

## Three-Year Population-Based Evaluation of Standardized Mycobacterial Interspersed Repetitive-Unit–Variable-Number Tandem-Repeat Typing of *Mycobacterium tuberculosis*<sup>∇</sup>

Caroline Allix-Béguec,<sup>1†</sup> Maryse Fauville-Dufaux,<sup>1‡</sup> and Philip Supply<sup>2,3‡\*</sup>

Scientific Institute of Public Health, Department Institut Pasteur de Bruxelles, Laboratoire Tuberculose et Mycobacteries, rue Engeland 642, 1180 Bruxelles, Belgium,<sup>1</sup> and Institut Pasteur de Lille<sup>2</sup> and Laboratoire des Mécanismes Moléculaires de la Pathogénèse Bactérienne,<sup>3</sup> INSERM U629-I, rue du Professeur Calmette, BP 245, 59019 Lille Cedex, France

Received 29 October 2007/Returned for modification 10 December 2007/Accepted 22 January 2008

**Standardized mycobacterial interspersed repetitive-unit–variable-number tandem repeat (MIRU-VNTR) typing based on 15 and 24 loci recently has been proposed for *Mycobacterium tuberculosis* genotyping. So far, this optimized system has been assessed in a single, 1-year population-based study performed in Germany (M. C. Oelemann, R. Diel, V. Vatin, W. Haas, S. Rusch-Gerdes, C. Locht, S. Niemann, and P. Supply, *J. Clin. Microbiol.* 45:691–697, 2007). Here, we evaluated these optimized formats in a much larger population-based study conducted during 39 months in the Brussels capital region of Belgium. Isolates from 807 patients were genotyped. The resolution power, cluster, and lineage identification by the standardized MIRU-VNTR sets were compared to those obtained using standardized IS6110-restriction fragment length polymorphism (RFLP), spoligotyping, and a previous 12-MIRU-VNTR-locus set. On a subset representing 77% of the cases during a 16-month period, a high concordance was observed between unique isolates or strain clusters as defined by standardized MIRU-VNTR and IS6110-RFLP (i.e., more than five IS6110 bands). When extended to the entire population-based collection, the discriminatory subset of 15 loci decreased the strain-clustering rate by almost twofold compared to that of the old 12-locus set. The addition of the nine ancillary MIRU-VNTR loci and/or spoligotyping only slightly further decreased this strain-clustering rate. Familial, social, and/or geographic proximity links were found in 48% of the clusters identified, and well-known risk factors for tuberculosis transmission were identified. Finally, an excellent correspondence was determined between our MIRU-VNTR-spoligotyping strain identifications and external reference strain lineages included in the MIRU-VNTR<sub>plus</sub> database and identified by, e.g., large sequence polymorphisms. Our results reinforce the proposal of standardized MIRU-VNTR typing as a new reference genotyping method for the epidemiological and phylogenetic screening of *M. tuberculosis* strains.**

The molecular typing of *Mycobacterium tuberculosis* has greatly improved the knowledge and control of tuberculosis (TB) by allowing the detection of unsuspected transmission, the identification of false-positive cultures, and the distinction between reinfection and relapse (4). Furthermore, genetic markers permit the identification of different genetic lineages and the study of their geographical distribution and physiopathology, which has implications for the development of new tools for TB control (5, 7, 17).

So far, IS6110-restriction fragment length polymorphism (RFLP) has been the gold standard method for genotyping *M. tuberculosis* (43). IS6110 fingerprinting has proven useful for conducting population-based studies of TB transmission (2, 32). However, this technique is laborious and time-consuming with this slow-growing mycobacterium. In addition, it has insufficient capacity to discriminate *M. tuberculosis* strains with

low copy numbers of IS6110 and is not straightforward enough for the identification of genetic lineages (10, 22, 23).

Several alternative PCR-based methods have been developed in order to overcome these problems. Spoligotyping often is used as a secondary method for typing low-copy-number IS6110 isolates, but it does little to improve strain differentiation in population-based studies (11). The most promising PCR-based methods are based on the analysis of multiple loci containing variable numbers of tandem repeats (VNTR) of different families of interspersed genetic elements, collectively called mycobacterial interspersed repetitive units (MIRU) (15, 24, 28, 31, 34, 35, 39, 40). Currently, the most used version of this method (designated MIRU-VNTR) is based on the analysis of 12 loci (25, 38). Especially when applied as a first-line typing test in combination with spoligotyping, this 12-locus-based method can provide adequate discrimination in many cases, including in large-scale studies (6, 9). Nevertheless, this 12-locus set still fails to discriminate a significant number of different strains, as revealed by more extensive MIRU-VNTR sets or IS6110-RFLP combined with contact tracing data (9, 27, 37).

A standardized MIRU-VNTR format with significantly improved discriminatory power has been proposed on the basis of the analysis of the clonal stability and evolutionary rates of MIRU-VNTR markers in primary genetic lineages of tubercle

\* Corresponding author. Mailing address: Molecular Mechanisms of Bacterial Pathogenesis, INSERM U629, Institut de Biologie de Lille/ Institut Pasteur de Lille, 1 rue du Professeur Calmette, F-59021 Lille Cedex, France. Phone: (33)320871154. Fax: (33)320871158. E-mail: philip.supply@pasteur-lille.fr.

† Present address: Genoscreen, Campus de l'Institut Pasteur de Lille, 1 rue du Professeur Calmette, 59000 Lille, France.

‡ These authors contributed equally to this work.

∇ Published ahead of print on 30 January 2008.

bacilli from around the world (37). This format is comprised of 24 loci, 15 of which were defined as composing a discriminatory subset based on higher variability within the different clonal complexes studied. So far, the predictive value of these standardized sets for estimating *M. tuberculosis* transmission has been evaluated only retrospectively, in a single population-based study conducted during 1 year in Hamburg, Germany (27).

Here, we evaluated this standardized format prospectively in a fivefold-larger population-based study with a different patient population, conducted during 39 months in the Brussels capital region of Belgium. The discrimination and cluster identification by these new MIRU-VNTR sets were compared to those obtained using the previous set of 12 loci, spoligotyping, and IS6110-RFLP and to available epidemiological information. Genetic lineages of *M. tuberculosis* were determined based on the congruence of the different markers and on the newly available MIRU-VNTRplus identification database (45). This database permits clonal identification based on comparisons to the reference lineages as defined by a comprehensive marker set, which is made up of large sequence polymorphisms (LSPs) and single-nucleotide polymorphisms (SNPs) in addition to MIRU-VNTR and spoligotyping. These data then were analyzed in terms of the geographic origin of the patients.

