

Novel Genetic Polymorphisms That Further Delineate the Phylogeny of the *Mycobacterium tuberculosis* Complex†

Richard C. Huard,^{1‡} Michel Fabre,² Petra de Haas,³ Luiz Claudio Oliveira Lazzarini,¹
Dick van Soolingen,³ Debby Cousins,⁴ and John L. Ho^{1*}

Division of International Medicine and Infectious Diseases, Department of Medicine, Joan and Sanford I. Weill Medical College, Cornell University, New York, New York¹; Laboratoire de Mycobactériologie, HIA Percy, 92140 Clamart, France²; National Institute of Public Health and the Environment, Bilthoven, The Netherlands³; and Australian Reference Laboratory for Bovine Tuberculosis, Department of Agriculture, South Perth 6151, Australia⁴

Received 22 November 2005/Accepted 3 April 2006

In a previous report, we described a PCR protocol for the differentiation of the various species of the *Mycobacterium tuberculosis* complex (MTC) on the basis of genomic deletions (R. C. Huard, L. C. de Oliveira Lazzarini, W. R. Butler, D. van Soolingen, and J. L. Ho, *J. Clin. Microbiol.* 41:1637–1650, 2003). That report also provided a broad cross-comparison of several previously identified, phylogenetically relevant, long-sequence and single-nucleotide polymorphisms (LSPs and SNPs, respectively). In the present companion report, we expand upon the previous work (i) by continuing the evaluation of known MTC phylogenetic markers in a larger collection of tubercle bacilli ($n = 125$), (ii) by evaluating additional recently reported MTC species-specific and interspecific polymorphisms, and (iii) by describing the identification and distribution of a number of novel LSPs and SNPs. Notably, new genomic deletions were found in various *Mycobacterium tuberculosis* strains, new species-specific SNPs were identified for “*Mycobacterium canettii*,” *Mycobacterium microti*, and *Mycobacterium pinnipedii*, and, for the first time, intraspecific single-nucleotide DNA differences were discovered for the dassie bacillus, the oryx bacillus, and the two *Mycobacterium africanum* subtype I variants. Surprisingly, coincident polymorphisms linked one *M. africanum* subtype I genotype with the dassie bacillus and *M. microti* with *M. pinnipedii*, thereby suggesting closer evolutionary ties within each pair of species than had been previously thought. Overall, the presented data add to the genetic definitions of several MTC organisms as well as fine-tune current models for the evolutionary history of the MTC.

Tuberculosis (TB) is an age-old, and now resurgent, infectious disease that has spread to nearly every corner of the globe. Mycobacteria that cause human and/or animal TB are grouped together within the *Mycobacterium tuberculosis* complex (MTC). The classical species of the MTC include *Mycobacterium tuberculosis*, *Mycobacterium africanum* (of various subtypes and variants), *Mycobacterium microti*, and *Mycobacterium bovis* (along with the widely used vaccine strain *M. bovis* bacillus Calmette-Guérin [BCG]) (7, 38, 51). Newly recognized additions to the MTC include *Mycobacterium caprae* and *Mycobacterium pinnipedii* (2, 12). Although it is not presently an officially described organism, “*Mycobacterium canettii*” is another widely accepted member of the MTC (85). Lastly, there exist rare MTC variants, the so-called dassie and oryx bacilli, whose standing within the MTC remains to be defined (50, 84). Historically, taxonomic segregation of the MTC has been based upon each species’ unique combination of growth, morphological, physiological, and biochemical characteristics (57).

Variations in geographical distribution, host preference, virulence, and relative human infectivity also differentiate the tubercle bacilli (38). Since *M. tuberculosis* sensu stricto is disproportionately responsible for the worldwide TB burden (38), it is of great interest to identify potential virulence-associated genetic differences between *M. tuberculosis* and the other members of the MTC, in order to better understand the molecular mechanisms underlying human TB. It is also equally important to identify the genetic origins of *M. tuberculosis* interstrain pathobiological differences in order to progress the eventual association of strain genotypes with patient clinical phenotypes.

As a result of recent technological advances and the publishing of the complete DNA sequence of *M. tuberculosis* (10), our understanding of genetic variability within the MTC is ever increasing, and genetic criteria are now routinely used to identify MTC clinical and veterinary isolates to the species level and fingerprint them (3). Commonly employed protocols such as IS6110-restriction fragment length polymorphism (RFLP) analysis and spacer oligonucleotide typing (spoligotyping) provide the means to differentiate reliably among strains of *M. tuberculosis* (and certain other MTC members) at the DNA level by targeting polymorphic chromosomal regions that are usually associated with mobile insertion sequences and/or other repetitive elements (3). These forms of genotyping clarify interstrain relationships and are an essential component of epidemiological investigations and TB control programs. Spoligotyping is also able to discriminate most of the MTC species

* Corresponding author. Mailing address: Division of International Medicine and Infectious Diseases, Department of Medicine, Joan and Sanford I. Weill Medical College, Cornell University, Room A-421, 525 East 68th St., New York, NY 10021. Phone: (212) 746-6316. Fax: (212) 746-8975. E-mail: jlho@med.cornell.edu.

† Supplemental material for this article may be found at <http://jba.asm.org/>.

‡ Present address: Clinical Microbiology Service and the Department of Pathology, Columbia University Medical Center, New York-Presbyterian Hospital, New York, NY 10032.

from one another by way of signature features within their individual spoligopatterns (8, 43, 53, 83, 87). However, standard fingerprinting data cannot be used to infer the genetic origins of interspecies and strain-to-strain phenotypic differences, such as virulence, transmissibility, and acquired drug resistance, nor do these data provide the type of information needed to unequivocally reconstruct the phylogeny of the entire MTC taxon (91); molecular-level DNA sequence information is required for these purposes. As it so happens, the current sum data indicate that the MTC organisms and their sublineages achieved genetic and biological diversity through a combination of discrete acquired single-nucleotide polymorphisms (SNPs) (30) as well as by way of gross gene alterations such as deletions, insertions, inversions, and duplications (collectively known as region of difference [RD] loci) (7, 9).

SNPs within the MTC are a well documented means of acquired drug resistance, but for reasons that remain to be determined, very few spontaneous unselected mutations become fixed within the clonal population structure of the MTC, despite their having a mutation frequency believed to be in the range of most other bacteria (77). In fact, the degree of synonymous (genetically neutral) nucleotide substitution has been estimated to be as low as 1 in 10,000, whereas nonsynonymous nucleotide changes within MTC genomes have been estimated to outnumber synonymous SNPs nearly 2 to 1 (21, 30, 77); such proportions support the assumption that positive selection is driving diversity. Once established in a clonal lineage, SNPs in the MTC organisms are believed to be relatively stable over time (56), and frequently identified SNPs in certain genes have been used, alone or in combination, as either specific markers for a given MTC species or to categorize isolates into defined MTC phylogenies (7, 38, 51). As examples, a unique signatory "*M. canettii*" SNP has been described in *hsp65* (27), while commonly observed SNPs located in *katG* and *gyrA* are routinely assessed, because together they permit all MTC strains to be assigned into one of three principal genetic groups (PGG) (77). The distribution of the SNPs in *katG* and *gyrA* indicates that PGG1 strains are ancestral to PGG2 strains, which are in turn ancestral to PGG3 strains. In large-scale evaluations of expansive data sets of synonymous SNPs differentially present in the MTC (first identified by in silico genome-wide cross-comparisons), the population structure of *M. tuberculosis* was found to consist of up to nine major clusters, with *M. bovis* terminating a separate evolutionary offshoot bridged by the remaining MTC species (20, 29, 30). Clearly, comprehensive SNP informatics is a powerful tool to reconstruct familial interrelationships within the MTC, and, with the addition of newly identified SNPs, the branches of the MTC phylogenetic tree will continue to be refined. However, it is the acquisition of long-sequence polymorphisms (LSPs) rather than single-nucleotide changes that is now emerging as possibly the key mechanism of adaptation in mycobacteria (9, 36, 73) and a significant source of MTC interstrain genetic variability.

RD loci often represent irreversible changes in genetic makeup that likely result either from errors in DNA replication, movement of mobile genetic elements, mycobacteriophage-mediated transduction, or recombination between adjacent homologous DNA fragments with loss of the intervening sequence (9). Single-nucleotide deletions (SNDs) (e.g., MT1558 and

MT1560 [Rv1510] of *M. tuberculosis* strain CDC1551 [accession no. AE000516 {21}]) to LSPs up to 19.5 kb in length (34) have been identified. To be viable, RD loci must either be neutral changes or be positively selected mutations, such as in the deletion of *katG* as a means of acquired resistance to isoniazid in some *M. tuberculosis* strains (64), or such as may result from the insertion/excision of *IS6110* elements (18). As with SNPs, LSPs become fixed when a clone with the variant genotype is successfully transmitted, thereby becoming the founder of a new lineage. When the LSP derives from a single event (as best determined by locus sequence analysis), strains can be phylogenetically positioned relative to whether or not they bear that mutation. For example, all *M. bovis* BCG isolates possess the unique RD1^{BCG} deletion, but other BCG-restricted LSPs are variably distributed among the different BCG strains (4). As a result, RD1^{BCG} is believed to have been the primary attenuation event in the derivation of *M. bovis* BCG from *M. bovis*, while other BCG-restricted LSPs arose later as the *M. bovis* BCG sublineages diverged with subculture (4). MTC taxon-wide PCR investigations of various RD loci that were known to be altered in *M. bovis* BCG but wild type in *M. tuberculosis* similarly found some RD loci to be restricted to one MTC strain or species, while others appeared to be differentially distributed among the MTC groupings (9). These data enabled the independent construction of matching phylogenetic maps for the evolution of the entire MTC, based upon the sequential stepwise accumulation of LSPs (7, 51). As a result, it was proposed that an *M. africanum*-like clone diverged from *M. tuberculosis* with the loss of the RD9 locus, and that the successive loss of DNA in this de novo lineage resulted in the derivation and clonal expansion of the *dassie* bacillus, the oryx bacillus, *M. microti*, *M. pinnipedii*, *M. caprae*, and finally *M. bovis* as they adapted to new hosts (7, 51). *M. tuberculosis* sublineage- and strain-defining RD loci have also been identified, indicating the ongoing adaptation and genetic diversification of *M. tuberculosis* and allowing the mapping of the phylogenetic interrelationships of the interrogated strains (21). Additional LSPs that are unique to but one of the other MTC species have been identified as well (6, 7, 38, 48, 50, 53). Importantly, the original LSP-based evolutionary maps of the MTC (7, 51) are highly congruent with a subsequent synonymous SNP-based phylogenetic tree (30). Likewise, no inconsistencies were found in a cross-comparison of the multiple MTC species- and lineage-restricted SNPs and RD loci from our (Cornell) laboratory (38). During the course of the latter investigation, the lesser known MTC members were also evaluated, and several additional novel MTC polymorphisms were identified but not reported. In this article, we describe these data and expand upon the previous work by evaluating a larger collection of MTC strains. The results define several novel MTC species-, lineage-, and strain-specific SNPs and RD loci that in turn suggest heretofore unanticipated genealogical relationships between certain MTC radiations.

