Identification and Genotyping of *Mycobacterium tuberculosis* Complex Species by Use of a SNaPshot Minisequencing-Based Assay

C. Bouakaze,1* C. Keyser,1 S. J. de Martino,2 W. Sougakoff,3,4 N. Veziris,4,5 H. Dabernat,6 and B. Ludes1

EA4438 Physiopathologie et Médecine Translationnelle, Université de Strasbourg, Institut de Médecine Légale, Strasbourg, France; Laboratoire de Bactériologie, Faculté de Médecine, Université de Strasbourg, Strasbourg, France; AP-HP, Hôpital Pitié-Salpêtrière, Laboratoire de Bactériologie-Hygienne, Paris, France; and Laboratoire d’Anthropologie Moleculaire et Imagerie de Synthèse (AMIS), CNRS FRE 2960, Université Paul Sabatier, Toulouse, France

Received 18 November 2009/Returned for modification 17 December 2009/Accepted 25 February 2010

The aim of the present study was to investigate the use of the SNaPshot minisequencing method for the identification of Mycobacterium tuberculosis complex (MTBC) isolates to the species level and for further genotyping of *M. tuberculosis* isolates. We developed an innovative strategy based on two multiplex allele-specific minisequencing assays that allowed detection of eight species-specific and eight lineage-specific single nucleotide polymorphisms (SNPs). Each assay consisted of an eightplex PCR amplification, followed by an eightplex minisequencing reaction with the SNaPshot multiplex kit (Applied Biosystems) and, finally, analysis of the extension products by capillary electrophoresis. The whole strategy was developed with a panel of 56 MTBC strains and 15 negative controls. All MTBC strains tested except one *M. africanaum* clinical isolate were accurately identified to the species level, and all *M. tuberculosis* isolates were successfully further genotyped. This two-step strategy based on SNaPshot minisequencing allows the simultaneous differentiation of closely related members of the MTBC, the distinction between principal genetic groups, and the characterization of *M. tuberculosis* isolates into one of the seven prominent SNP cluster groups (SCGs) and could be a useful tool for diagnostic and epidemiological purposes.

Although tuberculosis (TB) is an age-old disease, it still represents a major health problem worldwide, accounting for nearly 2 million deaths annually (World Health Organization, Tuberculosis Facts 2009 [http://www.who.int/tb/publications /factsheets/en/]). Besides the need for an improved therapy and for new vaccines, effective TB control requires the initiation of appropriate therapy as well as an increased understanding of its epidemiology.

The causative agents of TB in humans and animals, including *Mycobacterium tuberculosis*, *M. africanaum*, *M. bovis*, *M. canettii*, *M. microti*, *M. caprae*, and *M. pinnipedii*, form the *Mycobacterium tuberculosis* complex (MTBC). Although they differ widely in terms of their host tropisms, phenotypes, and pathogenicities, all members of the MTBC are closely related genetically (51). *M. tuberculosis* species, the most common pathogen in humans, can be further divided into genetic groups that also show differences in their levels of virulence, immunogeneticities, and geographical distributions (21). On the one hand, it is important to differentiate MTBC species to distinguish between strict human and zoonotic TB and to initiate an appropriate therapy (15). In particular, the distinction between *M. tuberculosis* and *M. bovis* is necessary, as the latter species is naturally resistant to the antituberculous drug pyrazinamide (48). On the other hand, genotyping of *M. tuberculosis* isolates is useful as a means of addressing evolutionary questions but also as a means of surveying the transmission dynamics of this pathogen and identifying new outbreaks.

Identification of MTBC isolates to the species level is so far routinely performed by analysis of the phenotypic and biochemical characteristics of the bacteria after culture. However, this is a time-consuming and subjective process, and that is why various molecular methods have been developed in recent years. Most molecular methods for MTBC species differentiation described in the literature are based on the analysis of genomic deletions by PCR, followed by agarose gel electrophoresis (30, 40, 56), or are based on the detection of single nucleotide polymorphisms (SNPs) by PCR-restriction fragment length polymorphism (PCR-RFLP) analysis, which consists of digestion of the PCR product with a restriction enzyme, followed by agarose gel electrophoresis (3, 13, 23, 33, 38). A commercial kit based on a DNA strip assay for the analysis of gyrB gene SNPs, 23S rRNA, and the RD1 deletion, named GenoType MTBC (Hain Lifescience GmbH, Nehren, Germany), has recently become available (37, 42, 50). This line probe assay enables identification of the presence of various members of the MTBC but does not differentiate *M. canettii* from *M. tuberculosis* and *M. africanaum* type I from *M. pinnipedii*.