#### MATERIALS AND METHODS

**Study isolates.** From 1 September 2002 to 31 December 2005, 807 *M. tuberculosis* cultures were received from the five accredited hospitals of the Brussels capital region and were genotyped at the Tuberculosis and Mycobacteria Reference Center, Pasteur Institute, Brussels, Belgium; one sample from each patient was used. During the same period, 32 cultures, obtained from (an)other isolate(s) from some of the same patients, additionally were genotyped. Speciation was performed using either the AccuProbe *M. tuberculosis* test (GeneProbe, San Diego, California), the Amplicor MTB assay (Roche Diagnostics, Vilvoorde, Belgium), PCR amplification of the IS6110 element, or classical biochemical tests.

**Molecular typing method.** IS6110-RFLP analyses were performed on 258 isolates according to an internationally agreed-upon standard method (43). Spoligotyping was used as previously described by Kamerbeek et al. (22). Twenty-four-locus-based MIRU-VNTR typing was routinely applied using a four-capillary-based ABI 3100-Avant genetic analyzer as described by Allix et al. (3) and Supply et al. (37).

**Computer-assisted analysis of patterns.** IS6110-RFLP patterns, spoligotyping, and MIRU-VNTR profiles were analyzed using the Bionumerics package, version 4.5 (Applied Maths, St-Martin-Latem, Belgium). Dendrograms based on IS6110 fingerprints were generated using the dice coefficient, the unweighted-pair group method using average linkages, and a position tolerance of 1.8%. Dendrograms based on MIRU-VNTR patterns were generated using the categorical coefficient and neighbor-joining method and were rooted using a "*Mycobacterium prototuberculosis*" C/D genotype (also called "*Mycobacterium canettii*") (20). A strain cluster was defined as two or more patients infected by isolates having identical genotypes depending on the typing method(s) used. Assuming that one patient from each strain cluster corresponded to the index case at the origin of infection, the strain-clustering rate (or recent transmission index) was calculated with the following equation: strain-clustering rate<sub>n-1</sub> =  $(n_c - c)/n$ , where  $n_c$  is the total number of strain-clustered cases,  $c$  is the number of strain clusters, and  $n$  is the total number of cases in the sample (33). For cluster analysis, isolates with mixed populations, identified by double alleles, in two or more MIRU-VNTR loci were excluded, and for isolates with clonal variants, the single locus displaying a double allele was not considered.

**Statistical analysis.** Pearson  $\chi^2$  or the Fisher exact test (depending on the number of subjects) was used to test pairwise differences in strain genetic lineages between patients of various geographic origins.

#### RESULTS

**Study population.** From 1 September 2002 to 31 December 2005, 1,157 new TB cases were identified in the Brussels capital region. Positive cultures were obtained from 947 of the patients, of which 814 cases (86% of culture-positive patients) were genotyped. Of these, seven cases were excluded because of probable laboratory cross-contamination, as confirmed by genotyping. Contamination was assumed to have occurred when two isolates with identical fingerprints were obtained from the same laboratory less than 1 day apart (except in one case) and were coupled with the absence of clinical TB symptoms.

The remaining patients (343/1,150) were not typed due to their producing culture-negative samples ( $n = 210$ ), contamination by other microorganisms ( $n = 3$ ), and culture unavailability ( $n = 130$ ). As expected, young patients (less than 15 years of age;  $n = 48$ ) and patients with extrapulmonary TB ( $n = 129$ ), notoriously known to give paucibacillary samples, were overrepresented among the culture-negative patients (data not shown).

**Strain typeability, clonal variants, and mixed populations.** The 807 isolates were fully typeable for the 24 MIRU-VNTR loci. Ninety-seven percent of the alleles were obtained after the first round of multiplex PCR. The remaining 3% were obtained after an additional round(s) of injection on the DNA analyzer or of multiplex or simplex PCR amplification. For one isolate, the spoligotype could not be obtained despite repeated attempts. IS6110-RFLP was not systematically applied to the full collection (see below).

Of the 807 isolates, only 8 reproducibly displayed a double allele in a single MIRU-VNTR locus, thus identifying the simultaneous presence of two closely related clonal variants, as defined previously (3, 18, 30, 37). The corresponding loci were 2163b ( $n = 3$ ), 4052 ( $n = 2$ ), 2165 ( $n = 1$ ), 2461 ( $n = 1$ ), and MIRU 27 ( $n = 1$ ). For each of these isolates, the corresponding locus initially was treated as missing data for cluster identification (i.e., only the 23 other loci of these isolates were considered). The double allele subsequently was considered for the detection of possible overlaps of clusters identified (see Discussion). Only five isolates reproducibly displayed double alleles in two ( $n = 1$ ) or more ( $n = 4$ ) loci, thereby identifying the simultaneous presence of independent clones in accordance with previous studies (3, 18, 30, 37). These isolates were excluded from subsequent analyses.

**Discriminatory power and cluster identification in a test panel.** As an initial step, the resolution power of MIRU-VNTR typing alone or in combination with spoligotyping was compared to that of IS6110-RFLP by analyzing 258 isolates (Table 1). This panel represented 77% (258/334) of all genotyped isolates from 1 September 2002 to 31 December 2003.

Among these isolates, 89% ( $n = 230$ ) showed IS6110-RFLP profiles classified as high copy numbers (six or more IS6110 bands), whereas the remaining 11% ( $n = 28$ ) had fingerprints with one to five IS6110 bands. In all, IS6110-RFLP defined 167 unique and 91 clustered isolates, giving a strain-clustering rate of 24.4%. By applying MIRU-VNTR typing, a better resolution was obtained no matter which standardized format (i.e., 15 or 24 locus based) was used. Actually, both the 24- and 15-locus-based sets identified the same 180 unique and 78 clustered isolates, thus yielding an identical strain-clustering rate

TABLE 1. Discriminatory power and cluster identification in a test panel (77% of all genotyped isolates from Brussels during a 16-month period;  $n = 258$ )<sup>a</sup>

Technique and no. of loci	No. of different profiles	No. of isolates with unique profiles	No. of clusters	No. of isolates in cluster	Strain-clustering rate (%)
Spoligotyping	113	76	37	182	56.2
IS6110-RFLP	195	167	28	91	24.4
Low copy no. ( $\leq 5$ IS6110 bands [ $n = 28$ ])	12	7	5	21	57.1
High copy no. ( $< 5$ IS6110 bands [ $n = 230$ ])	183	160	23	70	20.4
IS6110-RFLP and spoligotyping	206	180	26	78	20.2
15 MIRU-VNTR loci	207	180	27	78	19.8
Spoligotyping and 15 MIRU-VNTR loci	207	180	27	78	19.8
IS6110-RFLP and 15 MIRU-VNTR loci	213	189	24	69	17.4
IS6110-RFLP, spoligotyping, and 15 MIRU-VNTR loci	213	189	24	69	17.4

<sup>a</sup> The techniques referred to as 15 and 24 MIRU-VNTR loci correspond to the optimized formats defined in reference 37.

of 19.8%. The addition of spoligotyping had no effect on the resolution power obtained by using MIRU-VNTR typing alone in this panel, as none of the MIRU-VNTR-defined clusters was split by spoligotyping.