MATERIALS AND METHODS

Strains analyzed. A total of 125 MTC strains were evaluated in this report and included 71 MTC isolates utilized in a preceding companion paper (38). Most strains were previously characterized by various typing methods (42, 46) and were confirmed using a modified MTC PCR typing panel, described below (Table 1) (38). For the purposes of this study, the major subdivisions within the

TABLE 1. Summary amended MTC PCR typing panel results^a

MTC species	PCR-targeted locus [no. in profile] ^b								General profile ^e
	16S rRNA ^c [1]	<i>cfp32</i> (Rv0577) [2]	MiD3 (IS1561') [3]	RD4 (Rv1510) [4]	RD7 (Rv1970) [5]	RD1 (Rv3877 to Rv3878) [6]	RD9 ^d (Rv2073c) [7]	RD12 (Rv3120) [8]	
<i>M. canettii</i> (n = 5)	+	+	+	+	+	+	+	- ^f	1234567●
"Ancestral" <i>M. tuberculosis</i> [PGG1b] (n = 6)	+	+	+	+	+ ^g	+	+	+	12345678
"Modern" <i>M. tuberculosis</i> (PGG1b) (n = 14)	+	+	+	+	+	+ ^h	+	+	12345678
"Modern" <i>M. tuberculosis</i> (PGG2) (n = 19) ⁱ	+	+	+ ^j	+	+	+ ^k	+	+	12345678
"Modern" <i>M. tuberculosis</i> (PGG3) (n = 5)	+	+	+	+	+	+	+	+	12345678
<i>M. africanum</i> subtype Ib (n = 12)	+	+ ^l	+	+	+	+	- ^m	+	123456●8
<i>M. africanum</i> subtype Ia (n = 18)	+	+	+	+	-	+	-	+	1234●6●8
Dassie bacillus (n = 4)	+	+	+	+	-	- ⁿ	-	+	1234●●●8
Oryx bacillus (n = 2)	+	+	+	+	-	+	-	+	1234●6●8
<i>M. microti</i> (n = 10)	+	+	- ^o	+	-	+	-	+	12●4●6●8
<i>M. pinnipedii</i> (n = 7)	+	+ ^o	- ^o	+	-	+	-	+	12●4●6●8
<i>M. caprae</i> (n = 1)	+	+	+	+	-	+	-	-	1234●6●●
<i>M. bovis</i> (n = 14)	+	+	+	-	-	+	-	-	123●●6●●
<i>M. bovis</i> BCG (n = 8)	+	+	+	-	-	- ^o	-	-	123●●●●

^a Includes strains previously described (38) as well as strains evaluated herein for the first time.
^b Based upon the presence of (+) or a failure to amplify (-) a PCR fragment of the expected size; confirmatory PCR amplification of secondary targets from all strains (n = 125) was also performed as previously described for the loci MiD3, RD4, RD1, and RD9 (38), and the results were coincident with those of the primary targets (data not shown).
^c In addition to this gene, all MTC strains (n = 125) also amplified for the genetic elements *hsp65*, *rpoB*, *IS1081*, *mpb70*, *gyrA*, *gyrB*, *katG*, *oxyR'*, *pncA*, and Rv0911.
^d The locus introduced to expand the MTC PCR typing panel (38).
^e Based upon the presence or absence of a PCR fragment for each numerically assigned locus; not included are the isolates with independent overlapping strain-specific LSPs.
^f As a result of RD^{can}.
^g As an exception, *M. tuberculosis* strain CA-74 failed to amplify; resulting profile, 1234●678.
^h An alternate RD1 target amplicon (Rv3879c; 1,000 bp) was truncated by 57 bp (943 bp) in *M. tuberculosis* strains 97-803, 97-1177, and 97-1438.
ⁱ Includes Uganda genotype *M. tuberculosis* strains (n = 4).
^j As exceptions, *M. tuberculosis* strains 94-1055, 2002-1330, 2002-1384, and Percy47 failed to amplify; resulting profile, 12●45678.
^k *M. tuberculosis* strain CA-56 failed to amplify for an alternate RD1 target amplicon (Rv3879c) because of the RD1^{thA} LSP (62).
^l As an exception, *M. africanum* subtype Ib strain Percy13 failed to amplify; resulting profile, 1●3456●8.
^m Failure to amplify from *M. africanum* subtype Ib strain 15082 was supported by a flanking PCR that produced a bridge amplicon of the predicted size.
ⁿ As a result of RD1^{das}; the alternate RD1 target amplicon (Rv3879c) was PCR positive.
^o Failure to amplify was supported by a flanking PCR that produced a bridge amplicon of the predicted size.

MTC were considered to be "*M. canettii*" (n = 5), "ancestral" PGG1b *M. tuberculosis* (n = 6), "modern" PGG1b *M. tuberculosis* (n = 14), "modern" PGG2 *M. tuberculosis* (n = 19), "modern" PGG3 *M. tuberculosis* (n = 5), *M. africanum* subtype Ia (n = 18), *M. africanum* subtype Ib (n = 12), the dassie bacillus (n = 4), the oryx bacillus (n = 2), *M. microti* (n = 10), *M. pinnipedii* (n = 7), *M. caprae* (n = 1), *M. bovis* (n = 14), and *M. bovis* BCG (n = 8), as defined in the text below. A complete list of strains by final MTC species determination, isolate code, and country of isolation is provided in Table S1 of the supplemental material. Strains newly added to the collection and tested in this study are indicated by boldface type in Table S1. Strains with "Percy" designations were isolated at the Percy Hospital, Clamart, France (46). On the basis of genetic criteria (Tables 1 to 3), such as the fact that they retain the RD9 locus intact (7, 38), some strains formerly listed as either "*M. africanum* subtype II" or "indeterminate" were reclassified as *M. tuberculosis* in this study. Spoligotyping and IS6110-RFLP data additionally indicated that former *M. africanum* subtype II strains 133/92, 178/92, 290/92, and 1167/93 were of the recently recognized *M. tuberculosis* Uganda genotype (58; Stefan Niemann, personal communication).

PCR amplification primers and conditions. Purified DNA was prepared for PCR as previously described (38). For some strains, culture thermolysates (80°C for 30 min) were used as the source of DNA in PCR amplifications. The target gene loci and their primer names, primer sequences, and various amplification product sizes, as well as the programs used for amplification, are listed in Table S2 of the supplemental material. Primer sequences were either obtained from earlier publications (7, 27, 38, 45, 53, 57, 82) or created as previously described (38). In the case of PCR experiments to bridge novel deletions, primer pairs were generated to amplify targets upstream and downstream of, as well as successively outwards from, putative RD loci. When PCR amplicons were produced on either side of the locus of interest, the appropriate forward and reverse primers were combined to amplify across the altered site. Only the primer pairs that successfully generated products bridging putative LSPs are listed in Table S2. Similarly, in searching for novel SNPs by PCR amplification of an array of targets, only those primer pairs targeting loci with polymorphisms of interest are listed in

Table S2. The general PCR recipe was identical to that used previously (38), with the following exceptions: (i) when thermolysates were used as the source DNA, 2.5 to 5.0 µl of sample was added, and (ii) PCR for Rv3879c and RD701 required additional dimethyl sulfoxide (5.0 µl, 10% of total volume) for optimal amplification. In both cases, the volume of water was adjusted accordingly for a final volume of 50 µl. PCR amplification from purified DNA was performed using various programs: program 1a (with an initial denaturation step of 5 min at 94°C followed by either 25 or 35 cycles of 1 min at 94°C, 1 min at 60°C, and 1 min at 72°C, and ending with a final elongation step for 10 min at 72°C), program 2a (program 1a but with an annealing temperature of 65°C), program 3a (program 1a but with an annealing temperature of 55°C), or program 4a (program 1a but with an annealing temperature of 50°C). PCR testing of DNA thermolysates was performed in a similar manner using various programs: program 1b (with an initial denaturation step of 5 min at 94°C, followed by either 40 or 45 cycles of 1 min at 94°C, 1 min at 60°C, and 4 min at 72°C, and ending with a final elongation step for 10 min at 72°C), program 2b (program 1b but with an annealing temperature of 65°C), program 3b (program 1b but with an annealing temperature of 55°C), or program 4b (program 1b but with an annealing temperature of 50°C). Programs 1b and 2b were also used to amplify from purified DNA when potential target PCR fragments were larger than 1,250 bp. PCR products were visualized as previously described (38). Importantly, all PCR tests included a positive control for amplification of the target locus as well as a positive control for the test DNA. Furthermore, all negative and unexpected positive PCR results were repeated and confirmed at least once. The previously described loci RD1^{BCG}, RD4, RD8, RD9, MiD3, and TbD1 were each PCR evaluated by targeting at least two DNA fragments that cross a deletion junction site and/or that are internal to the deletion. Strains that failed to amplify from both targets were considered to bear the LSP in question. RD7, RD10, RD12, RD13, and N-RD25^{bovis/cap} were evaluated by a single deletion-internal PCR test per LSP. RD701, RD702, RD711, and RD713 were PCR screened using primers flanking these LSPs to generate PCR products bridging each locus and that are of different specific sizes depending upon whether the expected mutations are present or not. RD bridge amplicons of the same approximate size as those

