

High Genetic Diversity Revealed by Variable-Number Tandem Repeat Genotyping and Analysis of *hsp65* Gene Polymorphism in a Large Collection of “*Mycobacterium canettii*” Strains Indicates that the *M. tuberculosis* Complex Is a Recently Emerged Clone of “*M. canettii*”

Michel Fabre,¹ Jean-Louis Koeck,² Philippe Le Flèche,³ Fabrice Simon,^{4†} Vincent Hervé,¹ Gilles Vergnaud,³ and Christine Pourcel^{5,6*}

Laboratoire de Mycobactériologie, HIA Percy, 92140 Clamart,¹ HIA Val de Grâce, 75005 Paris,² Centre d'Etude du Bouchet, 91710 Vert le Petit,³ Génome, Polymorphisme et Minisatellites (GPMS), Institut de Génétique et Microbiologie, Université Paris XI, 91405 Orsay Cedex,⁵ and Institut Pasteur, 75015 Paris,⁶ France, and CHA Bouffard, Djibouti, Republic of Djibouti⁴

Received 2 February 2004/Returned for modification 15 March 2004/Accepted 6 April 2004

We have analyzed, using complementary molecular methods, the diversity of 43 strains of “*Mycobacterium canettii*” originating from the Republic of Djibouti, on the Horn of Africa, from 1998 to 2003. Genotyping by multiple-locus variable-number tandem repeat analysis shows that all the strains belong to a single but very distant group when compared to strains of the *Mycobacterium tuberculosis* complex (MTBC). Thirty-one strains cluster into one large group with little variability and five strains form another group, whereas the other seven are more diverged. In total, 14 genotypes are observed. The DR locus analysis reveals additional variability, some strains being devoid of a direct repeat locus and others having unique spacers. The *hsp65* gene polymorphism was investigated by restriction enzyme analysis and sequencing of PCR amplicons. Four new single nucleotide polymorphisms were discovered. One strain was characterized by three nucleotide changes in 441 bp, creating new restriction enzyme polymorphisms. As no sequence variability was found for *hsp65* in the whole MTBC, and as a single point mutation separates *M. tuberculosis* from the closest “*M. canettii*” strains, this diversity within “*M. canettii*” subspecies strongly suggests that it is the most probable source species of the MTBC rather than just another branch of the MTBC.

The *Mycobacterium tuberculosis* complex (MTBC) includes *M. tuberculosis*, *M. africanum*, *M. bovis*, and *M. microti*. Using molecular data, van Soolingen et al. recently proposed to consider “*M. tuberculosis* subsp. *canettii*” as a new subspecies of the MTBC, although its colonies are eugonic and have a smooth appearance in contrast to the other bacteria of the MTBC (25).

The first such strain was isolated by Georges Canetti in 1969. About 20 years later, two additional strains were isolated from patients with tuberculosis (TB) lymphadenitis (16, 25). In 2002 Miltgen et al. described the characteristics of two new strains of “*M. canettii*” isolated from two French patients with pulmonary TB (14). Considering the very small number of “*M. canettii*” infections reported to date, no difference with the classical *M. tuberculosis* infection can be identified. Nonetheless, some authors suggest that “*M. canettii*” could induce milder pneumonia than classical TB (12, 14).

Studies by Brosch et al. and Marmiesse et al. of 20 regions where insertion-deletion events took place in the genome of *M. tuberculosis* suggested that “*M. canettii*” diverged first from the rest of the MTBC (1, 13). The “*M. canettii*” taxon can also

be easily differentiated from the other members of the MTBC on the basis of a restriction site polymorphism in the *hsp65* gene (8).

All the described “*M. canettii*” strains show only two bands by spoligotyping (25), corresponding to spacers 30 and 36, which indicates that the direct repeat (DR) locus in “*M. canettii*” is very different from that of the other members of the MTBC. By sequencing, van Embden et al. showed the existence of at least 26 spacers that appear to be unique to “*M. canettii*” (24).

Frothingham et al. (7) and Supply et al. (21) were the first to describe the use of polymorphic tandem repeat loci for genotyping MTBC strains. Variable-number tandem repeat (VNTR) analysis is a very promising approach, since it successfully discriminates low-copy-number IS6110 strains (2), which are poorly resolved by the otherwise quite efficient IS6110 typing (23). In addition, tandem repeat typing can be standardized, and common databases are easily set up (3, 10). However, typing by multiple methods is still required to attain maximum specificity and more importantly to correlate multiple-locus VNTR analysis (MLVA)-derived data sets with data based on previous methods. One reason for this is that mutation rates and evolution of tandem repeat loci in bacteria are not yet fully understood, so that MLVA data interpretation is still to some extent a research area.

Based upon the available genomic sequences for *M. tuber-*

* Corresponding author. Present address: GPMS, Institut de Génétique et Microbiologie, Bat. 400, Université Paris XI, 91405 Orsay Cedex, France. Phone: (33) 1 69 15 30 01. Fax: (33) 1 69 15 66 78. E-mail: Christine.Pourcel@igmors.u-psud.fr.

† Present address: Service des maladies infectieuses et tropicales, HIA Laveran, 13013 Marseille, France.

culosis, we previously selected and analyzed a larger collection of VNTRs, giving rise to a high-resolution MLVA typing assay (10). Using a selection of 21 VNTRs we have typed 43 smooth variants of *M. tuberculosis* and compared their genotype with one of the previously described "*M. canettii*" strains and with the collection of MTBC strains described by Le Flèche et al. (10).

