JOURNAL OF CLINICAL MICROBIOLOGY, Aug. 2008, p. 2692–2699 Vol. 46, No. 8
0095-1137/08/$08.00+0 doi:10.1128/JCM.00540-08

Evaluation and Strategy for Use of MIRU-VNTR*plus*, a Multifunctional Database for Online Analysis of Genotyping Data and Phylogenetic Identification of *Mycobacterium tuberculosis* Complex Isolates[▽]

Caroline Allix-Béguec,[1] Dag Harmsen,[2] Thomas Weniger,[2] Philip Supply,[4,5]†* and Stefan Niemann[3]†

Genoscreen, 1, rue du Professeur Calmette, Lille 59019 Cedex, France[1]; Department of Periodontology, University Hospital Münster, Waldeyerstrasse 30, Münster D-48149, Germany[2]; Forschungszentrum Borstel, National Reference Center for Mycobacteria, Parkallee 1-40, Borstel 23845, Germany[3]; and INSERM U629[4] and Institut Pasteur de Lille,[5] 1, rue du Professeur Calmette, Lille 59019 Cedex, France

Received 20 March 2008/Returned for modification 17 May 2008/Accepted 31 May 2008

Because of its portable data, discriminatory power, and recently proposed standardization, mycobacterial interspersed repetitive-unit–variable-number tandem-repeat (MIRU-VNTR) typing has become a major method for the epidemiological tracking of *Mycobacterium tuberculosis* complex (MTBC) clones. However, no public MIRU-VNTR database based on well-characterized reference strains has been available hitherto for easy strain identification. Therefore, a collection of 186 reference strains representing the primary MTBC lineages was used to build a database, which is freely accessible at http://www.MIRU-VNTRplus.org. The geographical origin and the drug susceptibility profile of each strain were stored together with comprehensive genetic lineage information, including the 24-locus MIRU-VNTR profile, the spoligotyping pattern, the single-nucleotide- and large-sequence-polymorphism profiles, and the IS*6110* restriction fragment length polymorphism fingerprint. Thanks to flexible import functions, a single or multiple user strains can be analyzed, e.g., for lineage identification with or without the use of reference strains, by best-match or tree-based analyses with single or combined marker data sets. The results can easily be exported. In the present study, we evaluated the database consistency and various analysis parameters both by testing the reference collection against itself and by using an external population-based data set comprising 629 different strains. Under the optimal conditions found, lineage predictions based on typing by 24-locus MIRU-VNTR analysis optionally combined with spoligotyping were verified in >99% of the cases. On the basis of this evaluation, a user strategy was defined, which consisted of best-match analysis followed, if necessary, by tree-based analysis. The MIRU-VNTR*plus* database is a powerful tool for high-resolution clonal identification and has little equivalent in terms of functionalities among the bacterial genotyping databases available so far.

Mycobacterium tuberculosis is among the most successful human pathogens worldwide and is responsible for extensive morbidity and mortality, with approximately 2 million deaths each year (51). The importance of tuberculosis (TB) as a major public health problem has been dramatically reinforced due to the human immunodeficiency virus coepidemic and the emergence of (multi)drug-resistant *M. tuberculosis* strains.

Methods for genotyping of clinical *M. tuberculosis* complex (MTBC) strains have proven to be valuable tools for TB control. At the individual clinical management level, the application of genotyping enables the detection (1) or exclusion (25) of laboratory errors and the follow-up of relapse cases to identify treatment failures, reactivations of latent disease, and exogenous reinfections (46). At the public health level, genotyping enables the detection of unsuspected outbreaks and the identification of transmission chains and secondary cases of infection (4, 46).

Furthermore, the application of genotyping has unraveled the clonal population structure of MTBC, which comprises distinct phylogenetic lineages characterized by differences in their geographical distributions, immunogenicities, virulence, and associations with multidrug-resistant TB (10, 14, 19, 32, 37, 42, 47). Therefore, the recognition of specific *M. tuberculosis* clones (e.g., some clones of the W/Beijing lineage) can be predictive of (multi)drug-resistant TB in certain contexts and can provide indications of the TB case source (which is importation for differentiation between an infection acquired abroad and local transmission). Quantitative analysis of genotyping data may also help with the identification of emerging strains (44). From a research perspective, the accurate identification and study of specific clones worldwide may contribute to the development of new diagnostic, prophylactic, and therapeutic tools for TB control (14; T. Wirth, F. Hildebrand, C. Allix-Béguec, F. Wölbeling, T. Kubica, K. Kremer, D. van Soolingen, S. Rüsch-Gerdes, C. Locht, S. Brisse, A. Meyer, P. Supply, and S. Niemann, submitted for publication).

These objectives require powerful genotyping methods that are phylogenetically informative and standardized and that generate easily comparable data. Multilocus sequence typing is used to genotype many bacterial pathogens but is not applicable to MTBC isolates because of their highly restricted gene sequence variation (37). IS*6110*-based restriction fragment length polymorphism (IS*6110* RFLP) analysis, which has been

* 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.

† P.S. and S.N. contributed equally to the work.

▽ Published ahead of print on 11 June 2008.

the "gold standard" for the genotyping of *M. tuberculosis* for more than a decade, does not meet the conditions of speed and having an easily exchangeable format of fingerprinting data (45). In contrast, the more recently introduced PCR-based method of mycobacterial interspersed repetitive-unit–variable-number tandem-repeat (MIRU-VNTR) typing allows the high-throughput and discriminatory analysis of clinical isolates (12, 38–41). This method generates easily comparable numerical genotypes, similar to other multilocus-VNTR typing methods used for other organisms (21–24, 30, 31). Recently, a MIRU-VNTR typing scheme has been proposed for international standardization on the basis of analysis of the clonal stability and evolutionary rates of MIRU-VNTR markers in a primary genetics lineage of tubercle bacilli collected worldwide (38). This format includes 24 loci, 15 of which were defined as composing a discriminatory subset on the basis of their higher degrees of variability within the different clonal complexes studied. First reports have already shown the appropriateness of its use for the population-based study of TB transmission (3, 29) and for MTBC strain lineage identification based on allelic profiles (3) (Wirth et al., submitted). In addition to standardized MIRU-VNTR typing, PCR-based spoligotyping (20) is proposed as a quick and convenient secondary typing method. Although it is much less discriminatory, this method is especially useful for the easy recognition of MTBC lineages on the basis of the presence or the absence of some specific spacer sequences in the target direct repeat locus and the availability of large databases on the distribution of spoligotypes worldwide (7, 9).