TABLE 2. Summary differential distribution of known SNPs among the MTC species^a

MTC species	PGG ^c	Result of SNP analysis by locus ^b											
		<i>katG</i> ²⁰³	<i>katG</i> ⁴⁶³	<i>gyrA</i> ⁹⁵	<i>hsp65</i> ⁶³¹	<i>mmpL6</i> ⁵⁵¹	<i>gyrB</i> ^{Δd}					<i>oxyR</i> ²⁸⁵	<i>pncA</i> ¹⁶⁹
							675	756	1311	1410	1450		
" <i>M. canettii</i> " (n = 5)	1b	ACC	CTG	AGC	T	AAC	C	G	T	C	G	G	C
"Ancestral" <i>M. tuberculosis</i> (n = 6)	1b	ACC	CTG	AGC	C	AAC	C	G	T	C	G	G	C
"Modern" <i>M. tuberculosis</i> (n = 14)	1b	ACC	CTG	AGC	C	Fail ^f	C	G	T	C	G	G	C
"Modern" <i>M. tuberculosis</i> (n = 19) ^e	2	ACC	CGG	AGC	C	Fail	C	G	T	C	G	G	C
"Modern" <i>M. tuberculosis</i> (n = 5)	3	ACC	CGG	ACC	C	Fail	C	G	T	C	G	G	C
<i>M. africanum</i> subtype Ib (n = 12)	1b	ACC	CTG	AGC	C	AAC	C	G	T	C	T	G	C
<i>M. africanum</i> subtype Ia (n = 18)	1a	ACT	CTG	AGC	C	AAC	C	G	T	C	T	G	C
Dassie bacillus (n = 4)	1a	ACT	CTG	AGC	C	AAC	C	G	T	C	T	G	C
Oryx bacillus (n = 2)	1a	ACT	CTG	AGC	C	AAG	C	G	T	C	T	G	C
<i>M. microti</i> (n = 10)	1a	ACT	CTG	AGC	C	AAG	T	G	T	C	T	G	C
<i>M. pinnipedii</i> (n = 7)	1a	ACT	CTG	AGC	C	AAG	C	G	T	C	T	G	C
<i>M. caprae</i> (n = 1)	1a	ACT	CTG	AGC	C	AAG	C	A	G	C	T	A	C
<i>M. bovis</i> (n = 14)	1a	ACT	CTG	AGC	C	AAG	C	A	T	T	T	A	G
<i>M. bovis</i> BCG (n = 8)	1a	ACT	CTG	AGC	C	AAG	C	A	T	T	T	A	G

^a Includes strains previously described (38) as well as newly evaluated isolates and a newly targeted locus (*mmpL6*⁵⁵¹).

^b As determined by sequence analysis and/or PCR-RFLP.

^c Principle genetic group (expanded) as determined by SNP analysis of *katG*²⁰³, *katG*⁴⁶³, and *gyrA*⁹⁵ (24, 77).

^d By nucleotide position within *gyrB* according to previous nomenclature.

^e Includes Uganda genotype *M. tuberculosis* strains (n = 4).

^f Fail, locus is deleted and no PCR products were amplified.

confirmed by sequence analysis were held to possess the LSP in question as well. In the case of RD713, failure to amplify from certain MTC subdivisions was reasoned to be due to either an alternate mutation or an excessive DNA fragment size, consistent with available genomic database information (listed below).

In addition to the above, the deletion loci *MiD3* in all *M. microti* and *M. pinnipedii* strains, *TbD1* in all "ancestral" *M. tuberculosis* strains as well as *RD1*^{BCG} in all *M. bovis* BCG strains, and *N-RD25*^{bovis/cap} in all *M. caprae*, *M. bovis*, and *M. bovis* BCG strains were also reconfirmed using LSP-flanking PCR

TABLE 3. Differential distribution of additional RD loci among the MTC species^a

MTC species	PCR-targeted locus ^b (no. of strains evaluated)											
	<i>TbD1</i> (<i>mmpL6</i>)	<i>RD8</i> (Rv3617)	<i>RD10</i> (Rv0222)	<i>RD13</i> (Rv1257c)	<i>N-RD25</i> ^c (<i>PPE67</i>)	<i>RD701</i> (Rv0124)	<i>RD702</i> (Rv0186)	<i>RD711</i> (Rv1333 to Rv1336)	<i>RD713</i> ^d (Rv1977 to Rv1979c)	<i>PPE55</i> (Rv3343c)	<i>3' cfp32</i> (Rv0577 to Rv0578c)	<i>pks15/1</i> ^e
" <i>M. canettii</i> "	+ (5)	+ (5)	+ (5)	+ (5)	+ (5)	+ (2)	+ (2)	+ (2)	- (2)	+ ^f (5)	- ^d (5)	0 (2)
"Ancestral" <i>M. tuberculosis</i> (PGG1b)	+ (6)	+/- ^g (6)	+ (6)	+ (6)	+ (6)	+ (2)	+ (2)	+ (2)	- (2)	+ (6)	+ (6)	0 (2)
"Modern" <i>M. tuberculosis</i> (PGG1b)	del (14)	+ (14)	+ (14)	+ (14)	+/-del ^h (13)	+ (3)	+ (2)	+ (3)	- (3)	+ (14)	+ (14)	0 (2)
"Modern" <i>M. tuberculosis</i> (PGG2) ^f	del (19)	+ (19)	+ (19)	+ (19)	+ (19)	+ (4)	+ (4)	+ (4)	+ (4)	+/- ^{d,j} (19)	+ (19)	7 (4)
"Modern" <i>M. tuberculosis</i> (PGG3)	del (5)	+ (5)	+ (5)	+ (5)	+ (5)	+ (3)	+ (3)	+ (3)	+ (3)	+ (5)	+ (5)	7 (2)
<i>M. africanum</i> subtype Ib	+ (12)	+ (12)	+ (12)	+ (12)	+ (12)	+ (6)	+ (3)	+/-del ^k (12)	del (12)	+ (12)	+/- ^l (12)	0 (3)
<i>M. africanum</i> subtype Ia	+ (18)	- (18)	- (18)	+ (18)	+ (18)	del (18)	del (18)	+ (3)	- (3)	+ (18)	+ (18)	6 (2)
Dassie bacillus	+ (4)	- (4)	- (4)	+ (4)	- ^m (4)	+ (4)	+ (2)	+ (2)	- (2)	+ (4)	+ (4)	6 (4)
Oryx bacillus	+ (2)	- (2)	- (2)	+ (2)	+ (2)	+ (2)	+ (2)	+ (2)	- (2)	+ (2)	+ (2)	6 (2)
<i>M. microti</i>	+ (10)	- (10)	- (10)	+ (10)	+ (10)	+ (2)	+ (2)	+ (2)	- (2)	del ⁿ (10)	+ (10)	6 (2)
<i>M. pinnipedii</i>	+ (7)	- (7)	- (7)	+ (7)	+ (7)	+ (2)	+ (2)	+ (2)	- (2)	del ⁿ (7)	+ (7)	6 (2)
<i>M. caprae</i>	+ (1)	- (1)	- (1)	- (1)	del (1)	+ (1)	+ (1)	+ (1)	- (1)	+ (1)	+ (1)	6 (1)
<i>M. bovis</i>	+ (14)	- (14)	- (14)	- (14)	del (14)	+ (5)	+ (2)	+ (2)	- (2)	+ (14)	+ (14)	6 (2)
<i>M. bovis</i> BCG	+ (8)	- (8)	- (8)	- (8)	del (8)	+ (3)	+ (2)	+ (2)	- (2)	+ (8)	+ (8)	6 (2)

^a Includes strains previously described (38) as well as newly evaluated isolates and loci.

^b +, PCR fragments of the expected undisturbed sizes were observed from a deletion-associated target locus, either internal, bridging, or both; -, no amplicons were generated, either internal, crossing a deletion junction, or both; del, deleted, meaning no amplicons that were internal to the LSP were generated and/or a truncated bridge PCR fragment of the predicted size was seen.

^c *N-RD25*^{bovis/cap} was targeted.

^d Locus sequence to be described in greater detail elsewhere.

^e Given as the number of nucleotides deleted within the polymorphic locus as determined by sequence analysis (11).

^f *PPE55*^{can} PCR amplification success was primer sequence dependent.

^g *M. tuberculosis* strain Percy211 failed to amplify across the 5' junction of the *RD8* locus as a result of the *RD236a* LSP (82).

^h *M. tuberculosis* strains 97-742, 97-803, 97-818, 97-1177, 97-1289, 97-1438, Percy7, and Percy91 possessed the *N-RD25*^{tbA} LSP; strains 13876 and Percy209 possessed the *N-RD25*^{tbB} LSP.

ⁱ Includes Uganda genotype *M. tuberculosis* strains (n = 4).

^j *M. tuberculosis* strains 94-1055, 2002-1330, 2002-1384, and Percy47 failed to amplify.

^k Of the *M. africanum* subtype Ib isolates, strains Percy13, Percy16, Percy17, Percy56, Percy58, Percy60, and Percy204 possessed the *RD711* LSP, while strains 15082, AF0236, AF0271, Percy59, and Percy99a produced a full-length amplicon.

^l *M. africanum* subtype Ib strain Percy13 failed to amplify.

^m As a result of *N-RD25*^{das} (50).

ⁿ As a result of *MiD3*.

primers to bridge these respective deletions. The RD9 locus in *M. africanum* subtype Ib strain 15082 was confirmed to be deleted by way of a PCR that bridged the truncated locus, while the N-RD25^{bovis/cap} locus in the oryx bacillus strains was confirmed to be intact by a flanking PCR as well.