MATERIALS AND METHODS

Strains. The 43 smooth strains investigated in this study were all niacin negative, nitrate reductase positive, urease positive, and thiophen-2-carboxylic acid hydrazide positive. They were isolated from sputum, gastric fluid, lymph node, or ascites fluid.

Most patients were Djiboutian, and the others were expatriated patients living in Djibouti. No epidemiological link could be established between the patients. The detailed description of the clinical features of the cases will be published elsewhere (J.-L. Koeck et al., unpublished data). DNA from "*M. canettii*" strain CIPT140010060 (referred to in this work as CIPT060) was kindly provided by Véronique Vincent. percy8 (see Fig. 1) is the same as strain 990161 in reference 13. The reference strain *M. tuberculosis* H37Rv was used as a control in the MLVA (10).

VNTR analysis. PCR amplification of 21 VNTR loci and electrophoresis of products on agarose gels were carried out as described in reference 10. MIRU26 (22) was used instead of Mtub38, which is difficult to type because of the presence of repeats of different unit length at this locus (10).

Data management and analyses. Gel images were analyzed using a bionumerics software package (version 3.5; Applied-Maths, Sint-Martens-Latem, Belgium) as previously described (10). The number of repeats in each allele was deduced from the amplicon size. The resulting data were analyzed with bionumerics as a character data set. Clustering analysis was done using the categorical parameter and the unweighted pair group method with arithmetic averages coefficient. The minimum spanning tree (5) was constructed with the following options: (i) in case of equivalent solutions in terms of calculated distances, the selected tree was the one containing the highest number of links between genotypes differing at only one locus ("Highest number of single locus variants" option); (ii) the creation of hypothetical types (missing links) reducing the total length of the tree was allowed.

Analysis of the DR locus by PCR amplification and characterization of new spacers. PCR amplification was performed using primers selected in the spacers described by van Embden et al. (24): primer pair Mcan80For (5'-TAGCGA GCTGTGCGGCAGTA) and Mcan83Rev (5'-TAAGCACACCAGCACCT CCC) and primer pair Mcan73For (5'-TCGGTGCTGACCCCATGGAT) and Mcan76rev (5'-GATTCGCCCGTCGTGCAAT). These amplifications are predicted to produce, respectively, a 202- and a 213-bp fragment in the sequenced "*M. canettii*" strain (see Fig. 3A).

Primers DRa (5'-GGTTTTGGGTCTGACGAC) and DRb (5'-CCGAGAG GGGACGGAAC) were used to amplify the DR region as formerly described in the spoligotyping method (9). The PCR product was purified and size selected by electrophoresis on a 1% agarose gel. Cloning was performed using the pGEM-T Easy Vector system (Promega, Charbonnières, France). Inserts were analyzed by PCR and sequenced using the SP6 and T7 primers. EZ Load 100-bp ruler (Bio-Rad) was used as a DNA size marker. It contains 10 fragments from 100 bp to 1 kb.

hsp65 polymorphism and sequencing of PCR products. PCR amplification of part of the *hsp65* gene was performed as described previously (8) using primers Tb11 (5'-ACCAACGATGGTGTGCCAT) and Tb12 (5'-CTTGTCGAACCG CATACCCT). The 441-bp fragment was digested by HhaI or DdeI (Roche, Meylan, France) and run on a 3% Nusieve 3:1 agarose gel (FMC Bioproducts, TEBU, Le Perray en Yvelines, France).

For sequencing, a PCR with 45- μ l mixture was performed and the product was purified by polyethylene glycol precipitation as described previously (4). Sequencing was done by MWG Biotech (Courtabouef, France).

RESULTS

VNTR analysis. The strains were genotyped by PCR using 21 markers, ETR-A, -B, -C, -D, and -E (7); MIRU 02, 10, 16, 23, 26, 27, 39, and 40 (22); Mtub01, Mtub02, Mtub12, Mtub21, Mtub29, Mtub30, and Mtub39 (10); and QUB11a (18). The

allele sizes were converted to repeat unit numbers using the conventions defined elsewhere (10, 19) with one exception. Mtub39, which is coded 5 in reference strain H37Rv according to reference 10 is coded 6 in the present report to avoid the ambiguous code 0 in percy79 (consequently, the Mtub39 values indicated in reference 10 should be increased by 1 when comparing the two data sets). Such new "zero repeat unit" alleles (which are not PCR amplification failures) are observed (Table 1) at loci MIRU02 (genotype 4, strain percy79) and MIRU27 (genotype 14, strain percy65). In a dendrogram produced with the MLVA data and some previously obtained *M. tuberculosis* and *M. bovis* data, all the smooth strains, including the previously characterized "*M. canettii*" strain CIPT060, cluster into a very distant and separate group (not shown). When a dendrogram is produced with the smooth strains and the H37Rv control alone (Fig. 1), one group (group A) comprising 31 strains has a quite-homogenous genotype (genotypes 1 to 3; one or two differences), whereas the remaining 13 strains (including the previously described "*M. canettii*" strain CIPT060) display more diversity. Among the latter, five strains are clustered into a second group (group B). The diversity is shown for example with markers ETR-C and Mtub30, which possess only one allele in group A strains and, respectively, six and four different alleles in the remaining strains (Table 1).