To date, published, freely accessible databases for strain lineage identification have been developed only on the basis of spoligotype signature matching (6, 7, 9, 48). Although they are useful, these tools present significant limitations. The single-locus nature of spoligotyping renders this marker more sensitive to convergence (50), and the interpretation of patterns with various degrees of similarity to prototypes often remains arbitrary. Because of the highly clonal nature of *M. tuberculosis* (19, 37, 42), the use of combinations of multiple phylogenetically informative markers is the best approach to the identification of strain lineages. Here, we present and evaluate a novel freely accessible database that combines standardized MIRU-VNTR typing, spoligotyping, single-nucleotide-polymorphism (SNP), and large-sequence-polymorphism (LSPs) profiles for use for MTBC strain lineage identification.

MATERIALS AND METHODS

Bioinformatics. The bioinformatics tools developed for the MIRU-VNTR*plus* database will be detailed in a separate report. Briefly, the MIRU-VNTR*plus* Web application is implemented in Java programming language (Sun Microsystems Inc., Santa Clara, CA). The user interface consists of pages in hypertext markup language and is generated with JavaServer Faces (Sun) and the RichFaces extension (http://www.jboss.com/; Red Hat Inc., Raleigh, NC) by using Apache Tomcat (http://tomcat.apache.org/; The Apache Software Foundation, Forest Hill, MD) as a servlet container. Asynchronous JavaScript and XML (AJAX; http://www.openajax.org/; OpenAjax Alliance) technology is used to increase the performance of the user interface by reloading only parts of a Web page. The MIRU-VNTR*plus* database makes extended use of JavaScript, e.g., for generating the menus of the application or for using the AJAX technology. Therefore, it is not possible to use the database without JavaScript activated in the browser. Browser cookies are employed to store user settings, e.g., selected distance measure and genotyping methods. However, it is possible to use the service without cookies. The MIRU-VNTR*plus* database works with most modern Web browsers, as it was successfully tested with the Firefox 2 (http://www.mozilla.com/en-US/; Mozilla Corporation, Mountain View, CA), Explorer versions 6 and 7 (http://www.microsoft.com/en/us/default.aspx; Microsoft Corporation, Redmond, WA), Opera 9 (http://www.opera.com/; Opera Software ASA, Oslo, Norway), and Safari 3 (http://www.apple.com/fr/safari/; Apple Inc., Cupertino, CA) Internet browsers. Eclipse (http://www.eclipse.org/; IBM Corporation, New York, NY) was used as the integrated development environment for the whole project. Extensive documentation (online, Adobe PDF manual, and Flash tutorials) on the service and the genotyping methods is available at the MIRU-VNTR*plus* website (http://www.MIRU-VNTRplus.org).

Genetic relationship analysis. Five distance measures are available for strain comparison: categorical distance for all types of data; Jaccard's distance for spoligotyping data; and D_C (chord distance), $(\delta\mu)^2$, and D_{SW} (stepwise weighted distance) distance measures for MIRU-VNTR data only (8, 15, 35). As the most straightforward distance, categorical distance is proposed by default. This distance simply scores the number of markers with a different allele divided by the total number of markers used. It is identical to the D_A distance of Nei and colleagues, which is especially appropriate for phylogenetic analysis with VNTR markers (26, 43). The other genetic distances are provided for exploration purposes under different marker evolution models (for details, see the online help at the MIRU-VNTR*plus* website). Missing data are ignored, in order to accommodate the incorporation of user strains with incomplete data. Calculated distances for each typing method can be combined by using a weighting for each method.

Depending on the distance coefficient chosen, best matching identifies reference genotypes with the closest distance to the test isolate. The cutoff for identification is adjustable. The distance cutoff of 0.17 proposed by default corresponds to a tolerance of, at most, four locus differences or seven spacer differences when 24-locus-based MIRU-VNTR typing or 43-spacer-based spoligotyping is used alone, respectively. When 24-locus MIRU-VNTR typing and 43-spacer spoligotyping methods are combined with equal weights, the respective tolerances are i loci and j spacers, with i equal to 0 to 8, j equal to 0 to 14, and $[0.5 \times (i/24)] + [0.5 \times (j/43)] \leq 0.17$.

Phylogenetic trees are calculated and drawn by using either the unweighted pair group method with arithmetic means (36) or the neighbor-joining (34) algorithm.

Database reference strain collection. A collection of 186 well-characterized strains representing the primary MTBC lineages, as defined by LSPs and spoligotyping (5, 7, 13), was used to build the Internet-based database. This collection was partly described by Supply et al. (38) and comprised 122 *M. tuberculosis* reference strains (of the W/Beijing, Cameroon, Delhi/Central Asian, East African-Indian, Ghana, Haarlem, Latin American-Mediterranean, Turkish, S, Uganda I and II, Ural, and X lineages), 30 *M. africanum* reference strains (of the West African 1 and 2 lineages), 10 *M. bovis* reference strains, 2 *M. canettii* (*M. prototuberculosis*) reference strains, 11 *M. caprae* reference strains, 6 *M. microti* reference strains, and 2 *M. pinnipedii* reference strains. In addition, the ATCC type strains of *M. tuberculosis* H37Rv (ATCC 27294), *M. bovis* (ATCC 19210), and *M. africanum* (ATCC 25420) were included. For each strain, information on the country of isolation and the drug susceptibility profile (if one was available) was collected along with genotype information. The latter comprised the numbers of repeat copies of 24 MIRU loci (38); spoligotypes (20); reference region-of-difference (RD)/LSP profiles, as identified by Brosch et al. (5) and Gagneux et al. (13); SNPs in *gyrB* (27) and *katG* (37); and IS*6110* RFLP fingerprints (45). The correspondence between different strain lineage nomenclatures (7, 9, 13) was included in the database.