Automated DNA sequencing and PCR product restriction enzyme digest analysis. Direct sequencing of PCR fragments was performed as previously detailed, with the PCR amplification primers acting as sequencing primers (38). Additional sequencing primers internal to the target amplicons were also occasionally used and are listed in Table S2. Combined, a minimal single overlap from two directions was usually achieved across SNP sites, as well as deletion junction sites, upon completion of sequence analysis. The identities of all “wild-type” amplicons were confirmed by sequence analysis in *M. tuberculosis* strain H37Ra or an appropriate substitute, and each amplicon bridging a deletion was sequenced from at least one strain bearing the LSP. The Lasergene program (DNASTAR Inc., Madison, Wis.) was used to compare the derived sequence data with DNA sequence information taken from sequence databases (listed below) and supported the conclusions expressed herein. As with LSPs, SNPs are mostly described relative to the *M. tuberculosis* strain H37Rv full-genome sequence (accession no. AL123456 [10]), even though in some cases it is recognized that this perspective may be opposite to the actual direction of nucleotide change, especially with respect to “*M. canettii*.” In the case of loci deleted from *M. tuberculosis* strain H37Rv, the associated SNPs in other MTC organisms are described relative to the *M. bovis* strain AF2122/97 full-genome sequence (accession no. BX248333 [25]). Previously identified SNPs are numbered in this article by the nucleotide or codon at which they are located in the full coding sequence of the gene, as indicated. Unless otherwise noted, all novel SNPs are likewise numbered according to gene position (5' to 3'). SNPs within the 3' *cpf32*, *aroA*, *Tbd1*, and *RD13* loci are numbered according to the site of mutation within the respective PCR amplicons. This was done in order to enumerate SNPs that map to noncoding regions. Comparative analysis of key lineage- and species-specific nucleotide substitutions were further evaluated in this study by sequencing and/or by previously described protocols for RFLP analysis of restriction enzyme-digested PCR products (PCR-RFLP) in the PCR fragments for *katG*²⁰³ (codon 203; ACC→ACT), *katG*⁴⁶³ (codon 463; CTG→CGG), *hsp65*⁵³¹ (gene position 399; 399C→T), *oxyR*²⁸⁵ (gene position 285; 285G→A), *pncA*¹⁶⁹ (gene position 169; 169G→C), and *gyrB*^Δ (Δ represents the differential combinations of SNPs for base substitution sites in *M. tuberculosis* [*gyrB*^{tb}], *M. africanum* subtype I [*gyrB*^{af}; 1450G→T], *M. microti* [*gyrB*^{mic}; 675C→T, 1450G→T], *M. caprae* [*gyrB*^{cap}; 756G→A, 1311G→T, 1450G→T], and *M. bovis* [*gyrB*^{bovis}; 756G→A, 1410C→T, 1450G→T]) (24, 27, 38, 57, 71, 77). Note that the numerical positions provided for the known and novel SNPs within *gyrB*^Δ are given relative to previous nomenclature and not according to the *gyrB* (Rv0005) gene sequence of *M. tuberculosis* H37Rv (accession no. AL123456 [10]). SNP analysis of the commonly targeted *gyrA*⁹⁵ (codon 95; AGC→ACC) amplicon was done by sequence analysis and/or a novel PCR-RFLP strategy. PGG segregation was accomplished by comparative analysis of the commonly observed SNPs in *katG*⁴⁶³ and *gyrA*⁹⁵ (77). SNP analysis of *katG*²⁰³ was used to further segregate PGG1a isolates from PGG1b strains (24). As a complement to, or in lieu of, comprehensive sequence analyses, additional new PCR-RFLP protocols were constructed to evaluate the distribution of various other novel MTC species- and lineage-specific SNPs. The tested loci, their respective restriction enzyme(s) utilized, and their resulting expected restriction fragment sizes in the MTC species that are differentiated are listed in Table S3 in the supplemental material. Note that the *Sau3AI* enzymatic digestion of *rpoB*¹⁰⁴⁹, which segregates the *M. africanum* subtype Ia strains, was first developed by Lee et al. (45), but only the *M. africanum* subtype Ia type strain (ATCC 25420) was reported. All restriction enzymes listed above were purchased from New England Biolabs (Beverly, Mass.), and digest protocols followed the manufacturer's recommendations. The visualization of RFLP patterns was performed in a manner similar to that of previous descriptions (38), and the outcomes approximated the predicted band sizes for each RFLP strategy. Some PCR products containing novel SNPs of interest were evaluated by sequence analysis alone (from those strains that amplified in PCR) and included the *mmpL6*⁵⁵¹, *pncA*¹³⁸, and Rv0911²⁴⁹ SNP targets as well as the *PPE55*, *PPE55*^{can} (360-bp fragment), 3' *cpf32*, *pks15/1*, *aroA*, *Tbd1* (500-bp fragment), and *RD13* amplicons. Note that, in addition to the full evaluation of all strains within a major MTC subdivision associated with a given intraspecific polymorphism, the *M. caprae* type strain (CIP 105776) and at least two strains from every other major subdivision, as well as two *M. tuberculosis* Uganda genotype strains, were also evaluated by PCR or sequence analysis and/or PCR-RFLP for the genetic variation(s) in question. PCR targets that were predominantly evaluated by PCR-RFLP were also sequenced from at least one strain of each major MTC subdivision. With respect to interspecific SNPs, at least two strains from every major subdivision, as well as two *M.*

tuberculosis Uganda genotype strains and the *M. caprae* type strain, were evaluated. Exceptions occurred only when an SNP-bearing locus was deleted. The total number (*n*) of MTC species evaluated per locus is provided in the appropriate tables.

Web addresses utilized in this study were the following: GenBank and BLAST, <http://www.ncbi.nlm.nih.gov>; the Sanger Center, http://www.sanger.ac.uk/Projects/M_tuberculosis, http://www.sanger.ac.uk/Projects/M_bovis, http://www.sanger.ac.uk/Projects/M_microti, and http://www.sanger.ac.uk/Projects/M_leprae; The Institute for Genomic Research, <http://www.tigr.org> (for *M. tuberculosis* strains 210 and CDC1551); and Tuberculist, <http://genolist.pasteur.fr/TubercuList>.

RESULTS

A modified MTC PCR typing panel. In a previous report, we described the first solely PCR-based strategy for the genetic differentiation of the MTC species (38). In this protocol, a set of seven intra- or interspecific RD loci and controls were targeted for PCR amplification in separate but simultaneous reactions. The reactions either amplified successfully or failed, depending upon the genomic content of the test strain. When separated by agarose gel electrophoresis, the resulting pattern of products was shown to be indicative of MTC species identity. Because overlapping but otherwise unrelated LSPs in different MTC species are known to occur, secondary confirmatory genetic targets were also provided to strengthen the results as needed. The robustness of the protocol was validated by the fact that several isolates previously misidentified and subsequently reclassified in our study were similarly tagged as likely incorrect by Brosch et al. (7) and were outliers in name, but correctly positioned, within the synonymous SNP-based MTC phylogenetic tree of Gutacker et al. (30).

For the current study, we incorporated an amplification for RD9 as an eighth target locus in the routine MTC PCR typing panel in order to better differentiate the *M. africanum* subtypes, as was previously recommended (38). This addition was placed at position seven, and the amplification for RD12 was shifted to position eight of the series in order to maintain the sloping cascade of differentially sized amplicons and to ease visual inspection. An example of the modified MTC PCR typing panel using *M. tuberculosis* strain H37Ra is provided in Fig. 1. In essence, the MTC PCR typing panel patterns for all strains that were tested were similar to that illustrated, except when the PCR failed for a given locus and no product was generated. Only *M. tuberculosis* and “*M. canettii*” strains produced a band for the newly incorporated RD9 PCR. The results from testing of all MTC strains in our collection by the revised protocol is summarized in Table 1 and includes many species and strains evaluated for the first time. For instance, the *dassie* bacillus and *M. africanum* subtype Ib (defined below) each gave unique patterns, which will be useful in the future identification of strains. In contrast, the oryx bacillus pattern was identical to that of *M. africanum* subtype Ia, while the *M. pinnipedii* pattern was identical to that of *M. microti*, and these results illustrate limitations of this identification strategy. In possessing similar deletion profiles, however, these data also indicated a proximity in the evolutionary history of these particular sets of MTC species. Several *M. tuberculosis* strains were also identified that unexpectedly failed to amplify from certain loci. These strains were clearly *M. tuberculosis*, because they were PCR positive for RD9 and RD12 (failure to amplify RD12 excludes “*M. canettii*,” which possesses the overlapping RD12^{can} LSP) (38, 48) and so provided the initial impetus to investigate further all

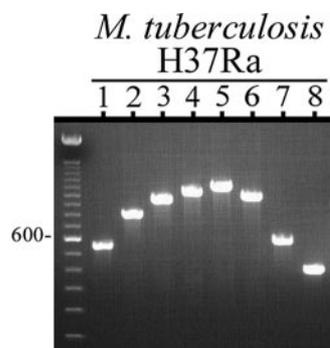


FIG. 1. Composite revised MTC PCR typing panel. Illustrated is an example of the MTC PCR typing panel output pattern for a single *M. tuberculosis* isolate (strain H37Ra). The protocol was amended from reference 38 to include an amplification test for the RD9 locus. A total of 125 MTC isolates were tested, as summarized in Table 1. Patterns differed by the presence or absence of the various PCR fragments in ways that were generally consistent per MTC species, but unexpected lineage- or strain-specific patterns sometimes resulted for certain isolates, as described in Table 1. PCR products and the 100-bp ladder (unlabeled lane) were visualized by agarose gel electrophoresis and ethidium bromide staining. Images were captured with the Nighthawk Imaging System (PDI Inc.) and Quality One software package (PDI Inc.). Lanes: 1, 16S rRNA gene; 2, *cfp32* (Rv0577); 3, MiD3 (IS1561'); 4, RD4 (Rv1510); 5, RD7 (Rv1970); 6, RD1 (Rv3877-3878); 7, RD9 (Rv2073c); and 8, RD12 (Rv3120).

new presumptive LSPs as well as other polymorphisms that had been observed.

As described below, novel SNPs were also found by sequence analysis of amplicons generated for investigational purposes both within and originally outside of this study, and their MTC distribution was subsequently characterized. Figure 2 summarizes the main findings of the overall study from a phylogenetic perspective. As an added point of interest, all MTC strains evaluated ($n = 125$) amplified for the targeted mycobacterial 16S rRNA, *hsp65*, and *rpoB* loci and were also PCR positive for the IS1081, *mpb70* (Rv2875), *gyrA*, *gyrB*, *katG*, *oxyR'*, and *pncA* elements. In addition, Rv0911 was found here for the first time to be a gene present in all MTC species tested ($n = 125$).