As a complement, a minimum spanning tree analysis was performed (Fig. 2). This kind of analysis is applicable to categorical data sets (5). The resulting tree minimizes the summed distance of all branches of the tree (the distance between two strains being the number of markers by which they differ). The group A strains (genotypes 1 to 3, light blue, Fig. 2) are closely related and are grouped into a single complex. Genotype 2, which includes 28 isolates, is the center of this cluster. Group B strains (genotypes 5 to 7) are similarly clustered. The eight remaining genotypes are very loosely connected to each other, suggesting a very high diversity. The CIPT060 "*M. canettii*" strain is genotype 12.

DR typing. van Embden et al. sequenced the DR locus in one "*M. canettii*" strain, strain SO93 (25), and described the existence of 26 new spacers (spacers 69 to 94) which were absent from all the *M. tuberculosis* tested. We first tested for the presence of this particular DR by PCR and sequencing using primers flanking the locus and primers corresponding to spacers 80 and 83 (McanFor plus Mcan83Rev and Mcan80For plus McanRev) (data not shown). The result suggested that the locus was present only in a subset of "*M. canettii*" strains and that two strains, percy32 and CIPT060, had additional spacers. We further investigated the locus by choosing two pairs of primers in spacers 80 and 83 and in spacers 73 and 76 and performing a PCR on all the smooth strains (Fig. 3 shows the result for a subset of strains). PCR using primers Mcan80For and Mcan83Rev produced an amplicon of the expected size (202 bp) in all the group A strains and in the reference strain CIPT060 and strain percy32 (Fig. 3A). The same result was obtained with primers Mcan73For and Mcan76Rev (expected amplicon size, 213 bp) except for the two strains CIPT060 and percy32 (lanes 16 and 20). The larger amplicon (505 bp) obtained in these two strains was sequenced, leading to the identification of the same four (and previously unknown) spacers (Fig. 3B). In 11 strains this approach failed to detect any spacer (Fig. 1).

TABLE 1. Genotyping data obtained by MLVA in this study

Genotype	No. of units with marker																					
	ETR-A	ETR-B	ETR-C	ETR-D	ETR-E	MIRU-02	MIRU-10	MIRU-16	MIRU-23	MIRU-26	MIRU-27	MIRU-39	MIRU-40	Mtub01	Mtub02	Mtub12	Mtub21	Mtub29	Mtub30	Mtub39	Qub11a	
1	10	4	6	3	4	3	3	2	3	2	2	8	8	9	8	4	3	5	2	3	3	9
2	10	4	6	3	4	3	3	2	3	2	2	8	8	9	8	4	3	5	2	3	3	11
3	10	4	6	3	4	3	3	2	3	2	2	9	8	9	8	4	3	5	2	2	3	11
4	10	1	6	2	6	0	3	1	3	1	2	5	8	9	8	4	3	5	1	1	1	11
5	10	4	6	2	2	1	3	1	3	1	2	3	7	9	7	5	1	2	3	2	2	11
6	10	4	6	2	2	1	3	1	3	1	2	3	7	9	7	5	1	2	3	2	2	14
7	10	4	7	2	2	1	3	1	3	1	2	3	7	9	7	5	1	2	3	2	2	11
8	10	5	3	3	5	3	3	1	1	1	2	6	8	9	8	4	3	4	3	2	4	11
9	10	4	3	2	6	1	3.5	1	1	1	2	6	8	9	8	4	3	4	3	2	4	12
10	10	1.5	5	2	6	2	4	1	6	1	2	6	8	9	8	4	3	1	1	4	4	11
11	10	5	4	3	7	1	3	3	6	2	3	6	6	9	6	5	2	5	5	3	5	11
12 (CIPT060)	10	6	3	2	6	3	3	2	7	2	3	6	6	9	6	5	2	5	5	5	3	11
13	10	6	5	3	7	3	2	1	7	1	2	7	7	9	7	5	3	5	1	4	4	11
14	10	1	10	6	3	4	3.5	0	1	0	1.5	1	8	9	8	3	1	4	2	1	1	11
15 (H37Rv)	3	3	4	3.5	3	2	3	3	2	2	2	1	6	10	6	4	2	4	2	6	6	3