Not surprisingly, the five IS6110-RFLP clusters displaying fingerprints with one to five bands all were subdivided by MIRU-VNTR typing. Most (19/21) of the isolates grouped within these clusters were identified as unique by MIRU-VNTR.

Of the 23 IS6110-RFLP clusters with high copy numbers and grouping a total of 70 isolates, 20 clusters (including 61 isolates) were found to be completely identical by MIRU-VNTR typing. Of the three remaining IS6110-RFLP clusters, two (including two Haarlem and three LAM isolates with 9 and 11 IS6110 bands, respectively) were fully subdivided both by four to seven MIRU-VNTR loci and by spoligotyping, whereas one cluster (including four isolates) was distinguished into two pairs by a single-locus change by MIRU-VNTR typing.

Conversely, only five clusters defined by MIRU-VNTR typing based on the 15 and the 24 loci and grouping a total of 12 isolates were subdivided by IS6110-RFLP. Four of these MIRU-VNTR clusters included two isolates, whereas one, containing four isolates, was distinguished into two pairs by IS6110-RFLP. In all cases, the IS6110-RFLP differences consisted of a single band difference.

**Discriminatory power of the PCR-based methods in the entire population-based collection.** Given the higher-resolution power of MIRU-VNTR typing compared to that of IS6110-RFLP and the excellent correlation for the strain cluster definition (excluding nonrelevant IS6110-RFLP clusters

with low numbers of bands) in the initial test panel, standardized MIRU-VNTR typing was applied in combination with spoligotyping for the screening of the *M. tuberculosis* isolates subsequently collected from 2004 and 2005. The analysis of the discriminatory power and cluster definition by the PCR-based methods thus was extended to a total of 802 isolates. The isolates were obtained during the full study period; five isolates that corresponded to mixtures of two strains were excluded (see above).

The discriminatory power of MIRU-VNTR typing was evaluated by comparing results obtained using the discriminatory subset of 15 loci, the full set of 24 loci, and the old set of 12 MIRU-VNTR loci (Table 2). The discriminatory subset of 15 loci distinguished 596 different profiles, resulting in a strain-clustering rate of 25.8%. The number of additional profiles obtained with the full set of 24 loci was limited to 14, resulting in only a slight decrease in the strain-clustering rate (23.9%). In comparison, the number of profiles and the strain-clustering rate obtained with the old set of 12 loci were 418 and 47.9%, respectively.

In the full population-based sample, spoligotyping alone distinguished 282 profiles, yielding a strain-clustering rate of 65%. When used as a secondary method, spoligotyping nevertheless subdivided some clusters defined by MIRU-VNTR typing and added 28 and 19 supplementary profiles in combination with the 15- and 24-locus sets, respectively. The strain-clustering rates were slightly reduced accordingly, to 22.2 and 21.6%, respectively. The combined use of the old 12-MIRU-VNTR-locus set and spoligotyping consistently generated significantly fewer profiles than the 15-locus set, even when the latter

TABLE 2. Discriminatory power of the PCR-based methods in the entire population-based collection ( $n = 802$ )<sup>a</sup>

Technique and no. of loci	No. of different profiles	No. of isolates with unique patterns	No. of clusters	No. of isolates in cluster	Strain-clustering rate (%)
Spoligotyping	282	181	101	621	64.8
12 MIRU-VNTR loci	418	304	114	498	47.9
Spoligotyping and 12 MIRU-VNTR loci	544	440	104	362	32.2
15 MIRU-VNTR loci	596	499	97	304	25.8
24 MIRU-VNTR loci	610	514	96	288	23.9
Spoligotyping and 15 MIRU-VNTR loci	624	539	85	263	22.2
Spoligotyping and 24 MIRU-VNTR loci	629	547	82	255	21.6

<sup>a</sup> The technique referred to as 12 MIRU-VNTR loci corresponds to the original set defined in reference 40, while those referred to as 15 and 24 MIRU-VNTR loci correspond to the optimized formats defined in reference 37.

