 49. van Embden, J. D. A., L. M. Schouls, and D. van Soolingen. 1995. Molecular techniques: applications in epidemiologic studies, p. 15–27. *In* C. O. Thoen and J. H. Steele (ed.), *Mycobacterium bovis* infection in animals and humans. Iowa State University Press, Ames, Iowa.
 50. van Soolingen, D., P. E. W. de Haas, R. M. Blumenthal, K. Kremer, M. Sluijter, J. E. M. Pijnenburg, L. M. Schouls, J. E. R. Thole, M. W. G. Dessens-Kroon, J. D. A. van Embden, and P. W. M. Hermans. 1996. Host-mediated modification of *PvuII* restriction in *Mycobacterium tuberculosis*. *J. Bacteriol.* **178**:78–84.
 51. van Soolingen, D., P. E. W. de Haas, J. Haagsma, T. Eger, P. W. M. Hermans, V. Ritacco, A. Alito, and J. D. A. van Embden. 1994. Use of various genetic markers in differentiation of *Mycobacterium bovis* strains from animals and humans and for studying epidemiology of bovine tuberculosis. *J. Clin. Microbiol.* **32**:2425–2433.
 52. van Soolingen, D., P. E. W. de Haas, P. W. M. Hermans, P. Groenen, and J. D. A. van Embden. 1993. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **31**:1987–1995.
 53. van Soolingen, D., P. E. W. de Haas, P. W. M. Hermans, and J. D. A. van Embden. 1994. DNA fingerprinting of *Mycobacterium tuberculosis*. *Methods Enzymol.* **235**:196–205.
 54. van Soolingen, D., P. W. M. Hermans, P. E. W. de Haas, and J. D. A. van Embden. 1992. Insertion element IS1081-associated restriction fragment length polymorphism in *Mycobacterium tuberculosis* complex species: a reliable tool to recognize *Mycobacterium bovis* BCG. *J. Clin. Microbiol.* **30**:1772–1777.
 55. van Soolingen, D., T. Hoogenboezem, P. E. W. de Haas, P. W. M. Hermans, M. A. Koedam, K. S. Teppema, P. J. Brennan, G. S. Besra, F. Portaels, J. Top, L. M. Schouls, and J. D. A. van Embden. 1997. A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, *canettii*: characterization of an exceptional isolate from Africa. *Int. J. Syst. Bacteriol.* **47**:1236–1245.
 56. van Soolingen, D., L. Qian, P. E. W. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhasaikhan, P. Nymadawa, and J. D. A. van Embden. 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of East Asia. *J. Clin. Microbiol.* **33**:3234–3238.
 57. van Soolingen, D., A. G. M. van der Zanden, P. E. W. de Haas, G. T. Noordhoek, A. Kiers, N. A. Foudraïne, F. Portaels, A. H. J. Kolk, K. Kremer, and J. D. A. van Embden. 1998. Disclosure of *Mycobacterium microti* infections among humans using novel genetic markers. *J. Clin. Microbiol.* **36**:1840–1845.
 58. van Steenberghe, T. J. M., S. D. Colloms, P. W. M. Hermans, J. de Graaff, and R. H. A. Plasterk. 1995. Genomic DNA fingerprinting by restriction fragment end labeling (RFEL). *Proc. Natl. Acad. Sci. USA* **92**:3234–3238.
 59. Vera-Cabrera, L., S. T. Howard, A. Laszlo, and W. M. Johnson. 1997. Analysis of genetic polymorphism in the phospholipase region of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **35**:1190–1195.
 60. Wiid, I. J. F., C. Werely, N. Beyers, P. Donald, and P. D. van Helden. 1994. Oligonucleotide (GTG)₅ as a marker for *Mycobacterium tuberculosis* strain identification. *J. Clin. Microbiol.* **32**:1318–1321.
 61. Yang, Z. H., P. E. W. de Haas, C. H. Wachman, D. van Soolingen, J. D. A. van Embden, and Å. B. Andersen. 1995. Molecular epidemiology of tuberculosis in Denmark in 1992. *J. Clin. Microbiol.* **33**:2077–2081.
 62. Zhang, Y., B. Heym, B. Allen, D. Young, and S. Cole. 1992. The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. *Nature* **358**:591–593.