In this study, *M. africanum* subtype Ib strain Percy13 was the only isolate ($n = 125$) that failed to amplify for the *cfp32* locus (Tables 1 and 3), an MTC PCR typing panel gene previously described as being universally present among the MTC organisms (37, 38). For unknown reasons, attempts to bridge the locus failed, even though the downstream Rv0575c and upstream Rv0579c genes were successfully amplified by PCR (results not shown). As with all polymorphisms, it should be pointed out that, in the absence of additional strains that bear the same mutation isolated from separate hosts, it remains possible that the observed change is actually a dead-end mutation that may have been adaptive only in the short term but that obviates long-term fitness. For example, deletion of an antigenic epitope to escape immune recognition or the removal of genes unnecessary for growth in culture (e.g., as is believed to have happened with *M. bovis* BCG [54]) may lead to an attenuation of virulence in the emerging strain. It so happens that mutagenesis studies failed to identify *cfp32* as a critical gene for in vitro growth (70; our own unpublished

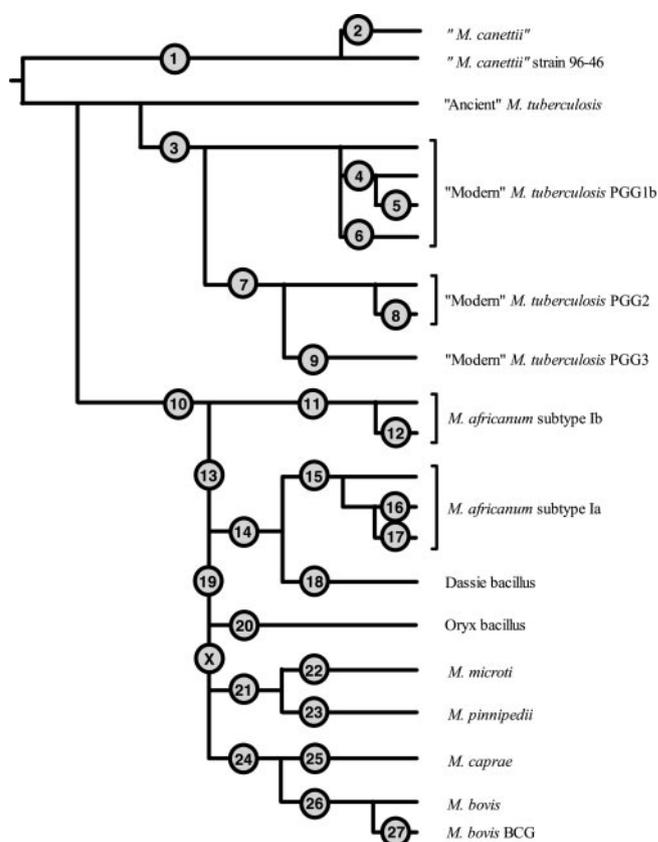


FIG. 2. Summary diagram and phylogenetic interpretation of data collected in the current study. Shown are the various major divisions of the MTC segregated according to the presence or absence of the investigated inter-, intra-, and lineage-specific polymorphisms. Not included are most strain-specific SNPs and LSPs that were identified, as well as the potentially novel RD loci that were noted but not characterized in this study. Circles are placed at points in evolutionary history beyond which each strain that was evaluated possessed a consistent set of polymorphisms. The circles are numbered in the figure to denote the following: circle 1, RD12^{can}, *hsp65*⁶³¹, *gyrB*^{can}, *mmpL6*¹⁸⁷⁹, *oxyR*¹⁸³, *pncA*¹³⁸, Rv0911²⁴⁹, RD13²⁵⁵, and *aroA*⁹¹; 2, *PPE55*^{can}; 3, TbD1; 4, N-RD25^{tbA}; 5, RD1^{tbB}; 6, N-RD25^{tbB}; 7, *pks15/1* (7-bp deletion) and *katG*⁴⁶³; 8, *aroA*¹¹⁷; 9, *gyrA*⁹⁵; 10, RD9 and *gyrB*^Δ (1450G→T); 11, RD713, TbD1¹⁹⁷, and *aroA*²⁸⁵; 12, RD711; 13, RD7, RD8, RD10, *pks15/1* (6-bp deletion), 3' *cfp32*³¹¹, RD13¹⁷⁴, *PPE55*²¹⁴⁸, and *PPE55*²¹⁵⁴; 14, Rv1510¹¹²⁹; 15, RD701, RD702, and *hsp65*⁵⁴⁰; 16, *rpoB*¹¹⁶³; 17, *rpoB*¹⁰⁴⁹; 18, RD1^{das}, N-RD25^{das}, 3' *cfp32*²²⁴, and Rv0911³⁸⁹; 19, *mmpL6*⁵⁵¹; 20, *gyrB*^{oryx}, TbD1¹⁷¹, *PPE55*²¹⁶², and *PPE55*²¹⁶³; 21, MiD3 and RD13³⁸⁰; 22, *gyrB*^{mic}, 16S rRNA¹²³⁴, and RD13⁶⁷; 23, 16S rRNA¹²⁴⁹ and RD13²²⁸; 24, RD12, RD13, N-RD25^{bovis/cap}, *gyrB*^Δ (756G→A), and *oxyR*²⁸⁵; 25, *gyrB*^Δ (1311T→G); 26, RD4, *gyrB*^Δ (1410C→T), and *pncA*¹⁶⁹; and 27, RD1^{BCG}. Note that distances are arbitrary and do not reflect the number of phylogenetically relevant polymorphisms present at each juncture. Although not evaluated in this study, as discussed in the text, data suggest that the -215 *narGHJI* SNP would also occupy circle 3. The loss of spoligotype spacers 33 to 36 may additionally correlate with circle 7, the absence of spacers 9 and 39 with circle 10, the deletion of spacers 40 to 43 with an event subsequent to divergence of the oryx bacillus putatively at circle X, and the deletion of spacers 3 and 16 with circle 24 of the above-proposed phylogeny (8, 40, 43, 55, 83).

data), so the meaning of the failure to PCR amplify *cfp32* from strain Percy13 cannot be determined at present.

"*M. canettii*." "*M. canettii*" is a very rare MTC biotype that was first isolated from a French farmer in 1969 (strain 96-46,

evaluated herein) (27). However, most subsequently isolated strains have come from patients with known links to east Africa, the presumed origin of this microbe (7, 17). “*M. canettii*” differs from all MTC strains described thus far in showing a smooth, white, and glossy colony morphology on solid media, as well as for having a unique direct repeat (DR) locus (the spoligotyping target) and a characteristic lipid content (17). “*M. canettii*” also exhibits a diversity of genotypes and sequences that appears to be in excess of all other MTC species (83, 85). As a result, it has been proposed that “*M. canettii*” represents the most phylogenetically distant member of the MTC and that this species probably diverged from a common mycobacterial ancestor prior to the evolutionary bottleneck thought to underlie the genesis of the remaining MTC species (7). Because of the sporadic incidence of human TB caused by “*M. canettii*,” it has been hypothesized that an as-yet-identified animal or environmental source may constitute the true reservoir of this organism (85).

In this study, all “*M. canettii*” strains evaluated were PGG1b and retained the TbD1, RD9, and *pks15/1* loci intact (Tables 1 to 3). These strains also possessed the previously described characteristic *hsp65*⁶³¹ polymorphism and failed to amplify for the RD12 locus as a result of the recently defined RD12^{can} LSP (Tables 1 and 2; Fig. 2, note circle 1) (7, 27, 48). During the course of this work, we further independently identified previously noted SNPs in *aroA* (*aroA*⁹¹) and *gyrB*^Δ (*gyrB*^{can}), showing that they were indeed unique to “*M. canettii*” (41, 42). Here, and as with all subsequent SNP data not otherwise provided in the text, the codon affected, nucleotide change, nature or location of change, and number of strains evaluated with and without the SNP in question are listed in Table 4. Strain-specific synonymous *gyrB*^Δ position 735 (C→T, codon 203) and 1011 (C→T, codon 295) SNPs were also identified in “*M. canettii*” strain 96-46. In addition to the above, we identified a series of new “*M. canettii*”-specific SNPs. These include nucleotide differences in *oxyR*¹⁸³ (*oxyR*¹⁸³), *pncA* (*pncA*¹³⁸), *mmpL6* (*mmpL6*¹⁸⁷⁹), Rv0911 (Rv0911²⁴⁹), and RD13 (RD13²⁵⁵) (Table 4; Fig. 2, note circle 1). Interestingly, the “*M. canettii*” *mmpL6*¹⁸⁷⁹ G→C SNP overlapped a nonsynonymous strain-specific *mmpL6*¹⁸⁷⁹ G→A SNP in the dassie bacillus isolate 68/7171 (Table 4).