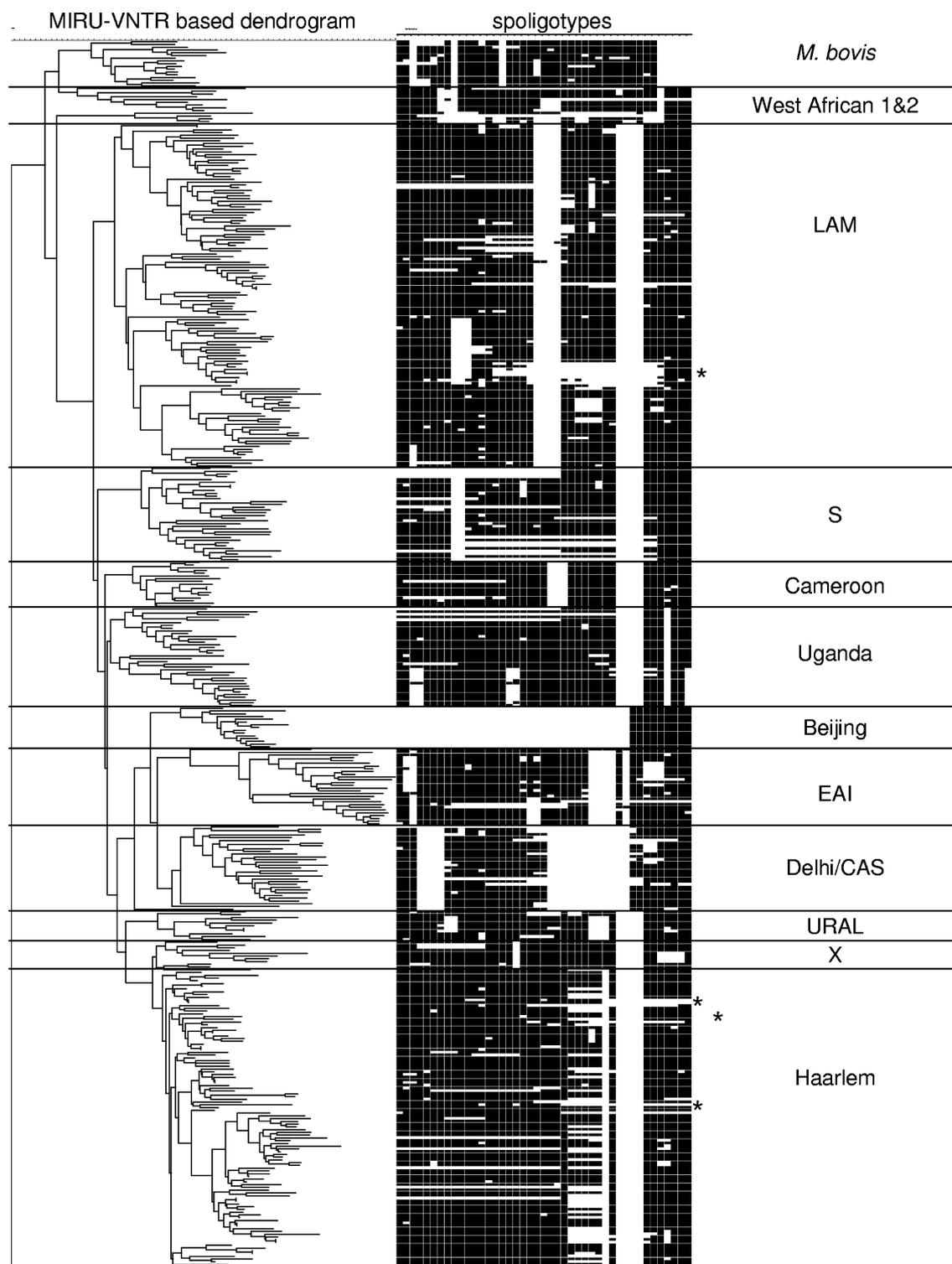


FIG. 1. Concordance observed between MIRU-VNTR groupings and spoligotype assignments among the 446 isolates from well-defined lineages. MIRU-VNTR groupings are visualized by a dendrogram calculated using a neighbor-joining algorithm and rooted using an “*M. prototuberculosis*” C/D genotype (20). Asterisks indicate examples in which the MIRU-VNTR groupings confirmed the genetic classifications that initially were presumed on the basis of spoligotypes sharing only partial spoligoprototype signatures.

MIRU-VNTR-locus set was considered alone (544 versus 596 profiles).

**Phylogeographic analysis.** The Brussels capital region is a cosmopolitan area, and 76% ( $n = 612$ ) of TB patients in this

study were born in 69 different countries, with the majority being from Africa ( $n = 381$ ). Thus, this population offers an opportunity for the study of global *M. tuberculosis* phylogeography, especially for Africa, which is understudied in this re-

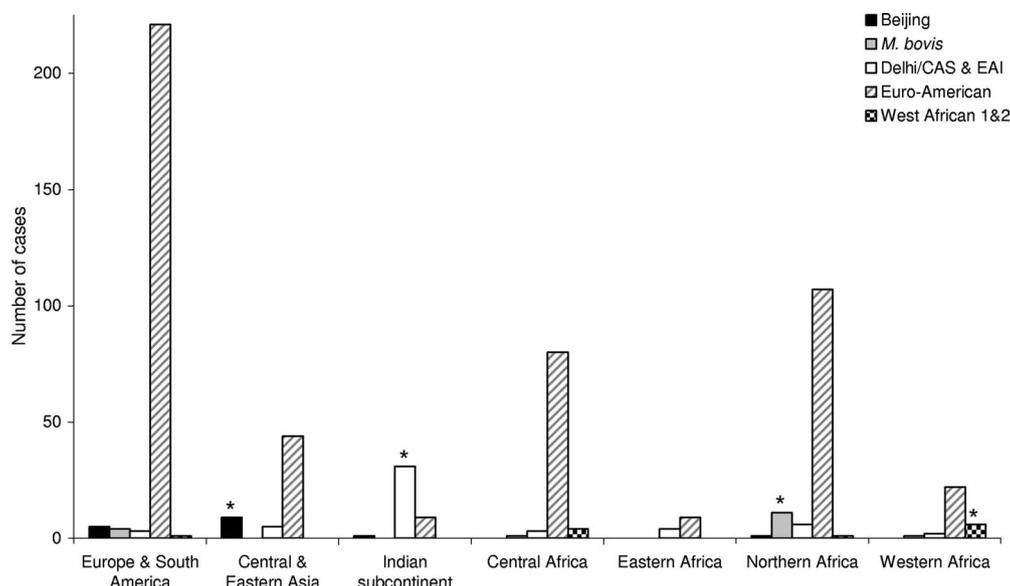


FIG. 2. Distribution of major *M. tuberculosis* lineages according to patients' regions of origin. Asterisks indicate statistically significant associations ( $P < 0.001$ ).

gard. The strain lineage distribution was determined for the 629 different genotypes, firstly by analyzing the congruence between spoligotyping and 24-MIRU-VNTR-locus typing within the study collection. Among them, 12 major spoligotype families, with profiles matching the signatures of classical prototypes (7, 13) and each containing 9 to 124 isolates, were identified (Fig. 1). Using a tree based on the 24-MIRU-VNTR-locus data (Fig. 1), excellent concordance was observed between MIRU-VNTR groupings and these spoligotype assignments, except for isolates with undefined or T spoligotypes, which is in keeping with results indicating that the latter actually conceal genetically heterogeneous strains (14).

In addition, the strain lineage designations based on this concordance were tested by searching for the best matches of their respective 24-MIRU-VNTR-locus genotypes, alone or in combination with their spoligotypes, among external reference lineages included in the MIRU-VNTR<sub>plus</sub> database. These results are detailed in a study presenting this database (C. Allix-Béguec, T. Weniger, D. Harmsen, P. Supply, and S. Niemann, unpublished data). In summary, internal designations were confirmed by consistent best matches in 88.7 to 94.8% of the cases when 24-MIRU-VNTR-locus types were considered alone or in combination with spoligotyping. Accordingly, absences of matches or apparent conflicts were detected in only 8.1 to 1.8% of the cases when 24-MIRU-VNTR-locus types were considered alone and in 1.6 to 3.4% of the cases when 24-MIRU-VNTR-locus types were considered in combination with spoligotyping.

The association between a patient's strain lineage and that patient's region of birth has been shown on the basis of LSP use (16). This analysis was performed after regrouping, based on a shared sequence deletion in *pks15/1*, all of the branches corresponding to SNP-based principal genetic group 2 (PGG2) and PGG3 (36) into a superlineage called Euro-American. This superlineage comprises lineages of Cameroon, Haarlem, LAM, Ural, S, and Uganda spoligotypes plus T spoligotypes

(17 and T. Wirth, F. Hildebrand, C. Allix, F. Wölbelling, T. Kubica, K. Kremer, D. van Soolingen, S. Rüsche-Gerdes, C. Loch, A. Meyer, P. Supply, and S. Niemann, unpublished data). We first similarly grouped our lineages, as identified by 24-MIRU-VNTR-locus types and confirmed by spoligotyping and MIRU-VNTR<sub>plus</sub> analysis, and analyzed their correlations with a patient's region of birth (Fig. 2). Significant associations ( $P < 0.001$  in all cases) were consistently found between CAS and EAI lineages and patients from the Indian subcontinent, between the Beijing lineage and those from central and east Asia, between *Mycobacterium africanum* (West African 1 and 2) and *Mycobacterium bovis* lineages and those from western and northern Africa, and between the Euro-American superlineage and those from Europe, Africa, and North and South America.