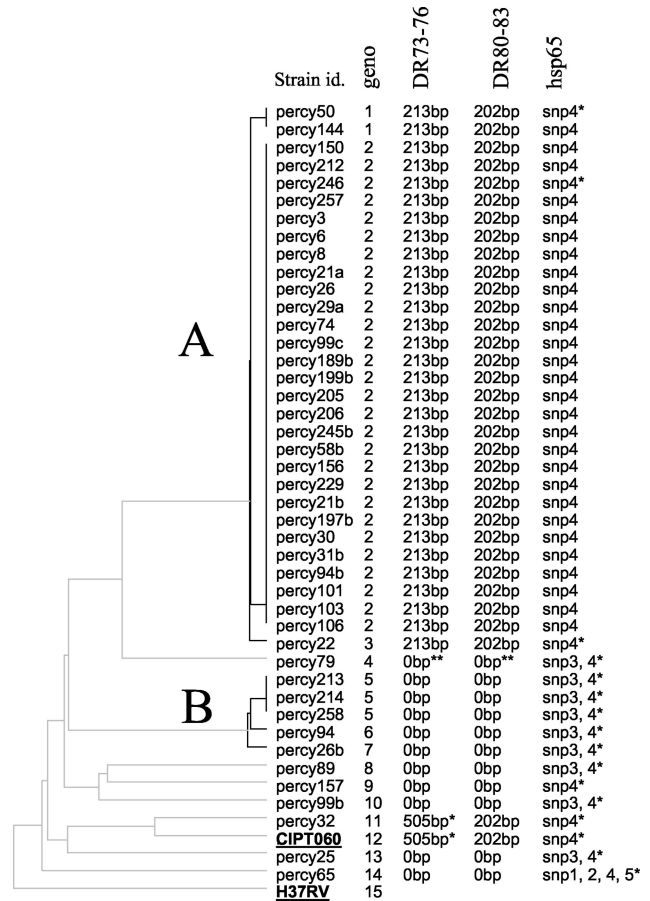


FIG. 1. MLVA cluster analysis. Clustering analysis was done using the categorical and unweighted pair group method with arithmetic averages options. From left to right, the columns designate the strains names, the genotype number (geno), the result obtained for the PCR amplification of part of the DR locus using “canettii-specific” spacers 73 to 76 or 80 to 83 (24), and the hsp65 profile. The asterisks indicate that sequencing data have been produced. The hsp65 fragment has been sequenced in at least one strain from each genotype (SNPs 2 and 3 [Fig. 5] are not detected by PCR-restriction fragment length polymorphism analysis). Strains H37Rv and CIPT060 are underlined.

To check whether a DR locus with unknown spacers might be present in some of the strains, we performed a PCR amplification with primers DRa and DRb localized in the constant region of the DR. No amplification was observed in 10 strains, confirming the complete absence of DR structure in these strains (Fig. 1). In contrast an amplification product in the form of a ladder was obtained with strain percy79 (Fig. 4A). We purified amplicons of a size ranging between 100 and 600 bp and cloned them into a plasmid vector (Fig. 4B and C). Sequencing of the insert from eight clones led to the identification of 20 spacers, all of them previously unknown.

hsp65 PRA. When analyzing the polymorphism of the hsp65 gene by PCR restriction analysis (PRA) using HhaI, we found that 43 of the 44 smooth strains studied (including CIPT060) had the same specific “*M. canettii*” profile formerly described by Goh et al. (8) and different from that of the MTBC strains. Figure 5A shows the constant pattern observed in representative strains from the MTBC. In H37Rv, *M. africanum* type I, *M.*

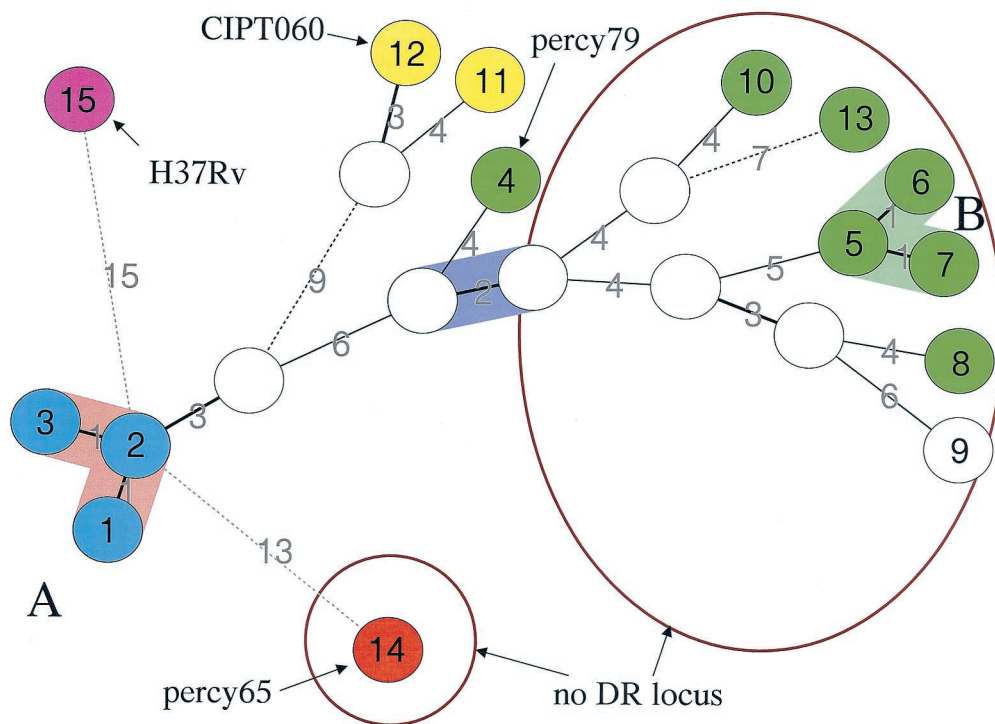


FIG. 2. Minimum spanning tree analysis and comparison with the other approaches. A minimum spanning tree was constructed using the genotyping data provided in Table 1. The parameters used for the construction of the tree authorized the introduction of hypothetical (missing) links which appear as open circles (data predicted by the software as intermediates, not present in the available collection). The genotype numbers listed in Fig. 1 are indicated in the circles. The three group A genotypes comprising 31 strains are shown in blue. The two closely related strains, including the "M. canettii" strain CIPT060 (genotype 12), and in which four additional spacers in the DR locus have been found are shown in yellow. Genotypes with the *hsp65* SNP 3 variant are presented in green. The exceptional strain percy65 (genotype 14; *hsp65* SNPs 1, 2, and 5) is colored in red. The two groups of strains devoid of DR regions are circled. Strain percy79 (genotype 4, absence of the "M. canettii"-specific set of spacers, presence of new spacers) is indicated. Dotted lines are used to link genotypes separated by seven or more differences in the MLVA genotype.