In PCR amplification screening, we also found that all “*M. canettii*” strains failed to amplify for the 3′ *cfp32* locus, while four of five strains failed to amplify a portion of the *PPE55* gene (Table 3); strain 96-46 was PCR positive for the *PPE55* fragment. The nature of the 3′ *cfp32* PCR amplification failure will be described in detail elsewhere. In the case of the *PPE55* amplification failure, PCR experiments to bridge this locus in “*M. canettii*” (*PPE55*^{can}) indicated that it was actually intact. Rather than being deleted, sequence analysis of the bridge amplicon revealed an extensive degree of polymorphism in the PCR-negative strains, including at the original primer annealing sites, thereby providing a plausible explanation for the initial failure to amplify *PPE55*^{can} by PCR. Overall, 54 SNPs and a 6-bp microdeletion were identified in this otherwise 346-bp amplicon (84.4% gene sequence homology, 81.7% predicted amino acid homology, sequence in-frame, and no termination signals introduced) (Table 4; Fig. S1 in the supplemental material). Identical *PPE55*^{can} polymorphisms at positions *PPE55*²¹⁴⁸, *PPE55*²¹⁶², and *PPE55*²¹⁶³ of other MTC

species are likely coincidental and not the product of direct descent, given the heightened substitution rate in this amplicon in “*M. canettii*” and the overall weight of phylogenetic data to the contrary, including the fact that the *PPE55* sequences of *M. africanum* subtype Ib isolates matched that of *M. tuberculosis* H37Rv (Tables 1 to 4). Similarly, the amplicon from strain 96-46 was identical to that of *M. tuberculosis* H37Rv, even though this strain bore each of the other “*M. canettii*”-specific markers (Tables 1 to 4). Alternative “*M. canettii*”-based primers for the same 340-bp amplicon successfully amplified the locus in the requisite four of five “*M. canettii*” strains, supporting the veracity of the above determination. Therefore, these data indicate that strain 96-46 is genetically closer to the remaining MTC species than are the other “*M. canettii*” strains evaluated in this study and are in accordance with similar conclusions previously expressed in regard to overall “*M. canettii*” genetic diversity (Fig. 2, note circle 2) (17). It should also be mentioned that, although “*M. canettii*” clearly exhibited an overall elevated rate of nucleotide substitution in the examined loci compared with the other MTC species, with the exception of the *PPE55*^{can} fragment, none of the sequenced loci contained more than one SNP, and several PCR fragments were nonpolymorphic.

Combined, these findings support the hypothesis that the other MTC species are actually clonally derived from an “*M. canettii*”-like organism (7). As such, “*M. canettii*” may offer a good genomic reference point to investigate how genes have evolved to greater virulence in *M. tuberculosis* as well as to research the biological significance of the augmented mutability of PPE genes relative to other tubercle bacillus genes. Of note, a recent publication describing a novel “prototuberculosis” designation of genotypically divergent MTC organisms, into which “*M. canettii*” was enveloped, arrived at similar conclusions with respect to the evolutionary meaning of the unique genetic character of these bacilli (31).

M. tuberculosis. *M. tuberculosis* is the predominant cause of human TB and arguably the most successful of human bacterial pathogens. The collection of MTC isolates investigated in this study, as segregated by PCR analysis for the TbD1 locus and *katG/gyrA* SNP analyses, included strains from the so-called “ancestral” (PGG1b) *M. tuberculosis* subgroup as well as each of the “modern” (PGG1b, PGG2, and PGG3) *M. tuberculosis* lineages, including the newly recognized *M. tuberculosis* Uganda genotype (PGG2) (Tables 2 and 3; Fig. 2, note circles 3, 7, and 9) (33, 58, 75). However, it should be noted that this collection is by no means a representative sample of the genetic and geographic diversity of all *M. tuberculosis* lineages. Nonetheless, we found that each *M. tuberculosis* strain retained the RD9 locus intact and also possessed the *gyrB*^{tb} allelic fragment (Tables 1 and 2). Of note, the “ancestral” and “modern” PGG1b *M. tuberculosis* strains also exhibited an undisturbed *pks15/1* locus, while “modern” PGG2 and PGG3 *M. tuberculosis* strains bore the characteristic 7-bp *pks15/1* microdeletion (Table 3; Fig. 2, note circle 7) (11, 48). None of the inter- and intraspecific SNPs and RD loci present in the other MTC species, and evaluated herein, were evidenced by the investigated *M. tuberculosis* strains (Tables 1 to 5). However, some additional lineage- and strain-specific RD loci that overlapped the LSPs of other MTC isolates were identified (Tables 1 and 3) and characterized (Table 5), as described below. Note that

TABLE 4. Description of novel species- and lineage-specific single-nucleotide changes

MTC organism(s) with polymorphism(s) and no. of strains evaluated	Locus ^{a,b}	Codon affected (no.)	Nucleotide change	Nature or location of change(s)	No. of strains with mutation(s)	No. of other MTC isolates tested ^c
"M. canettii," 5	<i>gyrB</i> ^{cano}	444 ^d	1458C→T	Synonymous	5	120
	<i>mmpL6</i> ¹⁸⁷⁹	627	1879G→C	Nonsynonymous	5	82
	<i>oxyR</i> ¹⁸³	61	183G→T	Synonymous	5	120
	<i>pncA</i> ¹³⁸	46	138A→G	Synonymous	5	34
	Rv0911 ²⁴⁹	83	249C→T	Synonymous	5	30
	RD13 ²⁵⁵	422 ^e	255C→T	Synonymous	5	45
	<i>aroA</i> ⁹¹		91T→C	Intergenic ^f	5	52
	<i>PPE55</i> ^{can}	Multiple	54 SNPs 6 SNDS	19 Nonsynonymous 2-Amino acid net decrease	4 ^p	35
"Modern" PGG2 <i>M. tuberculosis</i> , 19	<i>aroA</i> ¹¹⁷		117G→A	Intergenic ^f	8 ^q	49
<i>M. africanum</i> subtype Ib, 12	TbD1 ¹⁹⁷		197C→T	Intergenic ^g	12	20
	<i>aroA</i> ²⁸⁵	49 ^f	285G→A	Synonymous	12	45
PGG1a tubercle bacilli ^h						
23	3' <i>cfp32</i> ³¹¹	1279 ⁱ	311G→A	Synonymous	23	22
29	RD13 ¹⁷⁴	449 ^e	174G→A	Synonymous	29	21
16	<i>PPE55</i> ²¹⁵⁴	723	7321A→G	Synonymous	16 ^j	23
16	<i>PPE55</i> ²¹⁴⁸	725	7327A→G	Synonymous	20 ^j	19
<i>M. africanum</i> subtype Ia, 18	<i>hsp65</i> ^{540k}	180	540C→G	Synonymous	18	107
	<i>rpoB</i> ¹¹⁶³	388	1163C→T	Nonsynonymous	16 ^l	109
	<i>rpoB</i> ¹⁰⁴⁹	350	1049C→T	Nonsynonymous	5 ^m	46
Dassie bacillus, 4	3' <i>cfp32</i> ²²⁴		224A→G	Intergenic ⁱ	4	42
	Rv0911 ³⁸⁹	130	389G→deleted	Premature stop ⁿ	4	121
<i>M. africanum</i> subtype Ia, 18; dassie bacillus, 4	Rv1510 ¹¹²⁹	377	844G→A	Nonsynonymous	22	103
Oryx bacillus, 2	<i>gyrB</i> ^{oryx}	329 ^d	1113G→A	Synonymous	2	123
		442 ^d	1450G→T	Nonsynonymous		
	TbD1 ¹⁷¹		171C→G	Intergenic ^g	2	30
	<i>PPE55</i> ²¹⁶³	720	7312T→G	Synonymous	6 ^j	33
	<i>PPE55</i> ²¹⁶²	721	7313C→T	Nonsynonymous	6 ^j	33
<i>M. microti</i> , 10	16S rRNA ¹²³⁴		1234G→A	Intragenic	10	115
	RD13 ⁶⁷		67G→A	Intergenic ^e	10	40
<i>M. pinnipedii</i> , 7	16S rRNA ¹²⁴⁹		1249T→C	Intragenic	7	118
	RD13 ²²⁸	431 ^e	228G→A	Synonymous	7	43
<i>M. microti</i> , 10; <i>M. pinnipedii</i> , 7	RD13 ³⁸⁰	381 ^e	380A→G	Nonsynonymous	17	33

^a Nucleotide positions are given relative to the gene sequences of *M. tuberculosis* H37Rv (accession no. AL123456 [10]) except for *mmpL6*¹⁸⁷⁹, which is numbered according to the gene sequence of *M. bovis* strain AF2122/97 (accession no. BX248333 [25]).

^b SNPs in the *aroA*, 3' *cfp32*, RD13, and TbD1 PCR fragments are numbered by the nucleotide site within each respective amplicon; the provided *gyrB*^Δ SNP positions are relative to previously established nomenclature (57).

^c Does not include strains that failed to PCR amplify due to deletion of the respective loci.

^d According to the translated *gyrB* (Rv0005) gene sequence of *M. tuberculosis* H37Rv (accession no. AL123456).

^e RD13¹⁷⁴, RD13²²⁸, RD13²⁵⁵, and RD13³⁸⁰ are located within Rv1257c; RD13⁶⁷ is located intermediate of Rv1256c and Rv1257c.

^f *aroA*⁹¹ and *aroA*¹¹⁷ are located intermediate of Rv3226c and *aroA* (Rv3227); *aroA*²⁸⁵ is located within *aroA*.

^g Located intermediate of *M. bovis mb1581* (Rv1556) and *mmpL6* (*mb1582*) of the TbD1 locus.

^h Includes *M. africanum* subtype Ia, dassie bacillus, oryx bacillus, *M. microti*, *M. pinnipedii*, *M. caprae*, *M. bovis*, and *M. bovis* BCG.

ⁱ 3' *cfp32*³¹¹ is located within Rv0578c; 3' *cfp32*²²⁴ is located intermediate of *cfp32* (Rv0577) and Rv0578c.

^j Includes "M. canettii" strains (*n* = 4) with convergent SNPs.

^k Located 141 bp upstream of the "position 631" "M. canettii" *hsp65* SNP (actually nucleotide 399, codon 133).

^l *M. africanum* subtype Ia strains ATCC 25420, ATCC 35711, 1255/93, 1457/93, 1565/93, 1567/93, 17316, 17902, Percy18, Percy57, Percy61, Percy122b, Percy171, Percy199, Percy203, and Percy207.

^m *M. africanum* subtype Ia strains ATCC 25420, ATCC 35711, Percy57, Percy171, and Percy207.

ⁿ A premature termination signal is predicted to occur at the fourth subsequent codon as a result of frameshift.

^o Strain 96-46 possessed additional SNPs at this locus, as described in the text.

^p "M. canettii" strains 217-94, 17727, 2002-937, and 2002-1000.

^q "Modern" PGG2 *M. tuberculosis* strains CDC1551, Cb3.3, 97-279, 97-488, 2001-1255, CA-56, tb21, and AHTN13475.