Taking advantage of the resolution provided by our markers, we then sharpened the phylogeographical analysis by considering branches within the Euro-American superlineage individually and then refining geographic localizations. The LAM (19.7%) and Haarlem (17.0%) branches were the two most prevalent lineages in our study collection. Of these, a significant connection with a specific patient origin within the Euro-(African)-American region was found only for LAM isolates, which are associated with South American, North African, and central African patients (Fig. 3). Other PGG2 and PGG3 lineages more infrequently found in our setting were strongly linked with the patient's origin, as was the case for Cameroon isolates associated with western and northern African patients and Uganda isolates associated with patients from central Africa.

## DISCUSSION

This report represents the first large-scale population-based evaluation of the recently standardized MIRU-VNTR 15- and 24-locus-based formats. This evaluation entailed (i) a compar-

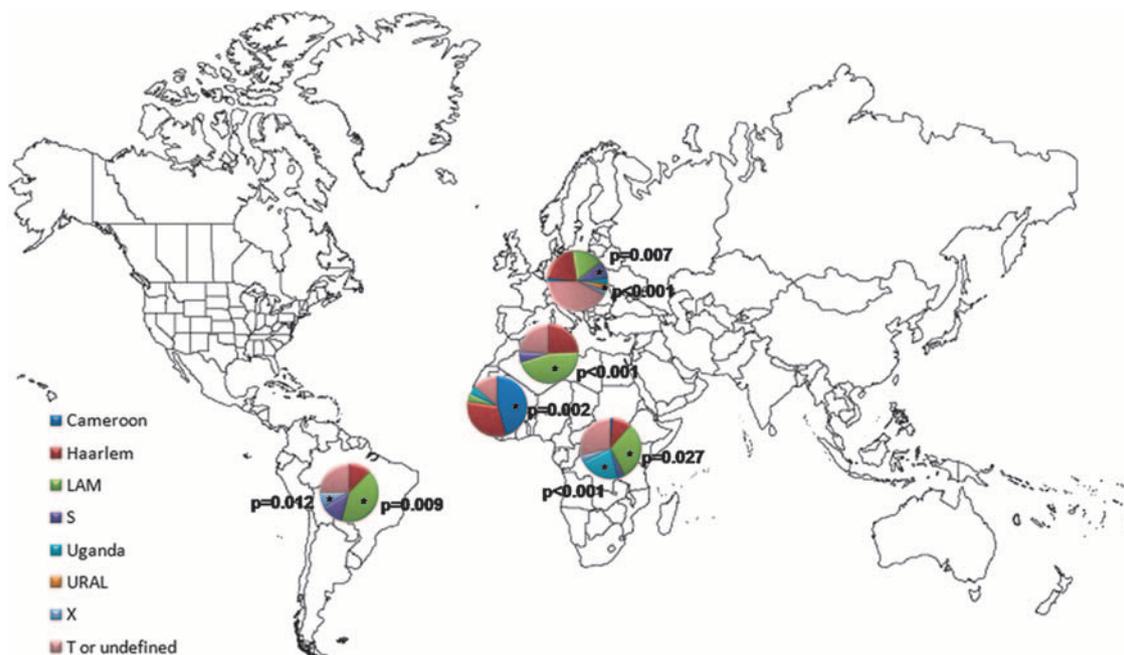


FIG. 3. Distribution of branches of the Euro-American *M. tuberculosis* superlineage according to patients' regions of origin. Asterisks indicate statistically significant associations, and respective *P* values are indicated.

ison of the results of the cluster analysis based on these formats to those obtained using standard IS6110-RFLP or MIRU-VNTR typing based on 12 loci plus spoligotyping, (ii) the assessment of standardized MIRU-VNTR typing for phylogeographical analysis, and (iii) a test of the robustness and operability in routine conditions of a reference clinical laboratory.

A high correlation was found between unique isolates or strain clusters defined by MIRU-VNTR and IS6110-RFLP (i.e., more than five IS6110 bands) in our test panel, representing 77% of all genotyped isolates during the 16-month period. Of the 23 IS6110-RFLP clusters with high copy numbers, 20 were found to be completely identical to each other by MIRU-VNTR typing. Of the three remaining IS6110-RFLP clusters, two were fully subdivided both by four to seven MIRU-VNTR loci and by spoligotyping. As shown by previous studies (27, 42), such discrimination by multiple MIRU-VNTR loci, a fortiori when independently corroborated by spoligotype differences, is predictive of the absence of a link between the patients involved and, thus, indicates that these IS6110 clusters are epidemiologically irrelevant. In contrast, only one IS6110-RFLP spoligotype cluster, containing four patient isolates, was subdivided into two pairs by a single-locus variation (SLV) by MIRU-VNTR typing. Interestingly, one of these pairs involved two Rwandese patients, while the other one grouped together a Belgian and a Somali patient who were identified as acquaintances by subsequent contact tracing. These two patient groups might be unrelated, since MIRU-VNTR SLVs have been observed in several instances among isolates with fully matching IS6110-RFLP results and spoligotypes originating from epidemiologically unlinked patients (9). However, it cannot be completely excluded that the SLV between our two patient isolate groups reflects a rare MIRU-VNTR mutation and genetic drift

in the clonal population originating from recent transmission (29, 37).

Not surprisingly, the five IS6110-RFLP clusters displaying fingerprints with one to five bands all were subdivided by MIRU-VNTR typing. Conversely, only five clusters defined by 24-MIRU-VNTR-locus typing were subdivided by IS6110-RFLP. In all cases, the IS6110-RFLP differences consisted of a single-band change among multibanded profiles, the epidemiological interpretation of which is often questioned (46). Regardless, the discriminatory power and the accuracy for the cluster analysis of MIRU-VNTR typing appeared to be slightly greater than that of IS6110-RFLP in this test panel. In these conditions, the maximal resolution power was achieved with the discriminatory subset of 15 loci without the influence of the additional inclusion of the 9 ancillary loci and/or spoligotyping.