Evaluation strain panel. The self-consistency of the database was evaluated by testing the genotyping data for the 186 isolates of the reference collection against the collection itself. Additionally, an external panel of strains from an independent population-based study was used for further assessment. This panel, referred to as the Brussels study panel, included 807 isolates from different notified TB cases in the Brussels-Capital Region of Belgium from 1 September 2002 to 31 December 2005 (3). For the present database evaluation, all the different strains based on 24-locus MIRU-VNTR typing and spoligotyping, corresponding to unique or clustered genotypes, were retained ($n = 629$). This panel was separated into two groups, with one containing preidentified lineage isolates ($n = 442$) and the other containing nonpreidentified lineage isolates or phylogenetically poorly informative spoligotypes (i.e., T spoligotypes) ($n = 187$).

Phylogenetic identification. The genetic lineages within the external evaluation strain panel were preidentified by using the Bionumerics package (Applied Maths, St-Martin-Latem, Belgium). Dendrograms based on the 24-locus MIRU-VNTR typing patterns were generated by using the categorical coefficient and the neighbor-joining algorithm and were rooted by using an *M. canetti* (*M. prototuberculosis*) strain of the C/D genotype (17). Genetic lineages were pre-

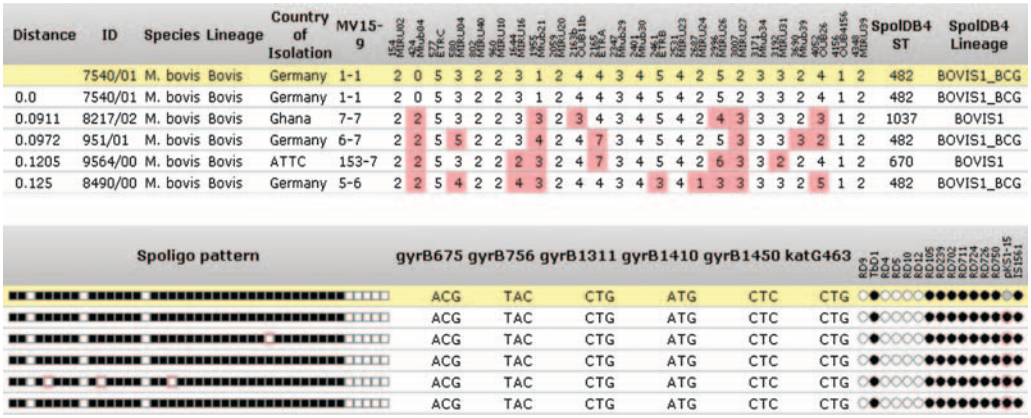

FIG. 1. Typical output of a best-match analysis. Visualization of the MIRU-VNTR copy numbers and spoligotype patterns of a test isolate and its best matches among the reference strain collection is presented along with information on the best-match reference strains, including the strain lineage, country of isolation, MIRU-VNTR type, SNP and RD types, and SpolDB4 shared type and lineage classification. Differences in genotype profiles between the test strain and strains in the reference database are highlighted in red for better recognition.

dicted on the basis of the congruence of the MIRU-VNTR groupings and the corresponding spoligotypes determined by using the existing nomenclatures (9, 14). These predictions were then tested as described in Results by comparison with the reference strains in the MIRU-VNTR*plus* database, whose lineages were additionally identified by their RD/LSP and SNP profiles.

RESULTS

Database organization and main functionalities. The functionalities of the MIRU-VNTR*plus* database will be detailed in a separate report. Briefly, the database allows users to analyze genotyping data for MTBC strains online. The user interface is dynamic and user-friendly, as it allows users to access different functions from either menus or icons. The use of the MIRU-VNTR*plus* database is intuitive, and in addition, online help and step-by-step tutorials are available at different levels.

Genotyping data for user strains can be uploaded either manually strain by strain or uploaded from Excel (Microsoft) files or files with comma-separated values. In the latter case, sets of data for up to 500 strains are accepted. Corresponding template importation files are downloadable. Depending on the genotyping method, various commonly used genotyping formats are accepted (e.g., binary or octal formats for spoligotyping and variant and double alleles for MIRU-VNTR typing).

In principle, two analysis options can be used. Genotyping data for user strains can be compared with those for the reference collection, which comprises 186 isolates representative of the main principal MTBC lineages, in order to identify the corresponding strain lineages. Alternatively, the user's strain genotyping data can be analyzed without using the database content.

Two analytical modes can then be chosen. A first mode relies on the identification of the best matches among genotypes from strains of the reference collection, based on an adjustable distance cutoff (Fig. 1). A second mode proposes dendrogram-based analysis, based on a selected distance coefficient and use of the unweighted pair group method with arithmetic means or the neighbor-joining algorithm. The distance coefficients that are available include Jaccard's and categorical coefficients. In both modes, comparisons can be based on the results of a single genotyping method (any MIRU-VNTR locus combination or spoligotyping, LSP, or SNP data) or on any combination of different data sets.

A summary statistics window displays the resulting clustering rates, allelic diversities, and distribution of the lineages identified. Different modes of visualization of the results can be chosen; for instance, the user can show or hide selected marker profiles corresponding to individual isolates on trees or in best-match groups. Dendrograms or best-match results (including lineage identification for the user's strains) can be exported in various data formats (e.g., MEGA, Newick, SVG, or Microsoft Excel).