TABLE 5. Description of various MTC species-, lineage-, and strain-specific LSPs

Locus	MTC species	MTC strain or no. examined (if more than 1)	Location ^a of deletion	ORF(s) affected ^b	Nucleotide sequence at the junction ^c
RD7 ^{tb}	“Ancestral” PGG1b <i>M. tuberculosis</i>	CA-74	2214.298–2217.261 [2,964 bp]	Rv1970 to Rv1973	GCAGCTGCCGGATGTCAACAACATCCAG CCGAACCTCGCGGGTTCGGGTTGCCGC GACGCCGCGGTTGCCCGTGCGGAA TCCGTGCGGGCCGCGACCGACG
RD236a	“Ancestral” PGG1b <i>M. tuberculosis</i>	Percy211	4056.946–4058.397 [1,452 bp]	<i>epHA</i> (Rv3617)	GTCATTATCTCCGATCATGTTTGATAAGG CGACGAAAACCCGACGATGGAACATA TCGGCGAGTTCACCCGACCGGTTT CACCGCGGCCTGAACCTGGT
N-RD25 ^{tbA}	“Modern” PGG1b <i>M. tuberculosis</i>	8 of 13 ^d	4189.258–4190.367 [1,110 bp]	Rv3737 to <i>PPE67</i> (Rv3739c)	CCGACGACGGCCGACGACGTGGACGCC GGCTATCGCGGATTGGCCCGCCGC AGTGAGCTCTTGGGCTGCCGAAGC ATACTCCGACTCAGCGATGTCC
N-RD25 ^{tbB}	“Modern” PGG1b <i>M. tuberculosis</i>	2 of 13 ^e	4189.234–4190.191 [951 bp]	Rv3737 to <i>PPE66</i> (Rv3738c)	AGCGTCCGCTGGAGCCGACGCGCC GACGACTCGCGGATCGGCTATCGCG GCATCGCCGCCAGCGCGCTGGCATA TGCCGTGGTCATAACCTCATGC
RD1 ^{tbB}	“Modern” PGG1b <i>M. tuberculosis</i>	3 of 13 ^f	4368.662–4368.718 [57 bp]	Rv3879c	AACGCGCGCCGCTCCACGCGGGTAC CGCTCGGCGCGGCGCCCGCCGC ATGGGCAGGCCCGACTGCGTCCCC CCGCCGCATGCTGGCCCGGCAC
MiD3 ^g	<i>M. pinnipedii</i>	7 of 7	3741.143–3755.779 [14,636 bp]	<i>PE-PGRS50</i> (Rv3345c) to <i>IS1561'</i> (Rv3349c)	GTGGGCGCGGTGCCGTTGCCCGCCG GAACCGTTTCAGGGCCGGTTCGCCCT CAGCCGCTAAACACGCCGACCAAG ATCAACGAGCTACCTGCCCGGTC

^a Relative to the *M. tuberculosis* strain H37Rv genome sequence (accession no. AL123456 [10]); given in kb.

^b ORF, open reading frame; affected by truncation or deletion.

^c Sequences in boldface indicate the 5' (left-hand) portion of the junction sequence.

^d *M. tuberculosis* strains 97-742, 97-803, 97-818, 97-1177, 97-1289, 97-1438, Percy7, and Percy91.

^e *M. tuberculosis* strains 13876 and Percy209.

^f *M. tuberculosis* strains 97-803, 97-1177, and 97-1438.

^g Identical to the *M. microti* MiD3 LSP ($n = 10$).

the absence of repeat sequences at the junction sites of all newly identified and defined RD loci in this study excludes insertion sequence homologous recombination-mediated deletion events as their cause (34).

(i) **PGG1 *M. tuberculosis***. During the course of this study, unexpected PCR amplification failures were observed from two “ancestral” *M. tuberculosis* strains. Isolate CA-74 was found to have a novel 2,964-bp deletion mapped within the RD7 locus (RD7^{tb}) (Table 5), while strain Percy211 was found to have the previously noted 1,452-bp RD236a deletion overlapping the RD8 locus (Table 5) (82). The latter deletion links Percy211, which is of European origin, with strains isolated in California (82).

With respect to “modern” PGG1b *M. tuberculosis* strains, 10 of 13 strains failed to amplify for N-RD25^{bovis/cap}, a region that overlaps N-RD25^{das} (50), while three W-Beijing isolates were wild type for this locus. We were able to bridge the putative deletion locus and delineate the boundaries of two apparently independent parallel RDs (Table 5). Of these 10 strains, the N-RD25^{tbA} LSP ($n = 8$) resulted in the loss of 1,110 bp of sequence, while the N-RD25^{tbB} LSP ($n = 2$) is the result of a 951-bp genomic lesion. These deletions align very closely, with the 5' end of the N-RD25^{tbA} LSP located just 15 bp downstream of the N-RD25^{tbB} LSP 5' breakpoint. The N-RD25^{tbA} deletion was also recently described by Rajakumar et al. (63) in a sizable cluster of strains causing an outbreak of TB in Leicester, United Kingdom. Given that the N-RD25^{tbA} *M. tuberculosis* isolates in our collection were of both African and

European origin, the combined data suggest that N-RD25^{tbA} may be at the node of an important branch of PGG1b *M. tuberculosis* and a useful phylogenetic marker (Fig. 2, note circles 4 and 6). In addition, we found that three of the strains of European origin also possessed a 57-bp minideletion within the RD1 gene Rv3879c (RD1^{tbB}), further segregating the N-RD25^{tbA} *M. tuberculosis* subset of isolates (Table 5; Fig. 2, note circle 5) (38). Interestingly, fingerprinting data from the Leicester strains indicate an Indian/south Asian origin for the N-RD25^{tbA} isolates (19, 63, 88).

(ii) **PGG2/PGG3 *M. tuberculosis***. In terms of PGG2 *M. tuberculosis*, strain CA-56 failed to amplify for Rv3879c as a result of a deletion known to overlap RD1 (RD1^{tbA}) and has been characterized elsewhere (Table 1) (62). In addition, four strains failed to amplify for the *IS1561'* element and a nearby *PPE55* fragment, both of which are deleted as a part of the MiD3 LSP in *M. microti* and *M. pinnipedii* (Tables 1 and 3). The characterization of this interesting RD locus, and the scope of this genotype's distribution, will be described elsewhere.

Lastly, we identified a unique intergenic SNP in the *aroA* fragment (*aroA*¹¹⁷) from 8 of 19 PGG2 *M. tuberculosis* isolates, including the Tennessee (CDC1551) and New York City (Cb3.3) outbreak strains (21, 22), but not the *IS1561'*-deleted ($n = 4$) and Uganda genotype ($n = 4$) PGG2 *M. tuberculosis* strains nor any other MTC isolate (Table 4). That the *aroA*¹¹⁷ polymorphism links the origin of these strains back to a common PGG2 progenitor is supported by a similar scenario de-

picted by Gutacker et al. (30), whose synonymous SNP analyses positioned strains Cb3.3 and CDC1551 in separate major clusters along a single PGG2 branch of the *M. tuberculosis* evolutionary tree (clusters IV and V, respectively) (Barry Kreiswirth, personal communication) (Fig. 2, note circle 8). As a final point of interest, the *M. tuberculosis* Uganda genotype strains were unremarkable within the context of this investigation, although unique RD loci, but not SNPs, have been identified previously in a few strains (53).

***M. africanum*.** *M. africanum* is predominantly isolated in Africa, and in certain regions it is thought to represent a significant proportion of the isolates obtained from patients with pulmonary TB (57). Reports on the sporadic isolation of *M. africanum* in Europe and the United States (15, 42) have also been made, including one outbreak of multidrug-resistant (MDR) *M. africanum* (32, 87). In the pregenomic era, traditional mycobacteriological testing segregated *M. africanum* into two subtypes based upon their phenotypic similarity to either *M. bovis* or *M. tuberculosis* and their respective corresponding geographic origin in west (subtype I) or east (subtype II) Africa (33, 35, 57). With the advent of new genetic marker-based definitions for each of the MTC species, true *M. africanum* (subtype I) strains have been divided into two distinct genotypic variants while most *M. africanum* subtype II strains have proven to be phenotypic variants of *M. tuberculosis*. As a result, calls have been made to discard the subtype II designation (7, 38, 75).

In this study, all previously designated *M. africanum* subtype II isolates that had been predetermined to retain RD9 and RD12 intact were considered *M. tuberculosis*. We noted that RD7-, RD8-, and RD10-deleted *M. africanum* subtype I strains were PGG1a, while *M. africanum* subtype I isolates with RD7, RD8, and RD10 intact were PGG1b (Tables 1 to 3). To ease discussion, we refer to these separate MTC lineages as *M. africanum* subtypes Ia and Ib, respectively. Both subtype I branches of *M. africanum* bore the same *gyrB*^{af} sequence (Tables 2) (57). In addition, a recent paper by Mostowy et al. described several novel RD loci within *M. africanum* organisms. Among these were RD711 and RD713, which were proposed to be *M. africanum* subtype Ib specific, as well as RD701 and RD702, which were validated in a larger evaluation of strains to be *M. africanum* subtype Ia defining (53). A reanalysis of these loci was undertaken as part of the current study.

With respect to distinctive genetic markers of *M. africanum* subtype Ib, RD713 was found to be a complex locus, because the expected amplification target contains sequence that is deleted from PGG2 and PGG3 *M. tuberculosis* and is partially deleted from *M. africanum* subtype Ib, but it remains intact in "*M. canettii*" and a proportion of PGG1b *M. tuberculosis* (our unpublished data). This locus also overlaps the 3' end of the RD7 locus and the 5' end of the RD2 locus. In PCR analysis, products were observed from each PGG2/PGG3 *M. tuberculosis* (4,248 bp) strain and all *M. africanum* subtype Ib (2,798 bp) strains that were evaluated (Table 3). In contrast, every "*M. canettii*" and PGG1b *M. tuberculosis* strain that was tested failed to amplify, most likely because the larger size of the anticipated RD713 amplicon is prohibitive under the PCR conditions utilized (a full-locus characterization will be described elsewhere). Each test MTC species on the *M. africanum* subtype Ia→*M. bovis* evolutionary track also failed to

amplify, because they are deleted in RD7. As a result of the extra genetic sequence at this location in *M. africanum* subtype Ib, the aforementioned data indicate that the genomic content lost with RD7 is not accurately reflected by the corresponding locus in the full *M. tuberculosis* H37Rv genome sequence. As was the case with RD713, sequence analysis of the TbD1 and *aroA* PCR fragments uncovered SNPs (TbD1¹⁹⁷ and *aroA*²⁸⁵, respectively) that were unique to all *M. africanum* subtype Ib strains in our collection and are the first intraspecific SNPs identified for this species (Table 4; Fig. 2, note circle 11). On the other hand, the expected RD711 942-bp amplicon was found in only seven *M. africanum* subtype Ib strains, while the remaining five *M. africanum* subtype Ib strains, as well as the other MTC isolates that were tested, bore the wild-type 2,885-bp amplicon (Table 3). Hence, rather than being a species-defining marker, the RD711 deletion appears to define a sublineage within *M. africanum* subtype Ib and likely occurred subsequent to RD713 and after the original derivation of this species (Fig. 2, note circle 12).