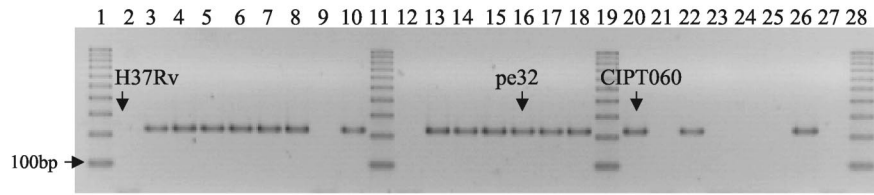
africanum type II, *M. bovis*, and *M. microti*, the characteristic restriction fragments are 186, 103, 72, and 63 bp (and a non-visible 17-bp fragment) (Fig. 5A, lanes 2 to 6, and B, lane 11). In all but one "M. canettii" strain, the 186- and 72-bp fragments are absent, replaced by a 258-bp fragment because one HhaI site is lacking (Fig. 5A, lane 7, and B). In strain percy65 the 258-bp fragment is absent and presumably replaced by a 235-bp fragment (Fig. 5B, lane 3). In order to further analyze the *hsp65* polymorphism we sequenced the PCR product of 16 "M. canettii" strains representing all 14 different genotypes and compared the sequences to that of H37Rv (Fig. 5D; a representative set of sequences is shown). All the strains showed the expected C-to-T nucleotide change compared to H37Rv, leading to the absence of an HhaI site (Fig. 5D, single nucleotide polymorphism [SNP] 4). A second C-to-T transition was observed in eight strains, percy26b, percy258, percy94, percy89, percy214, percy79, percy25, and percy99b (SNP 3, Fig. 1; the corresponding genotypes are shown in green in Fig. 2). In strain percy65, three other nucleotide differences were observed. An A-to-G change created a new HhaI site (SNP 1), explaining the appearance of the 235-bp fragment upon HhaI digestion (Fig. 5B). A C-to-T transition created a DdeI site (SNP 5). The existence of this predicted DdeI site was checked by digestion of the PCR amplicons. Two fragments of the expected size (330 and 111 bp) were indeed observed in strain percy65 (Fig. 5C, lane 3), whereas neither the representative

MTBC strains (Fig. 5A, lanes 9 to 13) nor the other "M. canettii" strains possessed a DdeI site. The last change in strain percy65 was a G-to-A change (SNP 2).

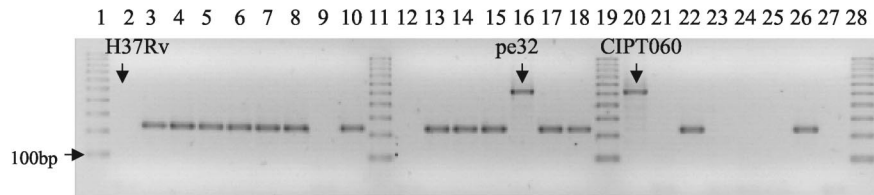
DISCUSSION

In the last few years, an exceptional collection of 43 smooth strain variants of *M. tuberculosis*, originating from the Republic of Djibouti, has been constituted. The very rare instances of such strains reported before have been estimated to be sufficiently different from the other MTBC species and sufficiently similar to one another to be given a specific name, "M. canettii." We have investigated this extended collection together with one previously identified "M. canettii" strain, using three complementary molecular tools, MLVA, DR locus investigations, and *hsp65* partial gene sequencing (and/or SNP typing). The results obtained clearly demonstrate that this strain collection represents a genetically homogeneous group, quite distinct from the MTBC. The observed relative homogeneity indicates that all these strains should indeed be considered as "M. canettii" strains. However, and very interestingly, this overall homogeneity does cover a diversity of genotypes and sequences which appears to be much larger than those observed across the whole MTBC. This is all the more remarkable in view of the very restricted geographic area from which all

A primer pair 80For and 83Rev



B primer pair 73For and 76Rev



New spacers
 CCGCGTCATCGGGCAGACCAGCAACAGGTCGTGGAAG
 ATACCGGAAGCACATTTCGCGCCAGTGGTCGTCGACAGAAG
 ATTCCGACTAGCGCGTCGTCGTTTCCAGCCTCAAT
 CGACAAGTTTTCGCGGGATCAAGTCCGCGCCGGTCAA

FIG. 3. Structure of the DR locus. PCR using primers Mcan80For and Mcan83Rev (A) or Mcan73For and Mcan76Rev (B) on 24 smooth strains and H37Rv. The 100-bp ruler is used as size marker (lanes 1, 11, 19, and 28). No amplicon is observed in H37Rv (lane 2) and in seven smooth strains (lanes 9, 12, 21, 23, 24, 25, and 27). (B) Strains percy32 and CIPT060 have four additional new spacers as shown by the presence of a 505-bp amplicon and by sequencing.

strains investigated originate, the Republic of Djibouti, on the Horn of Africa. We review here the present findings.