Various molecular methods for genotyping of *M. tuberculosis* have also been developed and were recently reviewed by...
Mathema and colleagues (36). The most commonly used methods include the “gold standard” analysis of highly conserved DNA fingerprinting patterns obtained by RFLP analysis of the insertion sequence IS6110 (IS6110-RFLP fingerprinting) (54), the study of variations within the genomic direct repeat region by spoligotyping (10, 32), and the typing of mycobacterial interspersed repetitive-unit–variable–number of DNA tandem repeats (MIRU-VNTR) (52). They are powerful molecular epidemiological typing methods since they are based on the analysis of mobile DNA elements (IS6110-RFLP fingerprinting) or repetitive DNA elements (spoligotyping and MIRU-VNTR typing) that change quite rapidly and thus provide a high degree of discriminatory power, but they are less useful for definition of phylogenetic groupings (14, 21). Recently, two studies have revealed that SNPs are valuable phylogenetically informative markers (18, 26), and the use of a minimal number of sets of SNPs has been proposed as a means of resolving M. tuberculosis and M. bovis isolates into seven genetic groups called SNP cluster groups (SCGs) (1, 18).

Thus, SNPs can be useful markers for both the identification of MTBC species and the genotyping of M. tuberculosis isolates. We describe here the development of an SNP typing-based strategy that was designed to simultaneously distinguish members of the MTBC to the species level and M. tuberculosis lineages. Among the various SNP typing methodologies described in the literature, we selected the SNaPshot minsequencing-based approach because of its high multiplexing capacity, robustness, and extreme sensitivity (22, 49). Moreover, it is a relatively simple and affordable method that requires the use of a thermal cycler and a genetic analyzer, which are types of equipment commonly available in molecular biology laboratories. This approach is based on the single-base extension (SBE) of an unlabeled minisequencing primer that anneals one base upstream of the relevant SNP with a fluorochrome-labeled deoxyoxynucleotide (ddNTP). The allelic state is then determined after separation of the extension products and the detection of fluorescence by capillary electrophoresis. This SNP typing method is currently widely used in the fields of forensic and population genetics (8, 9, 25, 47) and has recently been proposed for use by Alland and coworkers was chosen (the bases at the 5′ end of the sequencing primer were selected for use in the first SNP set). In our study, the first eight SNPs were selected as “species specific,” whereas the last eight were “lineage specific” or “genotype specific.”

**Materials and Methods**

**Bacterial strains and test reference tests.** The following 56 MTBC strains were used for the development of the assays: M. tuberculosis H37Rv ATCC 27294; M. bovis CIP 102426; M. bovis BCG CIP 105226; and clinical isolates of M. tuberculosis (n = 35), M. bovis (n = 6), M. bovis BCG (n = 4), M. africanum (n = 4), M. canetti (n = 1), M. caprae (n = 1), M. microti (n = 1), and M. pinnipedi (n = 1) (the M. pinnipedi isolate was recovered from a tapir, and the others were isolated from human clinical samples). Eleven clinical isolates of mycobacteria other than members of the M. tuberculosis complex (MOTT), comprising M. fortuitum (n = 2), M. kansasi (n = 2), M. abscessus (n = 1), M. avium (n = 3), M. chelonae (n = 2), and M. gordoneae (n = 1), as well as DNA samples from four bacteria that do not belong to the genus Mycobacterium (Nocardia nova, Corynebacterium amycolatum, Staphylococcus aureus, and Escherichia coli) were used as negative controls.

DNA samples of control strains M. tuberculosis H37Rv ATCC 27294, M. bovis CIP 102426, and M. bovis BCG CIP 105226 were provided by the Laboratory of Bacteriology of Strasbourg, Strasbourg, France. DNA extraction was performed from solid cultures by use of a MagNA Pure LC DNA III isolation kit (Roche Diagnostics, Indianapolis, IN), according to the manufacturer’s protocol. Samples of clinical isolates, which were prepared by simple thermolysis, were provided by the Laboratories of Bacteriology of Strasbourg, France, and Toulouse, France, and the Centre National de Référence des Mycobactéries in Paris, France. A loop of culture (i.e., bacteria grown on Lowenstein-Jensen or Coletos medium) was suspended in 300 to 500 μl water, and the mixture was heated at 95°C for 15 to 20 min. Mycobacterial cells were disrupted by sonication for 15 min or heat shock treatment (1 min at 95°C, followed by 1 min on ice, repeated five times). The supernatants obtained at full speed for 5 min contained the mycobacterial extracts used for the PCR amplification.