The benefit of the 15- and 24-locus formats was further evaluated in the entire population-based collection by comparing them to the 12-locus format currently used alone or in combination with spoligotyping for universal genotyping in the United States and elsewhere (6, 9). The use of the 15- and 24-locus formats resulted in a 50% drop in the strain-clustering rate compared to that obtained with the old set of 12 loci (25.8, 23.9, and 47.9%, respectively). The small 2% difference in strain-clustering rates between the 15- and 24-locus formats is consistent with the design of the discriminatory subset of 15 loci, which comprises the most variable markers across representative *M. tuberculosis* genetic lineages (37). It is noteworthy that the secondary use of spoligotyping slightly improved the resolution obtained with both the 15- and 24-locus formats and, interestingly, rendered the difference between these two formats negligible (22.2 and 21.6%, respectively). In contrast, the combined use of the old 12-locus set and spoligotyping

generated a significantly higher strain-clustering rate (32.2%) than did the 15- and 24-locus sets, even without considering spoligotyping results.

The epidemiological interpretation of the molecular results depends not only upon the discriminatory power of the markers but also upon their adequate clonal stability during infection and transmission. A total of 59 serial isolates, obtained from 27 different patients during the study period, were genotyped by using the 24 MIRU-VNTR loci and spoligotyping. Consistently with the low frequency of exogenous reinfection expected in this type of setting (8) and other analyses on MIRU-VNTR stability (29, 37), the genotypes were conserved across the full set of markers within the 27 groups.

The epidemiological significance of the clusters defined by 24 MIRU-VNTR loci and spoligotyping combined was further evaluated in a parallel study by analyzing epidemiological, demographic, and contact tracing data available for TB cases registered from 1 January 2003 to 31 December 2004 (C. Allix-Béguec, P. Supply, M. Wanlin, P. Bifani, and M. Fauville-Dufaux, unpublished data). These analyses confirmed familial or social links for at least 51/157 (32%) strain-clustered patients, including those defining a large ongoing outbreak. Patients were found to live in close proximity in 25/157 (16%) additional strain-clustered cases of unknown direct epidemiological links. The classical risk factors for TB transmission (e.g., being young and underprivileged) likewise were identified based on the clusters defined by MIRU-VNTR and spoligotyping. Furthermore, the strain-clustering rate (20% during the period considered) defined on this basis was remarkably identical to that established based on IS6110-RFLP in a similar study that involved a comparable population during the same 2-year period (41). Note that similar results would have been obtained with the 15 MIRU-VNTR loci and spoligotyping, as the strain clusters obtained in this case were virtually the same as those obtained with 24 MIRU-VNTR loci and spoligotyping (see above).

Interestingly, one isolate with a double allele in a single locus constituted the overlap of two clusters by sharing allele 3 with one isolate and allele 4 with another one in locus 2165, while the 23 other loci were fully identical for the three isolates. However, the plausibility of the possible corresponding epidemiological connection could not be confirmed by epidemiological or demographic data.

We additionally investigated the capacity of standardized MIRU-VNTR typing to identify different well-defined *M. tuberculosis* strain lineages. Because *M. tuberculosis* has a clonal population associated with patient geographic origins (16), strain lineage information is useful, as it can, for instance, provide indications regarding the source of the TB case (ongoing transmission versus the reactivation or the importation of an infection acquired abroad). Phylogeographical studies also have implications for the development of new tools for TB control (17). The phylogenetic information conveyed by MIRU-VNTR markers has been questioned (14, 17). Here, we found an excellent congruence between groupings based on 24 MIRU-VNTR loci and well-defined spoligotype signatures within our population-based collection. In fact, in many cases the MIRU-VNTR groupings even confirmed the genetic classifications that initially were only presumed on the basis of the spoligotypes sharing only partial spoligoprototype signatures

(see, for instance, the Haarlem and LAM profiles in Fig. 1). Moreover, the strain lineage designations defined on this basis were tested by searching for best matches of their respective 24-MIRU-VNTR-locus genotypes among external reference strain lineages included in the database at [www.miru-vntrplus.org](http://www.miru-vntrplus.org). In this database, these references were identified by combining extensive marker sets, including SNPs and LSPs, which are presented as the most robust markers for phylogenetic analysis (17). As detailed in another study (Allix et al., unpublished), the internal designations were confirmed by consistent best matches in 88.7% of the cases when 24-MIRU-VNTR-locus types were used alone for analysis. While the absence of a match was found in only 8.1% of the cases, conflicts were detected in no more than 3.2% of the cases. Finally, when grouping our strain lineages, which were identified by 24-MIRU-VNTR-locus types and confirmed by spoligotyping and MIRU-VNTR<sub>plus</sub>, similarly to previously published data (16) and taking into account the correspondence of lineage nomenclatures, we found an association between patients' strain lineages and the patients' regions of birth, as was seen with the use of LSPs or SNPs (16, 19).

In conclusion, our results extend inferences, which initially were based on a smaller study on a different patient population in Germany (42% foreign-born patients, mostly from Turkey and eastern countries) (1, 12, 27), regarding the wide applicability of standardized MIRU-VNTR typing optionally combined with spoligotyping for the analysis of TB transmission, at least in the numerous settings with epidemiological features similar to those of these two studies. It remains to be studied how useful the standardized MIRU-VNTR set of 15 or 24 loci would be in a stable rural area with clusters existing for a long time period and displaying much less strain heterogeneity, but it can be predicted that, even in such conditions, these sets most likely will outperform the old 12-locus format for strain discrimination. As W/Beijing strains were infrequently found in these two studies, the possible relevance of a few additional MIRU-VNTR loci to this standardized format in specific populations in which such strains are predominant awaits a population-based evaluation, including detailed epidemiological data and refined genetic analyses (21, 26, 44). However, we already notice that some additionally proposed loci clearly do not meet the conditions of robustness and stability for standardized screening in and among laboratories (37). In contrast, with its 807 fully typeable isolates and the consistent epidemiological analysis results, this study demonstrates the excellent operability and efficiency of the standardized MIRU-VNTR formats for molecular screening in routine laboratory conditions. Lastly, we show that this standardized method permits the accurate and high-resolution analysis of *M. tuberculosis* phylogeography, which is now further facilitated by the online accessibility of the multifunctional MIRU-VNTR<sub>plus</sub> database. Taking into account its speed and portability, standardized MIRU-VNTR typing represents a powerful tool with diverse applications.

#### ACKNOWLEDGMENTS

We thank Polydor Sonck for his technical assistance in IS6110-RFLP. We thank A. Dediste and A. Lafontaine from Saint-Pierre Hospital, O. Denis and M. Struelens from Erasme Hospital, M. Delmee and C. Christael-Degraux from Saint-Luc Hospital, B. Mu-

longo and M. Weckx from Saint-Jean Hospital, and D. Pierard and M. Seghers from AZ-VUB hospital for their kind and helpful collaboration by providing positive cultures and collecting some patient data. Maryse Wanlin and Patrick de Smet from the Belgian Lung and Tuberculosis Association also are acknowledged for their help in collecting and analyzing patient data. This study would not have been possible without their collaboration.