The 43 smooth variants of *M. tuberculosis* have characteristics in common with the representative “*M. canettii*” strain

analyzed, such as a specific polymorphism of the *hsp65* gene, but can be separated into several groups on the basis of the other molecular tests: (i) clustering by MLVA genotyping (Table 1; Fig. 1 and 2), (ii) the presence of a DR region in one

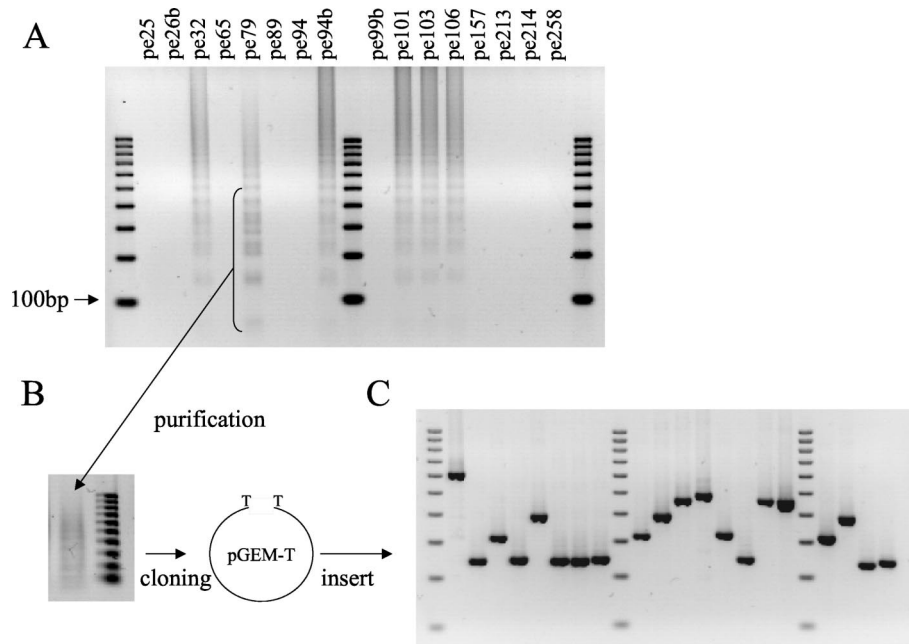


FIG. 4. Cloning of new spacers from strain percy79. (A) PCR amplification was performed using primers DRa and DRb on 16 “*M. canettii*” strains. Amplification product from strain percy79 was purified and cloned into the pGEM-T Easy vector (B), and inserts were analyzed by PCR amplification (C). Lanes 1, 10, and 19 in panels A and C contain 100-bp ruler DNA size markers.