Internal database evaluation. The self-consistency of strain lineage identification was first evaluated by testing the best matches of the genotyping data for the 186 isolates of the reference collection against the collection itself (Table 1). Best matching is a method which, depending on the distance coefficient chosen, identifies reference genotypes with the closest distance to the test isolate. In essence, it is a simple parsimony-like method because it uses an optimality criterion, defined as a minimal number of differences in allelic profiles reflected by the calculated distances. This analysis was performed by using the categorical distance with 24-locus MIRU-VNTR typing data alone or in combination with spoligotyping data, while the SNP, LSP, and IS*6110* RFLP profiles were used only as confirmatory markers. Under these conditions, each isolate submitted trivially best matched first against itself (data not shown). The objective was thus to determine which proportion of the isolates had a correct result of a second best match with another isolate of the same lineage, depending on the distance cutoff selected. Therefore, the sensitivity and the specificity for lineage detection were defined here as the proportion of correct best matches identified among the test samples and the proportion of correct best matches found among the total best matches identified, respectively.

A stringent distance cutoff of 0.17 was initially selected (Table 1) and corresponds to a tolerance of, at most, a four-locus difference when MIRU-VNTR typing is used alone. Under these conditions, a second-best-match result was obtained for 147 and 170 of the 186 strains for MIRU-VNTR typing alone

TABLE 1. Internal database evaluation[a]

Result	No. (%) of isolates with the indicated result by the following method(s) with the indicated cutoff					
	24-locus MIRU-VNTR typing		24-locus MIRU-VNTR typing + spoligotyping		Spoligotyping	
	0.17	0.3	0.17	0.3	0.17	0.3
Lineage confirmation[b]	147 (79.0)	177 (95.1)	170 (91.4)	183 (98.4)	175 (94.1)	178 (95.7)
No match	39 (21.0)	7 (3.8)	16 (8.6)	1 (0.5)	4 (2.2)	0 (0.0)
Conflict[c]	0 (0.0)	2 (1.1)	0 (0.0)	2 (1.1)	7 (3.8)	8 (4.3)

[a] The self-consistency of the strain lineage identification was evaluated by testing the best matches of the genotyping data for the 186 isolates of the reference collection against the collection itself.
[b] For lineage confirmation data, percentages correspond to sensitivity of lineage identification, defined here as the proportion of correct best matches identified among the test samples.
[c] The specificity for lineage identification equals 100 − conflict (in percent). Specificity is defined here as the portion of correct best matches found among the total best matches identified.

and MIRU-VNTR typing combined with spoligotyping, respectively. In both cases, these best matches always correctly occurred with another isolate from the same lineage, thus resulting in a specificity of 100% and sensitivities of 79.0 and 91.4%, respectively.

When the distance cutoff was relaxed to >0.3 (i.e., a tolerance of up to seven loci of difference when MIRU-VNTR typing was used alone), further best matches were detected, resulting in a sensitivity of 95.1% for MIRU-VNTR typing alone and a sensitivity of even 98.4% when MIRU-VNTR typing was combined with spoligotyping. However, two apparent mismatches were then observed in both cases, resulting in a specificity of 98.9%. For MIRU-VNTR typing alone, the two mismatches were a strain of the so-called New 1 genotype with a Haarlem strain and, less unexpectedly, another X-genotype strain with a Haarlem strain. When MIRU-VNTR typing was combined with spoligotyping, the two mismatches detected involved a Cameroon isolate with a Haarlem isolate and, to a lesser degree, an X-genotype strain with a Haarlem strain.

The use of spoligotyping alone with cutoffs of >0.3 and 0.17 resulted in more mismatches and, thus, lower specificities, which ranged from 95.7 to 96.2% (Table 1).

When 24-locus MIRU-VNTR typing data were used for tree-based analyses of the reference collection by using, e.g., the neighbor-joining algorithm and the categorical distance coefficient, consistent intralineage groupings were systematically obtained (by using the corresponding analysis parameters; see http://www.miru-vntrplus.org). The lineages were all monophyletic or, in the case of the Haarlem strains, singly paraphyletic by expectedly comprising a distinct subgroup of X genotypes (see Discussion). In tree-based analyses, the use of spoligotyping alone or in combination with MIRU-VNTR typing data resulted in more conflicting groupings (data not shown).

External database evaluation. MIRU-VNTR typing and spoligotyping data for 629 different strains from the population-based Brussels-Capital Region collection were used to externally evaluate the database. This data set comprised 442 different strains with a preidentified lineage determined on the basis of the congruence of both typing methods (Table 2).

First, the same best-match-based analyses described above were performed with these 442 genotypes, except that 15-locus MIRU-VNTR typing data were additionally considered. As for the internal evaluation, the specificity was almost perfect at a stringent distance cutoff of 0.17, with values ranging from 99.3% to 100%, regardless of the method(s) used (15- or 24-locus-based MIRU-VNTR typing combined or not combined with spoligotyping). However, the combination of 24-locus MIRU-VNTR typing and spoligotyping again offered the best results by combining a specificity of 100% and a sensitivity of 72.2%, corresponding to 319 of the 442 strains with a detected best match that systematically met the previous independent

TABLE 2. External evaluation with isolates whose lineages were preidentified[a]

Cutoff and result	No. (%) of isolates with the indicated result by:			
	24-locus MIRU-VNTR typing + spoligotyping	15-locus MIRU-VNTR typing + spoligotyping	24-locus MIRU-VNTR typing	15-locus MIRU-VNTR typing
Cutoff of 0.17				
Lineage confirmation[b]	319 (72.2)	278 (62.9)	240 (54.3)	160 (36.2)
No match	123 (27.8)	163 (36.9)	199 (45.0)	282 (63.8)
Conflict[c]	0 (0.0)	1 (0.2)	3 (0.7)	0 (0.0)
Cutoff of 0.3				
Lineage confirmation[b]	419 (94.8)	403 (91.2)	392 (88.7)	327 (74.0)
No match	8 (1.8)	32 (7.2)	36 (8.1)	106 (24.0)
Conflict[c]	15 (3.4)	7 (1.6)	14 (3.2)	9 (2.0)

[a] Best-match-based analyses were performed with the 442 isolates with a preidentified lineage from the Brussels-Capital Region collection.
[b] For lineage confirmation data, percentages correspond to sensitivity of lineage identification, defined here as the proportion of correct best matches identified among the test samples.
[c] The specificity for lineage identification equals 100 − conflict (in percent). Specificity is defined here as the portion of correct best matches found among the total best matches identified.