All isolates were previously identified to the species level by phenotypic and biochemical characterization methods after culture (43) or a gene probe assay, according to the manufacturer’s protocol (GenoType MTBC; Hain Lifescience GmbH). Ten samples, including seven clinical samples of M. tuberculosis, one clinical sample of M. africanum, M. tuberculosis H37Rv, and M. bovis control strains, were further genotyped by 24-locus MIRU-VNTR typing (52). The person who performed the MIRU-VNTR typing was blinded to the species identification and to the results of SNP typing.

**SNP selection.** A set of 16 well-characterized SNPs was chosen from the literature (1, 18, 19, 24, 39, 38, 51). The following eight SNPs were selected for use for the differentiation of MTBC species: hsp65<sup>240</sup> (C → T in M. canetti), gyrB<sup>675</sup> (numbers in parentheses refer to the SNP position on the genome) (C → T in M. microti), gyrB<sup>756</sup> (G → A in M. caprae and M. bovis), gyrB<sup>1410</sup> (C → T in M. bovis and M. bovis BCG DNA), dnaK<sup>203</sup> (ACT → ACC for PGG 1a and PGG 1b differentiation), katG<sup>463</sup> (CTG → CGG for PGG 2), and gyrA<sup>69</sup> (ACC → AGC for PGG 3). For genotyping of the M. tuberculosis isolates, the minimal SNP set that has recently been proposed for use by Alland and coworkers was chosen (the bases at the following SNP positions in H37Rv: 1977, 3532929 [instead of 54394], 74092, 105139, 264862, 3444390, 232574, 311631, and 913274) (1). These SNPs define the seven prominent SCGs that were first described by Filliol et al. (18). The sequence variation and lineage defined by them can be obtained from previously published reports (1, 18). It is important to note that the SNP at position 2154724, which also belongs to the minimal set, corresponds to katG<sup>463</sup>, which was already selected for use in the first SNP set. In our study, the first eight SNPs were qualified as “species specific,” whereas the last eight were “lineage specific” or “genotype specific.”

**Expanded validation of PCR and minsequencing primers.** Primer design was performed according to the recommendations of Sanchez et al. (45) so that the species-specific and lineage-specific SNPs were typed in two different multiplex assays: multiplex PCR 1 (mPCR 1) followed by SBE 1 and mPCR 2 followed by SBE 2, respectively. The sizes of the amplicons ranged from 28 to 76 nucleotides (nt) with SBE 1 and from 31 to 73 nt with SBE 2 and that differed in length from each other by 6 nt to allow sufficient separation by capillary electrophoresis. The sequences of the PCR and minsequencing primers used in this study are shown in Tables 1 and 2, respectively. All primers were tested in singleplex reactions with 1 ng template DNA under the conditions outlined by Sanchez et al. (45). Since the M. tuberculosis genome has a high GC content, 1 M betaine (Sigma-Aldrich, St. Louis, MO) was added to the singleplex PCR mixtures in order to reduce the formation of secondary structures in GC-rich regions (28). The amplification products were then verified by classical sequencing.

**Multiplex PCR amplifications and purification of PCR products.** Each mPCR was performed with 1 ng template DNA or 1 to 10 μl mycobacterial extract for clinical samples in a 50-μl final volume composed of 1× PCR Gold buffer (Applied Biosystems [AB], Foster City, CA), 8 mM MgCl<sub>2</sub>, 400 μM each deoxynucleoside triphosphate (dNTP), 0.2 to 0.5 μM each primer, and 2 U of AmpliTaq Gold DNA polymerase (AB). The concentrations of the PCR primers used in each reaction mixture are specified in Table 1. The thermal cycling consisted of a first denaturation step at 94°C for 5 min, followed by 33 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 30 s, and extension at 65°C for 30 s and with a final extension at 65°C for 7 min. Excess PCR primers and dNTPs were removed by using a NucleoSpin Extract II kit (Macherey-Nagel, Düren, Germany).
guishable from other species. Indeed, for all MTBC samples tested, the allele at the hsp65 locus was always C except in M. canetti (in which it was T), the allele at 16S rRNA locus was always T except in M. pinnipedii (in which it was C), and gyrB gene SNP position 675 was always C except in M. microti (in which it was T). Finally, three M. africanum samples were differentiated from the other species because the ancestral allele was detected for all SNPs, which also enabled us to determine that they belonged to PGG1a and thus probably belonged to M. africanum subtype Ib, as referred to by Huard and colleagues (29). The fourth M. africanum sample tested was assigned to PGG1b since it had a C allele at the katG SNP locus and thus probably belonged to M. africanum subtype Ib. However, the alleles detected from this M. africanum isolate and all PGG1b M. tuberculosis isolates tested were identical for all SNP loci analyzed in this study, showing that this assay does not allow discrimination between M. africanum subtype Ib and M. tuberculosis PGG1b. All negative controls always failed to amplify all loci except for, sometimes, hsp65 locus and/or 16S rRNA locus. Examples of electropherograms obtained from each MTBC species tested are shown in Fig. 1.