C.A.-B. was a fellow of the Brussels capital region.

P.S. is a researcher of the Centre National de la Recherche Scientifique (CNRS).

#### REFERENCES

- Achtman, M., G. Morelli, P. Zhu, T. Wirth, I. Diehl, B. Kusecek, A. J. Vogler, D. M. Wagner, C. J. Allender, W. R. Easterday, V. Chenal-Francois, P. Worsham, N. R. Thomson, J. Parkhill, L. E. Lindler, E. Carniel, and P. Keim. 2004. Microevolution and history of the plague bacillus *Yersinia pestis*. *Proc. Natl. Acad. Sci. USA* **101**:17837–17842.
- Alland, D., G. E. Kalkut, A. R. Moss, R. A. McAdam, J. A. Hahn, W. Bosworth, E. Drucker, and B. R. Bloom. 1994. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. *N. Engl. J. Med.* **330**:1710–1716.
- Allix, C., P. Supply, and M. Fauville-Dufaux. 2004. Utility of fast mycobacterial interspersed repetitive unit-variable number tandem repeat genotyping in clinical mycobacteriological analysis. *Clin. Infect. Dis.* **39**:783–789.
- Barnes, P. F., and M. D. Cave. 2003. Molecular epidemiology of tuberculosis. *N. Engl. J. Med.* **349**:1149–1156.
- Bifani, P. J., B. Mathema, N. E. Kurepina, and B. N. Kreiswirth. 2002. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. *Trends Microbiol.* **10**:45–52.
- Blackwood, K. S., J. N. Wolfe, and A. M. Kabani. 2004. Application of mycobacterial interspersed repetitive unit typing to Manitoba tuberculosis cases: can restriction fragment length polymorphism be forgotten? *J. Clin. Microbiol.* **42**:5001–5006.
- Brudey, K., J. R. Driscoll, L. Rigouts, W. M. Proding, A. Gori, S. A. Al-Hajjaj, C. Allix, L. Aristimuno, J. Arora, V. Baumanis, L. Binder, P. Cafrune, A. Cataldi, S. Cheong, R. Diel, C. Ellermeier, J. T. Evans, M. Fauville-Dufaux, S. Ferdinand, D. Garcia de Viedma, C. Garzelli, L. Gazzola, H. M. Gomes, M. C. Gutierrez, P. M. Hawkey, P. D. van Helden, G. V. Kadival, B. N. Kreiswirth, K. Kremer, M. Kubin, S. P. Kulkarni, B. Liens, T. Lillebaek, H. M. Ly, C. Martin, I. Mokrousov, O. Narvskaja, Y. F. Ngeow, L. Naumann, S. Niemann, I. Parwati, M. Z. Rahim, V. Rasolof-Razanamparany, T. Rasolonavalona, M. L. Rossetti, S. Rusch-Gerdes, A. Sajduda, S. Samper, I. Shemyakin, U. B. Singh, A. Somoskovi, R. Skuce, D. van Soolingen, E. M. Streicher, P. N. Suffys, E. Tortoli, T. Tracevska, V. Vincent, T. C. Victor, R. Warren, S. F. Yap, K. Zaman, F. Portaels, N. Rastogi, and C. Sola. 2006. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol.* **6**:23.
- Cacho, J., A. Perez Meixeira, I. Cano, T. Soria, A. Ramos Martos, M. Sanchez Concheiro, S. Samper, P. Gavin, and C. Martin. 2007. Recurrent tuberculosis from 1992 to 2004 in a metropolitan area. *Eur. Respir. J.* **30**:333–337.
- Cowan, L. S., L. Diem, T. Monson, P. Wand, D. Temporado, T. V. Oemig, and J. T. Crawford. 2005. Evaluation of a two-step approach for large-scale, prospective genotyping of *Mycobacterium tuberculosis* isolates in the United States. *J. Clin. Microbiol.* **43**:688–695.
- Cowan, L. S., L. Mosher, L. Diem, J. P. Massey, and J. T. Crawford. 2002. Variable-number tandem repeat typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS6110 by using mycobacterial interspersed repetitive units. *J. Clin. Microbiol.* **40**:1592–1602.
- Cronin, W. A., J. E. Golub, L. S. Magder, N. G. Baruch, M. J. Lathan, L. N. Mukasa, N. Hooper, J. H. Razeq, D. Mulcahy, W. H. Benjamin, and W. R. Bishai. 2001. Epidemiologic usefulness of spoligotyping for secondary typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS6110. *J. Clin. Microbiol.* **39**:3709–3711.
- Diel, R., S. Schneider, K. Meywald-Walter, C. M. Ruf, S. Rusch-Gerdes, and S. Niemann. 2002. Epidemiology of tuberculosis in Hamburg, Germany: long-term population-based analysis applying classical and molecular epidemiological techniques. *J. Clin. Microbiol.* **40**:532–539.
- Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. A. Dang, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, E. Kassa-Kelembho, M. L. Ho, A. Makristathis, C. Mammina, G. Martin, P. Mostrom, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasolof-Razanamparany, M. Ridell, M. L. Rossetti, F. Stauffer, P. N. Suffys, H. Takiff, J. Texier-Maugein, V. Vincent, J. H. de Waard, C. Sola, and N. Rastogi. 2003. Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. *J. Clin. Microbiol.* **41**:1963–1970.
- Filliol, I., A. S. Motiwala, M. Cavatore, W. Qi, M. H. Hazbon, M. Bobadilla del Valle, J. Fyfe, L. Garcia-Garcia, N. Rastogi, C. Sola, T. Zozio, M. I. Guerrero, C. I. Leon, J. Crabtree, S. Angiuoli, K. D. Eisenach, R. Durmaz, M. L. Jobaba, A. Rendon, J. Sifuentes-Osorio, A. Ponce de Leon, M. D. Cave, R. Fleischmann, T. S. Whittam, and D. Alland. 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J. Bacteriol.* **188**:759–772.
- 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.
- Gagneux, S., K. DeRiemer, T. Van, M. Kato-Maeda, B. C. de Jong, S. Narayanan, M. Nicol, S. Niemann, K. Kremer, M. C. Gutierrez, M. Hilty, P. C. Hopewell, and P. M. Small. 2006. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **103**:2869–2873.
- Gagneux, S., and P. M. Small. 2007. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect. Dis.* **7**:328–337.
- García de Viedma, D., N. Alonso Rodriguez, S. Andrés, M. J. Ruiz Serrano, and E. Bouza. 2005. Characterization of clonal complexity in tuberculosis by mycobacterial interspersed repetitive unit-variable-number tandem repeat typing. *J. Clin. Microbiol.* **43**:5660–5664.
- Gutacker, M. M., B. Mathema, H. Soini, E. Shashkina, B. N. Kreiswirth, E. A. Graviss, and J. M. Musser. 2006. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. *J. Infect. Dis.* **193**:121–128.
- Gutierrez, M. C., S. Brisse, R. Brosch, M. Fabre, B. Omais, M. Marmiesse, P. Supply, and V. Vincent. 2005. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog.* **1**:e5.
- Han, H., F. Wang, Y. Xiao, Y. Ren, Y. Chao, A. Guo, and L. Ye. 2007. Utility of mycobacterial interspersed repetitive unit typing for differentiating *Mycobacterium tuberculosis* isolates in Wuhan, China. *J. Med. Microbiol.* **56**:1219–1223.
- Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, and J. van Embden. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* **35**:907–914.
- Kremer, K., C. Arnold, A. Cataldi, M. C. Gutierrez, W. H. Haas, S. Panaiotov, R. A. Skuce, P. Supply, A. G. van der Zanden, and D. van Soolingen. 2005. Discriminatory power and reproducibility of novel DNA typing methods for *Mycobacterium tuberculosis* complex strains. *J. Clin. Microbiol.* **43**:5628–5638.
- Le Flèche, P., M. Fabre, F. Denoed, J. L. Koeck, and G. Vergnaud. 2002. High resolution, on-line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. *BMC Microbiol.* **2**:37.
- Mazars, E., S. Lesjean, A. L. Banuls, M. Gilbert, V. Vincent, B. Gicquel, M. Tibaeyre, C. Locht, and P. Supply. 2001. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc. Natl. Acad. Sci. USA* **98**:1901–1906.
- Nikolayevskiy, V., K. Gopaul, Y. Balabanova, T. Brown, I. Fedorin, and F. Drobniewski. 2006. Differentiation of tuberculosis strains in a population with mainly Beijing-family strains. *Emerg. Infect. Dis.* **12**:1406–1413.
- Oelemann, M. C., R. Diel, V. Vatin, W. Haas, S. Rusch-Gerdes, C. Locht, S. Niemann, and P. Supply. 2007. Assessment of an optimized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. *J. Clin. Microbiol.* **45**:691–697.
- Roring, S., A. Scott, D. Brittain, I. Walker, G. Hewinson, S. Neill, and R. Skuce. 2002. Development of variable-number tandem repeat typing of *Mycobacterium bovis*: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. *J. Clin. Microbiol.* **40**:2126–2133.
- Savine, E., R. M. Warren, G. D. van der Spuy, N. Beyers, P. D. van Helden, C. Locht, and P. Supply. 2002. Stability of variable-number tandem repeats of mycobacterial interspersed repetitive units from 12 loci in serial isolates of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **40**:4561–4566.
- Shamputa, I. C., L. Jugheli, N. Sadradze, E. Willery, F. Portaels, P. Supply, and L. Rigouts. 2006. Mixed infection and clonal representativeness of a single sputum sample in tuberculosis patients from a penitentiary hospital in Georgia. *Respir. Res.* **7**:99.
- Skuce, R. A., T. P. McCorry, J. F. McCarrroll, S. M. Roring, A. N. Scott, D. Brittain, S. L. Hughes, R. G. Hewinson, and S. D. Neill. 2002. Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. *Microbiology* **148**:519–528.
- 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.
- Small, P. M., N. B. McClelleny, S. P. Singh, G. K. Schoolnik, L. S. Tompkins, and P. A. Mickelsen. 1993. Molecular strain typing of *Mycobacterium tuber-*