TABLE 3. External evaluation with unknown or T-spoligotype isolates[a]

Cutoff and result	No. (%) of isolates with the indicated result by:			
	24-locus MIRU-VNTR typing + spoligotyping	15-locus MIRU-VNTR typing + spoligotyping	24-locus MIRU-VNTR typing	15-locus MIRU-VNTR typing
Cutoff of 0.17				
Best match	62 (33.2)	21 (11.2)	16 (8.6)	8 (4.3)
No match	125 (66.8)	166 (88.8)	171 (91.4)	179 (95.7)
Cutoff of 0.17				
Best match	179 (95.7)	172 (92.0)	164 (87.7)	57 (30.5)
No match	8 (4.3)	15 (8.0)	23 (12.3)	130 (69.5)

[a] Best-match-based analyses were performed with the 187 unknown or T-spoligotype isolates from the Brussels-Capital Region collection.

strain identification. Again, the sensitivity could be significantly improved to 94.8% by relaxing the distance cutoff to 0.3, but at the cost of 15 mismatches, which reduced the specificity to 96.6%.

We then examined how tree-based analysis could complement these best-match results. When 24-locus MIRU-VNTR typing data were used together with the neighbor-joining algorithm and categorical distance coefficient, consistent groupings within the expected reference lineages were obtained for 439 of the 442 test genotypes (data not shown). The three isolates grouped with unpredicted references were one West African 2 lineage isolate grouped with seal isolates, one Cameroon lineage isolate grouped with LAM isolates, and one X-lineage isolate grouped with New 1 isolates. Less unexpectedly, one X-lineage isolate was found to be grouped with Haarlem isolates. As for the tree-based analysis described above, the use of spoligotyping alone or in combination with MIRU-VNTR typing resulted in more conflicting groupings (data not shown).

When the 187 isolates for which no lineage or only T spoligotypes (including variants) were preassigned were considered, either no best match was found or the best matches found were distributed among the different lineages represented in the database, as expected (Table 3). For these isolates, the proportion of best matches found again increased by relaxing the distance cutoff from 0.17 to 0.3, e.g., from 33.2 to 95.7% for 24-locus MIRU-VNTR typing and spoligotyping combined. In any case, all best matches detected for these isolates were found among lineages that all belonged to the Euro-American superlineage of *M. tuberculosis*. Consistently, tree-based analysis by 24-locus MIRU-VNTR typing distributed all these isolates among the same Euro-American branches. However, a large set of isolates (which had T spoligotypes and which were in the majority) for which no best matches were detected was grouped into a large cluster that appeared to be distinct from the other Euro-American branches (Fig. 2).

DISCUSSION

In this report, we evaluate the first multifunctional Web-based database for the combined analysis of bacterial genotyping data obtained by up to five different typing methods (although reference IS*6110* RFLP patterns are provided only for visual comparison). This tool allows the genotype/phylogenetic lineage classification of MTBC isolates online by comparison

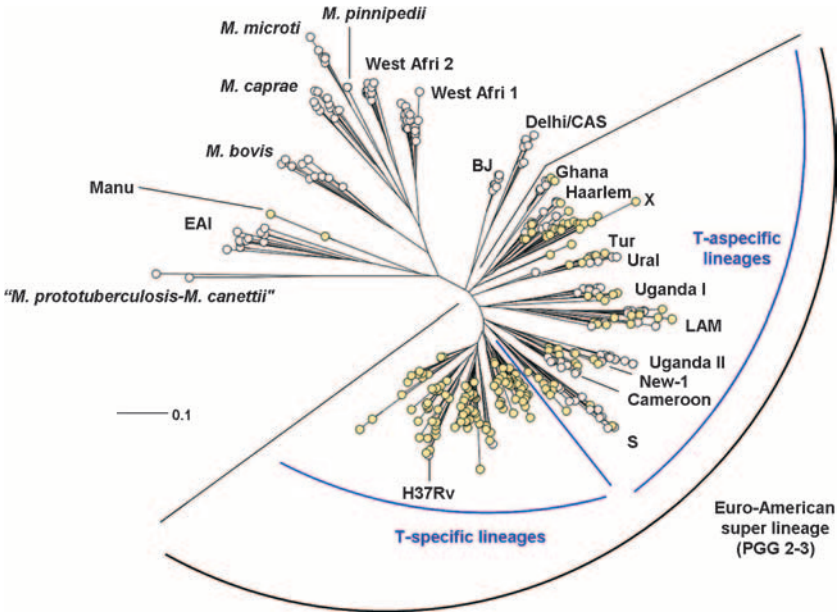

FIG. 2. Phylogenetic distribution of isolates with unassigned or T spoligotypes. A radial tree was constructed by using the 24-locus MIRU-VNTR typing data for the reference strains in the MIRU-VNTR*plus* database and 187 genotypes with unassigned or T spoligotypes from the Brussels population-based collection by using the neighbor-joining algorithm and categorical distance coefficient. The positions of Brussels genotypes with unassigned or T spoligotypes are indicated by yellow circles. Lineages containing exclusively T-spoligotype strains or lineages corresponding to well-characterized, classical clades of the Euro-American superlineage (such as the Haarlem lineage) comprising T-spoligotype strains are denoted T-specific and T-aspecific, respectively. EAI, East African-Indian; West Afri 2, West African 2 (*M. africanum* 2); West Afri 1, West African 1 (*M. africanum* 1); Tur, Turkish; BJ, Beijing; CAS, Central Asian; LAM, Latin American-Mediterranean.

with a reference strain collection covering the primary branches of this bacterial complex, as defined by LSP analysis (5, 13) and the global distribution of spoligotype families in SpolDB4 (7). All the reference strains have been characterized by quality-controlled, state-of-the-art 24-locus MIRU-VNTR typing and spoligotyping, which together tend to replace IS*6110* RFLP as the internationally standardized gold standard for molecular epidemiological studies (1, 29, 38). These strains have additionally been characterized by the use of reference SNP and LSP markers, which collectively ensures their accurate MTBC "species" and lineage identification (5, 13, 14, 27).