Multiplex reactions for genotyping of M. tuberculosis isolates

(SGC identification). The laboratory strain M. tuberculosis H37Rv, the M. bovis control strain, all clinical isolates of M. tuberculosis (n = 35), and one clinical M. africanum strain were further successfully characterized into one of the seven SCGs by means of the second multiplex assay: SCG1 (n = 2), SCG2 (n = 3), SCG3a (n = 2), SCG3b (n = 9), SCG5 (n = 14), SCG6a (n = 5), SCG6b (M. tuberculosis H37Rv), and SCG7 (M. bovis). All samples assigned to SCGs 1, 2 and 3a were found to belong to PGG1b, those assigned to SCGs 3b and 5 belonged to PGG2, and finally, those assigned to SCGs 6a and 6b belonged to PGG3, which is consistent with the results reported in the literature (18). Genotype data for the eight lineage-specific SNPs and the MIRU-VNTR typing results obtained for seven clinical M. tuberculosis samples, one clinical sample of M. africanum, M. tuberculosis H37Rv, and the M. bovis control samples are shown in Table 4. The SCGs inferred...
The work presented here describes the successful development of two simple, rapid, and specific assays for the differentiation of MTBC species, except M. africanum and M. canetti. The first assay, based on SNP detection by a multiplex minisequencing reaction, was found to be a reliable and reproducible method for SNP typing. Indeed, the SNP genotyping results were always confirmed by at least two independent PCRs for each mycobacterial sample, and the results were always in complete accordance with the results obtained by previous reference identification tests.

TABLE 2. Minisequencing primers for SBE 1 and SBE 2 used in this study

<table>
<thead>
<tr>
<th>Marker</th>
<th>allele specific sequence (5'-3')</th>
<th>Neutral sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBE 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBE 2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 3. Genotypes obtained for the eight species-specific SNPs typed for the 56 MTBC strains

<table>
<thead>
<tr>
<th>MTBC species (n = 56)</th>
<th>Nucleotide at SNP:</th>
<th>PGGa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>hsp65^563</td>
<td>katG^463</td>
</tr>
<tr>
<td><em>M. canetti</em> (n = 1)</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td><em>M. tuberculosis</em> (n = 7) and <em>M. africana</em> (n = 1)</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td><em>M. tuberculosis</em> (n = 23)</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td><em>M. africanum</em> (n = 3)</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td><em>M. pinnipedii</em> (n = 1)</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td><em>M. microti</em> (n = 1)</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td><em>M. caprae</em> (n = 1)</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td><em>M. bovis</em> (n = 7) and <em>M. bovis BCG</em> (n = 5)</td>
<td>C</td>
<td>T</td>
</tr>
</tbody>
</table>

* Inferred from the genotyping results for katG^203, gyrA^95, and katG^203 and results published previously (19, 51).

in a portable format, and alleles can be automatically called by the *ad hoc* software.

Besides the differentiation of MTBC species, the first assay presented here allows the simultaneous classification of MTBC isolates into three distinct genotypic groups, termed PGGs (PGGs 1a/b, 2 and 3), which is not achieved by previously developed MTBC species identification methods. All MTBC strains can be assigned to one of the three PGGs on the basis of the comparative analysis of three SNPs (*katG*^203, *katG*^463, and *gyrA*^95*) (19, 29, 51). To date, *M. tuberculosis* strains have been found to segregate to PGG1b, PGG2, and PGG3, while other MTBC members are restricted to PGG1a and PGG1b, which was also illustrated by our results (29). In the evolutionary pathway proposed by Sreevatsan et al., *M. tuberculosis* PGG1 strains are ancestral to PGG2 and PGG3 strains (51). Moreover, those authors observed through their study that PGG3 organisms are rather associated with sporadic TB cases, while PGG1 and PGG2 organisms are associated with clustered TB cases, suggesting differences in transmissibility or virulence between these genotypic groups (51). Thus, the determination of these major genetic groups, which is currently widely used, can be useful for both evolutionary and epidemiological applications.