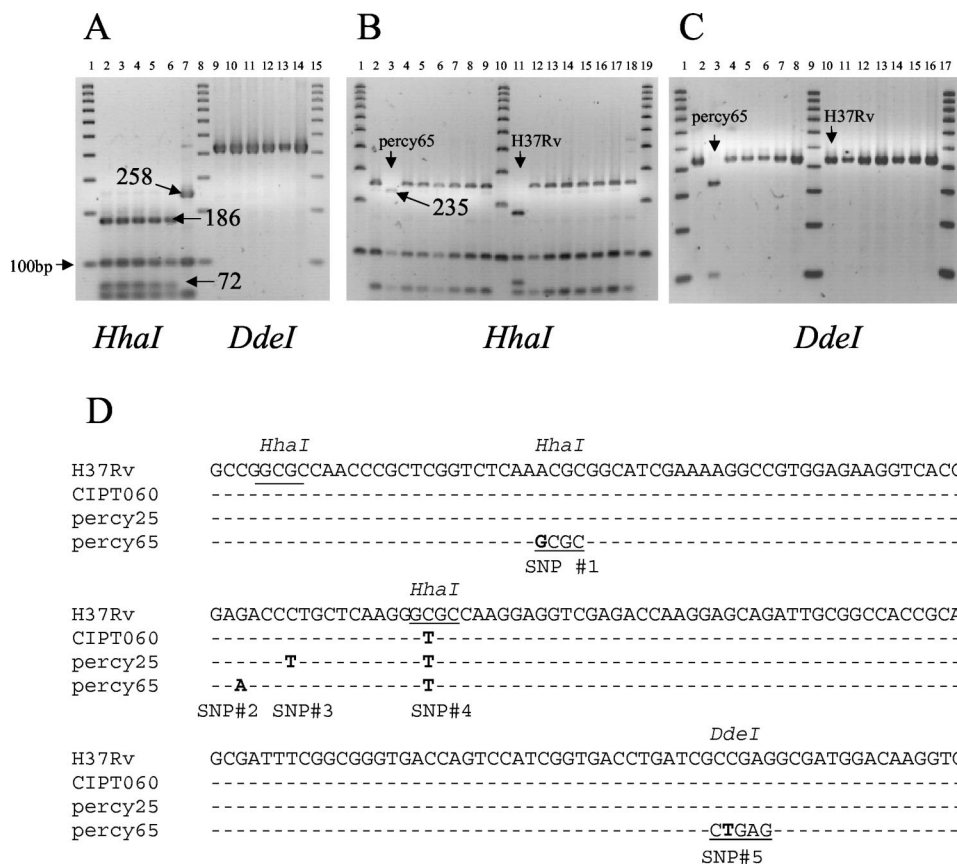


FIG. 5. PCR-restriction fragment length polymorphism analysis of the *hsp65* locus. (A) *hsp65* PRA patterns of representative strains from the MTBC after digestion with *HhaI* (left) or *DdeI* (right). The representative "*M. canettii*" strain CIPT060 shows a different pattern with *HhaI* as expected (lane 7) but not with *DdeI* (lane 14). (B) Fifteen smooth strains and H37Rv upon digestion of the 441-bp amplicon with *HhaI*. (C) Same as panel B, using *DdeI* instead of *HhaI*. A 100-bp ruler is used as size marker. Strain percy65 has a unique digestion pattern as exemplified. (D) Multiple alignments using the CLUSTALW algorithm of the 180-bp portion of the *hsp65* gene in H37Rv and smooth strains showing five single nucleotide changes, numbered 1 to 5. The SNP described by Goh et al. (8) is SNP 4.

group containing a collection of unique spacers (Fig. 3 and 4), (iii) the absence of DR in the second group (Fig. 2 to 4), and (iv) additional polymorphism in the *hsp65* gene (Fig. 5).

MLVA places the 44 strains together in a separate group compared to almost 200 strains of the MTBC similarly analyzed (10; unpublished data). For several markers, some alleles are unique to "*M. canettii*," such as ETR-A (allele 10), ETR-C (alleles 6 and 10), MIRU-02 (allele 3), MIRU-40 (allele 8), and Mtub29 (allele 5). Two markers are monomorphic in this collection, ETR-A and Mtub01. In contrast, ETR-A was found to be highly polymorphic in the MTBC (19). Here a unique "*M. canettii*"-specific allele is observed.

The strains percy32 and CIPT060, which possess a DR locus with additional spacers compared to those described by van Embden et al. (24), constitute a branch more closely related to group A strains than to the rest. Ten strains lacking a DR locus are loosely connected (Fig. 1 and 2). Analysis of the PCR amplification product obtained with primers corresponding to the conserved repeat of the DR locus shows that strain percy79 possesses a not previously described set of spacers. Previous work (24) suggested that a common ancestor bearing a large number of units has evolved by the interstitial deletion of

motifs to explain the hundreds of different combinations seen in contemporary MTBC strains (6). The fact that strains which lack a DR locus have highly diverged genotypes (here, for instance, strain percy65, [Fig. 2] and the genotypes 5 to 10 and 13) suggests that the deletion occurred at least twice independently in the different strains. Further molecular work such as spoligotyping and sequencing across the "missing DR" might improve the understanding of the origin and history of the DR locus. Considering the genome microdeletions that occurred during the MTBC evolution and separated *M. bovis* from *M. tuberculosis*, Brosch et al. (1) concluded that *M. canettii* is the species closest to the common ancestor. This analysis was based on the investigation of five "*M. canettii*" strains, which were shown to be devoid of the set of deletions which differentiate the other members of the MTBC. The relatively very high variability of the MTBC in terms of microdeletion events is quite interesting and probably reflects a sudden change of environment and selection pressure (some parts of the genome then becoming dispensable) rather than true evolutionary distance.

Our investigation on a relatively large collection of smooth strains, recruited in a very limited geographic area (the Republic of Djibouti), demonstrates that the genetic heterogene-

ity (not taking into account deletions) within “*M. canettii*” is much larger than within the whole MTBC. This is suggested first by the MLVA data (Fig. 1 and 2). In spite of the suspicion that tandem repeat loci may be under variable and specific evolutionary pressures, the clustering of strains by MLVA does make sense (15, 17). This is illustrated by a growing number of studies, not only of *M. tuberculosis*, in which MLVA clustering data have been compared to previous knowledge (3, 11) and which validate this approach as a clinical epidemiologic tool (19). In the present investigation, percy32 and CIPT060, for instance, which were predicted to be closely related by MLVA (Fig. 2, yellow circles), turned out to contain the same very rare four additional spacers in their DR regions. Similarly, strains carrying *hsp65* SNP 3 are grouped together by the MLVA (Fig. 2). However, it is the *hsp65* polymorphism data which provide the strongest evidence for the comparatively high “*M. canettii*” diversity. We have analyzed a portion of 441 bp from the *hsp65* gene which was previously shown to harbor a single nucleotide difference specific for the few “*M. canettii*” strains previously assayed. This polymorphism is present in the 44 strains of this study, but in addition we describe four new nucleotide differences in 6 strains (synonymous substitutions), one of them being present in 5 strains. Interestingly strain percy65 possesses three unique SNPs. This important polymorphism in a small genomic region confirms that some strains are highly diverged. In contrast, no sequence difference at this locus is seen between the three sequenced genomes, *M. bovis* strain AF2122, *M. tuberculosis* strains H37Rv, and CDC1551. Not a single SNP was observed among either the 267 strains representing the MTBC diversity (and including *M. microti*, *M. africanum*, *M. bovis*) or the *M. bovis* BCG strains (20). In addition only one SNP separates any member of the MTBC investigated so far from the closest “*M. canettii*” sequence (Fig. 5D), whereas four different SNPs have been identified within the present collection of “*M. canettii*.” This implies that the “*M. canettii*” group is much older than the MTBC. In other words, the available data strongly suggest that the MTBC is a recently emerged subspecies of “*M. canettii*.” Furthermore, all the “*M. canettii*” strains investigated originate from the Republic of Djibouti, i.e., a very limited geographic area, and were collected over a short time period, so that the diversity detected probably underestimates the true diversity of the species.