We have chosen to construct this reference database in a closed format; i.e., users can freely submit their genotypes for identification, but these genotypes are not added to the reference database. This simplifies maintenance of the database by the authors and ensures the reliability of the reference data set by avoiding the need to curate externally submitted data.

The database distinctively offers user-friendly interfaces, tutorials and help, and multiple convenient tools for clustering and phylogenetic analysis of users' isolates as well as for data import and export. In all analysis modes, the database can be used to visualize supplementary information, in addition to primary genotyping profiles, which can guide additional analyses (e.g., by using lineage-specific LSPs) to confirm the initial classification, if necessary. Therefore, the use of this database provides many functionalities of expensive commercial software packages such as the Bionumerics package beyond just the classical report of matching "genotype" codes. Nevertheless, it also includes functionality for automated online SpolDB4-based spoligotype sequence type and spoligotype clade assignation, as well as for MIRU-VNTR*plus* database-based MIRU-VNTR type codification. This standardized MIRU-VNTR type classification relies on the juxtaposition of two codes, corresponding to types based on 15 more discriminatory loci and 9 ancillary loci, respectively. This dual-code scheme thus accommodates the classification in a single system of genotypes obtained with either the 15-locus-based discriminatory subset or the complete 24-locus format (38). In addition, this codification system leaves the door open for the inclusion of an additional code based on a few other MIRU-VNTR loci that may prove specifically necessary for the discrimination of *M. bovis* strains (2, 33) and, perhaps, W/Beijing strains (18, 28, 49).

In contrast to other available MTBC genotyping databases, the multimarker basis of the MIRU-VNTR*plus* database also allowed us to test optimal parameters and define predictive values for phylogenetic identification. Evaluations were performed by simulating the most frequent utilization conditions, i.e., by basing comparisons on MIRU-VNTR typing data alone or in combination with spoligotyping data, and independently by using the SNP and LSP profiles of reference strains for confirmation. Both an internal evaluation (with the reference strains themselves) and an external evaluation were performed. The external evaluation made use of the largest population-based standardized MIRU-VNTR data set available so far. The corresponding study included 807 TB patients from the Brussels-Capital Region, 76% of whom were foreign born and were from 69 different countries. As a reflection of the MTBC phylogeography, 12 spoligotype families, including all the major strain lineages, were identified among the patient isolates. This population thus offered an opportunity to test the consistency of the MIRU-VNTR*plus* predictions with a wide range of strain lineages of diverse geographical origins (3).

Both the internal and the external evaluations indicated the self-consistency and the usefulness of the database. Consistent with the fact that parsimony-based methods generally perform well on restricted genetic perimeters, best-match analysis with a stringent distance cutoff was found to provide highly specific identifications when it was performed with strains with preidentified lineages. The selected cutoff of 0.17, equivalent to, at most, a four-locus difference among 24 MIRU-VNTR loci by using categorical distance, approximately corresponds to the most conservative definition of clonal complexes by multilocus sequence typing, based on the sharing of identical alleles at six of the seven gene sequence targets studied. Best-match results met expectations in 99.3 to 100% of the cases for MIRU-VNTR typing alone and in 100% of the cases when MIRU-VNTR typing was combined with spoligotyping. The use of these methods in combination also logically provided the highest sensitivity (defined as the proportion of predictions among the isolates tested), which, however, did not exceed 91.4% and 72.2% in the internal and external evaluations, respectively. This level of sensitivity is explained by the fact that the different lineages in the current database do not yet systematically include very closely related clones that could best match any test strain at this stringent distance cutoff. Relaxing the distance cutoff to 0.3 improved the sensitivity to 98.4% and 94.8% in the internal and external evaluations, respectively, but not surprisingly, this cost up to 15 mismatches in the external evaluation. However, it is noteworthy that these mismatches reflect just a lack of fine-tuning at this cutoff rather than true inconsistencies, as virtually all of these mismatches were concentrated among strains from different branches that all consistently belong to the Euro-American superlineage (14).

In order to maximize predictions and minimize fine-tune mismatching, a better compromise was obtained by using tree-based methods based on 24 MIRU-VNTR loci. In the internal evaluation, the groupings of the reference strains were fully consistent. All lineages except Haarlem were monophyletic; the Haarlem lineage included an distinct lineage X subgroup. However, this result was expected, as this branching is also recognized by other analytic methods (10, 16). In the external tree-based tests with strains of preidentified lineages, groupings within expected reference lineages were obtained for 439/442 (99.3%) test genotypes. It is noteworthy that in contrast to the results of best-match analysis, the use of spoligotyping in combination with MIRU-VNTR typing data (or alone) resulted in more conflicting groupings (data not shown). This reflects the approximation from the use of simple spoligotyping spacer number differences in distance-based methods, although this monolocus marker can evolve by the loss of a single spacer or a spacer block in a single event.

In the external evaluation, 30% of the isolates had no preassigned lineage, and the external evaluation included a majority of T-spoligotype strains and variants thereof, which are known to be poorly phylogenetically informative (10). They were considered separately, since their precise classification by MIRU-VNTR typing (possibly combined with spoligotyping) could not be extrinsically confirmed. Nevertheless, their broad

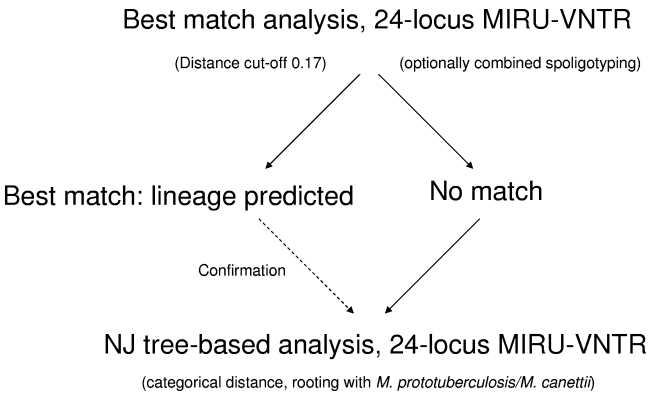

FIG. 3. Strategy for use of phylogenetic identification. If no match is detected after an initial best-match analysis with a stringent distance cutoff of 0.17, tree-based analysis is preferred upon best-match analysis with a relaxed distance cutoff of 0.3 as a second step, because the former method minimizes the fine-tuned mismatching that can occur as an exception among strains belonging to the Euro-American superlineage (see text). NJ, neighbor-joining.