In recent years, several additional comparative genomic studies of human-adapted members of the MTBC have identified other phylogenetically informative polymorphisms that define discrete strain lineages (1, 4, 10, 18, 20, 26). Because of the highly clonal population structure of the MTBC, the phylogenetic groupings recognized in the different studies are relatively concordant, even though they are based on the analysis of different markers (21). Although molecular epidemiological informative markers like MIRU-VNTR loci have been proposed for use in phylogenetic and population genetic analyses of MTBC, a recent comparative study conducted by Comas et al. revealed that SNPs and large sequence polymorphisms (LSPs) are the most robust and appropriate phylogenetically informative markers (14). The classification of MTBC strains into these distinct lineages is relevant for evolutionary purposes but also for TB control. Indeed, these discrete lineages are associated with particular human populations or geographical regions and show differences in virulence and disease outcomes (12, 20, 21, 35, 41, 53).

The second assay presented here allows the further accurate and unambiguous classification of *M. tuberculosis* isolates into six more resolved genotypic groups called SCGs. Because these SCGs are congruent with the lineages defined on the basis of other phylogenetically informative markers (21), interlaboratory comparisons can be done even if different markers are studied. This second assay can be run after the first assay, if the PGG determination is not sufficient, or it can be simultaneously performed.

The main advantage of the two-step strategy presented here is that it allows the reliable simultaneous identification of MTBC species and the classification of *M. tuberculosis* lineages into distinct lineages. It is easy to perform, gives unambiguous results, and can be easily introduced in a bacteriology laboratory with an automated sequencer, a detection platform that is also needed for other common applications like DNA sequencing and MIRU-VNTR typing.

Comas et al. recently suggested a combined SNP typing–MIRU-VNTR typing scheme to generate more accurate data for epidemiological and evolutionary applications (14). Indeed, on the one hand, SNPs are robust phylogenetic markers but offer insufficient discriminatory power for routine molecular epidemiological identifications, and on the other hand, the standard 24 MIRU-VNTR loci are powerful epidemiological markers but are unable to detect all strain lineages and their discriminatory powers vary by strain lineage. The new approach proposed by Comas et al. (14) consists of the identification of the main strain lineages by SNP typing, followed by further molecular epidemiological discrimination by the use of lineage-specific sets of the most discriminatory MIRU-VNTR markers for the lineage of interest. The SNAPSHOT mini-sequencing assay used in the present study seems to be an ideal method for SNP typing for such an approach because it uses the same equipment as MIRU-VNTR typing (i.e., a thermal cycler and a genetic analyzer). Furthermore, our study shows that the reliable identification of MTBC species can also be achieved by the SNAPSHOT minisequencing assay. This suggests that the identification of MTBC species could easily be added as a first step to the scheme proposed by Comas et al. (14).

It is also important to note that this two-step strategy was initially developed for paleomicrobiological analyses, i.e., investigation of ancient human tissues showing lesions suggestive of TB that are suspected to contain low levels of degraded mycobacterial DNA. That is why PCR primers were carefully designed in order to obtain amplicons of less than 150 bp. This feature could be of special interest in a clinical context when laboratories must deal with samples containing small amounts of degraded DNA (e.g., fixed tissues), since short DNA frag-
FIG. 1. Electropherograms obtained from the seven MTBC species by the multiplex assay for MTBC species identification. Each plot was obtained with GeneMapper ID (version 3.2.1) software (AB) and shows the relative fluorescence units (RFUs) versus the measured size (in nucleotides) of the SBE products relative to the GeneScan-120 LIZ internal size standard (AB). Mutated alleles are indicated by arrows.
ments are amplified more efficiently than long fragments in such cases. We are currently testing the use of this two-step strategy with ancient human remains. However, in recent studies describing the development and validation of a similar assay for the typing of human SNPs designed for the screening of ancient human remains, we already reported on the sensitivity and robustness of the SNaPshot minisequencing approach for SNP typing (6, 7). These advantages of this SNP typing method were also reported by other authors, who successfully used it to evaluate ancient human samples or archival clinical samples (6, 7, 11, 16, 22, 31, 46).

To conclude, the present study describes the successful development of reliable SNaPshot minisequencing-based SNP assays for the identification of MTBC species, determination of the principal genetic groups (PGGs 1a/b, 2, and 3), and the further classification of *M. tuberculosis* isolates into more resolved phylogenetic groups called SCGs. SNaPshot minisequencing for SNP genotyping is currently used for various applications; here, we propose the use of this technique as a complement to current methods for the clinical diagnosis of TB and epidemiological investigations.

**ACKNOWLEDGMENTS**

We thank Rosine Bauriaud, Cathie Barthel, and Marie Durand for preparation of the DNA from the MTBC strains and Daniel Montagnon and Gerald Millot for technical assistance.

**REFERENCES**