This finding clearly raises a number of interesting questions. Smooth *M. tuberculosis* strains have not been detected elsewhere in spite of the fact that they show a very remarkable colonial morphology and spoligotype (25) and could not have been missed. It is tempting to speculate that the MTBC originates from East Africa. One successful clone spread, whereas the rest of the progenitor species, “*M. canettii*,” remained in East Africa. It may be relevant in this view to recall that the group of ancestral *M. tuberculosis* strains without deletions of TbD1 (13) is associated with the East African Indian type of strains (19).

Finally, a reservoir must exist. In spite of the lack of epidemiological link between the different patients, we show here that one clone (genotype 2) is responsible for a large proportion of cases. Field investigations would be of great interest in trying to find this reservoir.

ACKNOWLEDGMENTS

We thank Philippe Dubrous (Military Hospital, Bordeaux), Eric Garnotel (Military Hospital, Marseille), Michaël Ponsoda (Merieux Laboratory, Lyon), and Véronique Vincent (Institut Pasteur, Paris) for providing strains or DNA. We thank Sabrina Ivol for excellent technical help.

Work on the typing and molecular epidemiology of dangerous pathogens is supported by the French ministry of defense.

REFERENCES

- 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.
- 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.
- Denoeud, F., and G. Vergnaud. 2004. Identification of polymorphic tandem repeats by direct comparison of genome sequence from different bacterial strains: a Web-based resource. *BMC Bioinformatics* **5**:4.
- Embley, T. M. 1991. The linear PCR reaction: a simple and robust method for sequencing amplified rRNA genes. *Lett. Appl. Microbiol.* **13**:171–174.
- Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J. Bacteriol.* **186**:1518–1530.
- 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. Mammia, 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. Sufys, 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.
- 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.
- Goh, K. S., E. Legrand, C. Sola, and N. Rastogi. 2001. Rapid differentiation of “*Mycobacterium canettii*” from other *Mycobacterium tuberculosis* complex organisms by PCR-restriction analysis of the *hsp65* gene. *J. Clin. Microbiol.* **39**:3705–3708.
- 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.
- Le Flèche, P., M. Fabre, F. Denoeud, 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.
- Le Flèche, P., Y. Hauck, L. Onteniente, A. Prieur, F. Denoeud, V. Ramiisse, P. Sylvestre, G. Benson, F. Ramiisse, and G. Vergnaud. 2001. A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. *BMC Microbiol.* **1**:2.
- Lopez, B., D. Aguilar, H. Orozco, M. Burger, C. Espitia, V. Ritacco, L. Barrera, K. Kremer, R. Hernandez-Pando, K. Huygen, and D. van Soolingen. 2003. A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes. *Clin. Exp. Immunol.* **133**:30–37.
- Marmiesse, M., P. Brodin, C. Buchrieser, C. Gutierrez, N. Simoes, V. Vincent, P. Glaser, S. T. Cole, and R. Brosch. 2004. Macro-array and bioinformatic analyses reveal mycobacterial “core” genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiology* **150**:483–496.
- Miltgen, J., M. Morillon, J. L. Koeck, A. Varnerot, J. F. Briant, G. Nguyen, D. Verrot, D. Bonnet, and V. Vincent. 2002. Two cases of pulmonary tuberculosis caused by *Mycobacterium tuberculosis* subsp. *canettii*. *Emerg. Infect. Dis.* **8**:1350–1352.
- 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.
- Pfyffer, G. E., R. Auckenthaler, J. D. van Embden, and D. van Soolingen. 1998. *Mycobacterium canettii*, the smooth variant of *M. tuberculosis*, isolated from a Swiss patient exposed in Africa. *Emerg. Infect. Dis.* **4**:631–634.
- Pourcel, C., Y. Vidgop, F. Ramiisse, G. Vergnaud, and C. Tram. 2003. Char-

- acterization of a Tandem Repeat Polymorphism in *Legionella pneumophila* and Its Use for Genotyping. *J. Clin. Microbiol.* **41**:1819–1826.
18. **Skuce, R. A., T. P. McCorry, J. F. McCarroll, 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.
 19. **Sola, C., I. Filliol, E. Legrand, S. Lesjean, C. Locht, P. Supply, and N. Rastogi.** 2003. Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. *Infect. Genet. Evol.* **3**:125–133.
 20. **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.
 21. **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.
 22. **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.
 23. **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.
 24. **van Embden, J. D., T. van Gorkom, K. Kremer, R. Jansen, B. A. van Der Zeijst, and L. M. Schouls.** 2000. Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. *J. Bacteriol.* **182**:2393–2401.
 25. **van Soolingen, D., T. Hoogenboezem, P. E. de Haas, P. W. Hermans, M. A. Koedam, K. S. Teppema, P. J. Brennan, G. S. Besra, F. Portaels, J. Top, L. M. Schouls, and J. D. van Embden.** 1997. A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. *Int. J. Syst. Bacteriol.* **47**:1236–1245.