classification met expectations on the basis of previous knowledge. Except for two isolates with a full 43-spacer MANU spoligotype (7, 11), both best-match analysis (when isolates were detected) and tree-based analysis classified all the isolates in the Euro-American superlineage of *M. tuberculosis*, which corresponds to principal genetic groups 2 and 3 (10, 14, 16, 37). Consistent with SNP-based analyses, a number of these T-spoligotype strains were associated with well-characterized lineages, such as Haarlem and Cameroon, while a large set of strains appeared to form a distinct T-spoligotype-specific group (Fig. 2), likely corresponding to SNP-defined cluster 6b or VIII (10, 16).

In conclusion, on the basis of these evaluations, we propose the following strategy for the use of optimal phylogenetic identification (Fig. 3). This scheme integrates best-match analysis at a high stringency based on MIRU-VNTR typing optionally combined with spoligotyping, followed by tree-based analysis based on 24 MIRU-VNTR loci to maximize the predictions, if necessary. This strategy blends the specificity and straightforward interpretation of best-match analysis with the sensitivity of tree-based identification. In our tests, preidentified lineages of strains were thereby verified in >99% of the cases. On the basis of these results and the consistency of the results with the SNP-based results (see above), we think that predictions of the phylogeny of *M. tuberculosis* strains with unknown or weakly informative spoligotypes such as the T spoligotype will reflect their true lineage connection. Depending on the cases, these strains will most often be classified with reference strains from different well-identified lineages or in a distinct group within the Euro-American superlineage. The upcoming inclusion of additional reference strains, including those with the T spoligotype, will further ensure these predictions and will concomitantly increase the sensitivity of best-match analysis at the stringent distance cutoff. In addition to the expansion of the reference strain database, we envision opening of a second, open database to which users would be allowed to add their strain genotypes. However, such an open database will require the development of efficient quality control algorithms. Regardless, on the basis of the results of the present evaluation, we believe that the current version of the MIRU-VNTR*plus* database already constitutes a powerful tool for the clonal identification of MTBC isolates. It particularly enables the better exploitation of the advantages of MIRU-VNTR typing, including its wide use and recently proposed international standardization. It may serve as a basis for the development of multilocus-VNTR typing-based databases for the typing of other species.

ACKNOWLEDGMENTS

We thank I. Radzio, T. Ubben, and P. Vock for excellent technical assistance.

Parts of this work were supported by INSERM and the Institut Pasteur de Lille, Lille, France. P.S. is a researcher of the Centre National de la Recherche Scientifique. S.N. was supported by the Germany Ministry of Health and the German Federal Ministry for Education and Research (BMBF) within the PathoGenomikPlus Network.

REFERENCES

1. **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.
2. **Allix, C., K. Walravens, C. Saegerman, J. Godfroid, P. Supply, and M. Fauville-Dufaux.** 2006. Evaluation of the epidemiological relevance of variable-number tandem-repeat genotyping of *Mycobacterium bovis* and comparison of the method with IS*6110* restriction fragment length polymorphism analysis and spoligotyping. J. Clin. Microbiol. **44:**1951–1962.
3. **Allix-Béguec, C., M. Fauville-Dufaux, and P. Supply.** 2008. Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit–variable-number tandem-repeat typing of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **46:**1398–1406.
4. **Barnes, P. F., and M. D. Cave.** 2003. Molecular epidemiology of tuberculosis. N. Engl. J. Med. **349:**1149–1156.
5. **Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole.** 2002. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc. Natl. Acad. Sci. USA **99:**3684–3689.
6. **Brown, J. R.** 2003. Ancient horizontal gene transfer. Nat. Rev. Genet. **4:**121–132.
7. **Brudey, K., J. R. Driscoll, L. Rigouts, W. M. Prodinger, A. Gori, S. A. Al-Hajoj, C. Allix, L. Aristimuño, 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, M. L. Ho, C. Martin, C. Martin, I. Mokrousov, O. Narvskaïa, Y. F. Ngeow, L. Naumann, S. Niemann, I. Parwati, M. Z. Rahim, V. Rasolofo-Razanamparany, T. Rasolonavalona, M. L. Rossetti, S. Rüsch-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.
8. **Cavalli-Sforza, L. L., and A. W. F. Edwards.** 1967. Phylogenetic analysis. Models and estimation procedures. Am. J. Hum. Genet. **19:**233–257.
9. **Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valétudie, D. D. Anh, 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, H. M. Ly, A. Makristathis, C. Mammina, G. Martin, P. Moström, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasolofo-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.
10. **Filliol, I., A. S. Motiwala, M. Cavatore, W. Qi, M. H. Hazbón, M. Bobadilla del Valle, J. Fyfe, L. García-García, N. Rastogi, C. Sola, T. Zozio, M. I. Guerrero, C. I. León, J. Crabtree, S. Angiuoli, K. D. Eisenach, R. Durmaz, M. L. Joloba, A. Rendón, J. Sifuentes-Osornio, A. Ponce de León, 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.