- culosis* to confirm cross-contamination in the mycobacteriology laboratory and modification of procedures to minimize occurrence of false-positive cultures. *J. Clin. Microbiol.* **31**:1677–1682.
34. **Smittipat, N., P. Billamas, M. Palittapongarnpim, A. Thong-On, M. M. Temu, P. Thanakijcharoen, O. Karnkawinpong, and P. Palittapongarnpim.** 2005. Polymorphism of variable-number tandem repeats at multiple loci in *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **43**:5034–5043.
  35. **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.
  36. **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.
  37. **Supply, P., C. Allix, S. Lesjean, M. Cardoso-Oelemann, S. Rusch-Gerdes, E. Willery, E. Savine, P. de Haas, H. van Deutekom, S. Roring, P. Bifani, N. Kurepina, B. Kreiswirth, C. Sola, N. Rastogi, V. Vatin, M. C. Gutierrez, M. Fauville, S. Niemann, R. Skuce, K. Kremer, C. Locht, and D. van Soolingen.** 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **44**:4498–4510.
  38. **Supply, P., S. Lesjean, E. Savine, K. Kremer, D. van Soolingen, and C. Locht.** 2001. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J. Clin. Microbiol.* **39**:3563–3571.
  39. **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.
  40. **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.
  41. **van Deutekom, H., S. P. Hoijng, P. E. de Haas, M. W. Langendam, A. Horsman, D. van Soolingen, and R. A. Coutinho.** 2004. Clustered tuberculosis cases: do they represent recent transmission and can they be detected earlier? *Am. J. Respir. Crit. Care Med.* **169**:806–810.
  42. **van Deutekom, H., P. Supply, P. E. de Haas, E. Willery, S. P. Hoijng, C. Locht, R. A. Coutinho, and D. van Soolingen.** 2005. Molecular typing of *Mycobacterium tuberculosis* by mycobacterial interspersed repetitive unit-variable-number tandem repeat analysis, a more accurate method for identifying epidemiological links between patients with tuberculosis. *J. Clin. Microbiol.* **43**:4473–4479.
  43. **van Embden, J. D., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. Shinnick, et al.** 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J. Clin. Microbiol.* **31**:406–409.
  44. **Wada, T., S. Maeda, A. Hase, and K. Kobayashi.** 2007. Evaluation of variable numbers of tandem repeat as molecular epidemiological markers of *Mycobacterium tuberculosis* in Japan. *J. Med. Microbiol.* **56**:1052–1057.
  45. **Weniger, T., D. Harmsen, P. Supply, and S. Niemann.** 2007. Mycobacteria MIRU-VNTR<sub>plus</sub>: online database and analysis tool for MIRU, spoligo, and regions of difference data. 105th Gen. Meet. Am. Soc. Microbiol., abstr. no. U-024. American Society for Microbiology, Washington, DC.
  46. **Yeh, R. W., A. Ponce de Leon, C. B. Agasino, J. A. Hahn, C. L. Daley, P. C. Hopewell, and P. M. Small.** 1998. Stability of *Mycobacterium tuberculosis* DNA genotypes. *J. Infect. Dis.* **177**:1107–1111.