11. **Flores, L., T. Van, S. Narayanan, K. DeRiemer, M. Kato-Maeda, and S. Gagneux.** 2007. Large sequence polymorphisms classify *Mycobacterium tuberculosis* strains with ancestral spoligotyping patterns. J. Clin. Microbiol. **45:**3393–3395.
12. **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.
13. **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.
14. **Gagneux, S., and P. M. Small.** 2007. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. Lancet Infect. Dis. **7:**328–337.
15. **Goldstein, D. B., A. R. Linares, L. L. Cavalli-Sforza, and M. W. Feldman.** 1995. An evaluation of genetic distances for use with microsatellite loci. Genetics **139:**463–471.
16. **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.
17. **Gutierrez, M. C., S. Brisse, R. Brosch, M. Fabre, B. Omaïs, M. Marmiesse, P. Supply, and V. Vincent.** 2005. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. PLoS Pathog. **1:**e5.
18. **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.
19. **Hirsh, A. E., A. G. Tsolaki, K. DeRiemer, M. W. Feldman, and P. M. Small.** 2004. Stable association between strains of Mycobacterium tuberculosis and their human host populations. Proc. Natl. Acad. Sci. USA **101:**4871–4876.
20. **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.
21. **Keim, P., A. M. Klevytska, L. B. Price, J. M. Schupp, G. Zinser, K. L. Smith, M. E. Hugh-Jones, R. Okinaka, K. K. Hill, and P. J. Jackson.** 1999. Molecular diversity in Bacillus anthracis. J. Appl. Microbiol. **87:**215–217.
22. **Klevytska, A. M., L. B. Price, J. M. Schupp, P. L. Worsham, J. Wong, and P. Keim.** 2001. Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. J. Clin. Microbiol. **39:**3179–3185.
23. **Koeck, J.-L., B.-M. Njanpop-Lafourcade, S. Cade, E. Varon, L. Sangare, S. Valjevac, G. Vergnaud, and C. Pourcel.** 2005. Evaluation and selection of tandem repeat loci for Streptococcus pneumoniae MLVA strain typing. BMC Microbiol. **5:**66.
24. **Lindstedt, B.-A., E. Heir, E. Gjernes, and G. Kapperud.** 2003. DNA fingerprinting of *Salmonella enterica* subsp. *enterica* serovar Typhimurium with emphasis on phage type DT104 based on variable number of tandem repeat loci. J. Clin. Microbiol. **41:**1469–1479.
25. **Loïez, C., E. Willery, J.-L. Legrand, V. Vincent, M. C. Gutierrez, R. J. Courcol, and P. Supply.** 2006. Against all odds: molecular confirmation of an implausible case of bone tuberculosis. Clin. Infect. Dis. **42:**e86–e88.
26. **Nei, M., F. Tajima, and Y. Tateno.** 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J. Mol. Evol. **19:**153–170.
27. **Niemann, S., D. Harmsen, S. Rüsch-Gerdes, and E. Richter.** 2000. Differentiation of clinical *Mycobacterium tuberculosis* complex isolates by *gyrB* DNA sequence polymorphism analysis. J. Clin. Microbiol. **38:**3231–3234.
28. **Nikolayevskyy, 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.
29. **Oelemann, M. C., R. Diel, V. Vatin, W. Haas, S. Rüsch-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.
30. **Onteniente, L., S. Brisse, P. T. Tassios, and G. Vergnaud.** 2003. Evaluation of the polymorphisms associated with tandem repeats for *Pseudomonas aeruginosa* strain typing. J. Clin. Microbiol. **41:**4991–4997.
31. **Pourcel, C., Y. Vidgop, F. Ramisse, G. Vergnaud, and C. Tram.** 2003. Characterization of a tandem repeat polymorphism in *Legionella pneumophila* and its use for genotyping. J. Clin. Microbiol. **41:**1819–1826.
32. **Reed, M. B., P. Domenech, C. Manca, H. Su, A. K. Barczak, B. N. Kreiswirth, G. Kaplan, and C. E. Barry III.** 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature **431:**84–87.
33. **Roring, S., A. N. Scott, R. G. Hewinson, S. D. Neill, and R. A. Skuce.** 2004. Evaluation of variable number tandem repeat (VNTR) loci in molecular typing of Mycobacterium bovis isolates from Ireland. Vet. Microbiol. **101:** 65–73.
34. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
35. **Shriver, M. D., L. Jin, E. Boerwinkle, R. Deka, R. E. Ferrell, and R. Chakraborty.** 1995. A novel measure of genetic distance for highly polymorphic tandem repeat loci. Mol. Biol. Evol. **12:**914–920.
36. **Sokal, R. R., and C. D. Michener.** 1958. A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull. **28:**1409–1438.
37. **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.
38. **Supply, P., C. Allix, S. Lesjean, M. Cardoso-Oelemann, S. Rüsch-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.
39. **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.
40. **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.
41. **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.
42. **Supply, P., R. M. Warren, A. L. Bañuls, S. Lesjean, G. D. van der Spuy, L.-A. Lewis, M. Tibayrenc, P. D. van Helden, and C. Locht.** 2003. Linkage disequilibrium between minisatellite loci supports clonal evolution of Mycobacterium tuberculosis in a high tuberculosis incidence area. Mol. Microbiol. **47:**529–538.
43. **Takezaki, N., and M. Nei.** 2008. Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. Genetics **178:**385–392.
44. **Tanaka, M. M., and A. R. Francis.** 2006. Detecting emerging strains of tuberculosis by using spoligotypes. Proc. Natl. Acad. Sci. USA **103:**15266–15271.
45. **van Embden, J. D. A., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. 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.
46. **Van Soolingen, D., K. Kremer, and E. Vynycky.** 2003. New perspectives in the molecular epidemiology of tuberculosis, p. 17–45. In S. H. E. Kaufmann and H. Hahn (ed.), Mycobacteria and TB: issues in infectious diseases, vol. 2. Karger, Berlin, Germany.
47. **van Soolingen, D., L. Qian, P. E. W. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, 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.
48. **Vitol, I., J. Driscoll, B. Kreiswirth, N. Kurepina, and K. P. Bennett.** 2006. Identifying Mycobacterium tuberculosis complex strain families using spoligotypes. Infect. Genet. Evol. **6:**491–504.
49. **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.
50. **Warren, R. M., E. M. Streicher, S. L. Sampson, G. D. van der Spuy, M. Richardson, D. Nguyen, M. A. Behr, T. C. Victor, and P. D. van Helden.** 2002. Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. J. Clin. Microbiol. **40:**4457–4465.
51. **World Health Organization.** 2006. Global tuberculosis control. WHO Report. World Health Organization, Geneva, Switzerland.