In this study, the previously identified 6-bp microdeletion in *pks15/1* (11, 48), as well as novel SNPs at *PPE55*²¹⁴⁸, *PPE55*²¹⁵⁴, RD13¹⁷⁴, and 3' *cfp32*³¹¹, were also investigated. The presence of interspecific polymorphisms at these loci appear to sharply divide the MTC isolates along the same line as the RD7, RD8, RD10, and *katG*²⁰³ polymorphisms that segregate "*M. canettii*," *M. tuberculosis*, and *M. africanum* subtype Ib from *M. africanum* subtype Ia and the remaining MTC species (Tables 1 to 4; Fig. 2, note circle 13; also see Fig. S1A in the supplemental material).

With respect to distinctive genetic markers for *M. africanum* subtype Ia, amplification for RD701 and RD702 yielded PCR fragments consistently shorter than those generated from the other MTC species tested (340 versus 2,081 bp and 732 versus 2,101 bp, respectively), thereby confirming the *M. africanum* subtype Ia specificity of these LSPs (Table 3). We also identified a novel SNP in the Telenti *hsp65* fragment (*hsp65*⁵⁴⁰) that was present in all *M. africanum* subtype Ia isolates but no other MTC strains that were assessed (Table 4). This SNP is the first to be reported that is specific for *M. africanum* subtype Ia, and it is conveniently situated within an amplicon frequently targeted in PCR-RFLP-based mycobacterial identification protocols (Fig. 2, note circle 15) (27, 80). A second SNP was identified in Rv1510 (Rv1510¹¹²⁹) of all *M. africanum* subtype Ia strains and, excepting the dassie bacillus, was absent from the remaining MTC species (Table 4; Fig. 2, note circle 14). Further SNPs were also found in the *M. africanum* subtype Ia *rpoB* gene (*rpoB*¹⁰⁴⁹ and *rpoB*¹¹⁶³) (Table 4). Although neither of the *rpoB* SNPs was found in non-*M. africanum* subtype Ia MTC organisms, these SNPs were not true *M. africanum* subtype Ia markers but defined secondary and tertiary branches within this species; whereas *rpoB*¹¹⁶³ was present in 16 of the 18 strains screened, the previously described *rpoB*¹⁰⁴⁹ SNP (46) was present in 5 *rpoB*¹¹⁶³-positive isolates, including the *M. africanum* subtype Ia type strain ATCC 25420 (Table 4; Fig. 2, note circles 16 and 17). These SNPs are not believed to be associated with rifampin resistance, since they are outside of the *rpoB* 81-bp rifampin resistance-determining region and because strain ATCC 25420 has previously been reported to be rifampin sensitive (65).

The dassie bacillus. The dassie bacillus was first identified in the late 1950s as an acid-fast microbe causing pulmonary TB in hyraxes (*Procavia capensis*) (74, 89, 90). The hyrax, or “dassie,” as it is locally known, is a small rabbit-like mammal commonly found in South Africa and the Middle East. The dassie bacillus was rediscovered in the late 1980s when it caused disease in a colony of captive South African hyraxes at the Perth Zoo, and it has also since been isolated from a suricat housed in a Swedish zoo (13, 51). Based upon certain overlapping laboratory characteristics, the dassie bacillus was originally proposed to be a highly attenuated variant of *M. microti*, and even though several additional biological features are now known to set it apart from *M. microti*, this perception has persisted to date (23). It should be noted, however, that in the early descriptions of the organism it was recognized that the dassie bacillus could also be a new species of mycobacteria (90).

In our PCR analyses, all dassie bacillus strains failed to amplify for the RD7, RD8, RD9, RD10, and RD713 regions but produced PCR fragments for the RD4, RD12, RD13, MiD3, and TbD1 LSPs, as well as full-length amplicons for RD701, RD702, and RD711 (Tables 1 and 3). A recent paper describing the gross genetic content of the dassie bacillus reported similar findings with respect to all the known *M. microti*-specific genomic deletions (50). We further noted in this study a location-specific failure to amplify from the RD1^{BCG} locus as well as the N-RD25^{bovis/cap} region, which indicated to us that the dassie bacillus also possessed its own unique set of LSPs, thereby positioning this organism apart from all other MTC species (Tables 1 and 3). Mostowy et al. (50) independently described these loci in addition to other dassie bacillus intraspecific LSPs. Notably, the RD1^{das} LSP overlaps the boundaries of RD1^{BCG}, while N-RD25^{das} is a unique polymorphism that overlaps the N-RD25^{bovis/cap} and N-RD25^{tbA/B} deletions of other MTC species and sublineages (Fig. 2, note circle 18).

The dassie bacillus was thoroughly evaluated for many other known and newly identified interspecific SNPs. The dassie bacillus was PGG1a and evidenced all the mutations associated with the polymorphic juncture that separates *M. africanum* subtype Ia from Ib, as noted above (Tables 1 to 4). We also evaluated for the previously identified *mmpL6*⁵⁵¹ SNP (codon 551; AAG→AAC) (Fig. 2, note circle 19) (7). The dassie bacillus clustered with “*M. canettii*,” “ancient” *M. tuberculosis*, and *M. africanum* subtypes Ia and Ib in being wild type rather than grouping together with *M. microti* and the remaining MTC species for this interspecific SNP (Table 2). None of the other previously known or novel MTC species- or sublineage-specific SNPs evaluated herein were simultaneously present in the dassie bacillus, with two exceptions. First, the dassie bacillus *gyrB*^Δ sequence was homologous to *gyrB*^{aftr} (as is that of *M. pinnipedii* [see below]) and contradicts the reported *M. africanum* subtype I specificity of this sequence (Table 2) (57). Of course, this point may matter little clinically, given the rarity of identification of the dassie bacillus and *M. pinnipedii* outside their normal animal hosts. The second shared point mutation, the interspecific Rv1510¹¹²⁹ SNP, is described below.

In this work, many SNPs were identified fortuitously, but in the case of the dassie bacillus we actively searched for novel intraspecific nucleotide differences by sequence analysis of PCR products generated from numerous loci. Collectively, ~19 kb of DNA was sequenced from strain 68/7171. The 16S

rRNA gene (1,431 bp), the 16S-23S internal transcribed spacer region, and the *hsp65* Telenti fragment were among the investigated loci that matched 100% to those of *M. tuberculosis* H37Rv. In fact, just two polymorphisms that were unique to the dassie bacillus strains and one shared SNP were identified. The substitution rate for the sequences evaluated was calculated to be ~0.016% and is in line with previous estimates on the pace of nucleotide change within the MTC (77). The first mutation was located in the intergenic region of *cfp32* and Rv0578c (3' *cfp32*²²⁴) (Table 4). None of the other MTC isolates tested carried this SNP. We also identified a single-nucleotide deletion (SND) in the Rv0911 gene (Rv0911³⁸⁹) (Table 4; Fig. 2, note circle 18). The resulting shift in codon frame leads to a premature termination signal and predicts a truncated protein. Interestingly, the Rv0911 homologue in *Mycobacterium leprae* (ML2112) has also undergone severe gene decay. All MTC isolates tested amplified for Rv0911, but only the dassie bacillus isolates had the Rv0911³⁸⁹ SND and so may be a contributing factor to the remarkable attenuation of the dassie bacillus in comparative infection experiments of laboratory animals (13, 89). Lastly, Rv1510¹¹²⁹ was found to be an SNP common to all dassie bacillus and all *M. africanum* subtype Ia strains evaluated (Table 4; Fig. 2, note circle 14); no other MTC species carried this SNP. Therefore, although the sum phenotypic, biochemical, and genetic data support that the dassie bacillus is a unique MTC species and not a strain of *M. microti*, the SNP data also appear to align the dassie bacillus phylogenetically with *M. africanum* subtype Ia. Indeed, the SNP data fit a model wherein these species share a common ancestor that had itself diverged earlier from the MTC evolutionary tree prior to the derivation of the current set of separate species-specific polymorphisms in current *M. africanum* subtype Ia and dassie bacilli (Fig. 2). Of course, whether or not the Rv1510¹¹²⁹ genotype shared by these organisms is merely happenstance remains to be determined by further genetic comparisons. However, our interpretation of the data is supported by a previous genome-wide synonymous SNP analysis that also placed *M. africanum* subtype Ia and the dassie bacillus very close together in their evolutionary history (30). Based upon the complement of data collected to date, we believe that the dassie bacillus will eventually be elevated to full species status within the current context of the existing MTC nomenclature.

The oryx bacillus. TB in oryxes (*Oryx leucorox*; an antelope species) has previously been reported (28, 47, 84), but to date the oryx bacillus has been but mentioned in passing as a phenotypically divergent subtype of *M. bovis* (7, 42, 84). The two isolates evaluated in this study were retrieved from oryxes in a wild park in Saudi Arabia and a zoo in The Netherlands. These strains were notable in that they carry about 18 or 19 copies of IS6110, which is unusual for *M. bovis*, and they share a unique spoligotype atypical of true *M. bovis* and *M. caprae* isolates (43, 83, 84).

In our analyses, the oryx bacillus was further distinguished because it has preserved the regions RD4, RD12, RD13, and N-RD25^{bovis/cap} intact, and it does not bear the *gyrB*^{bovis}, *gyrB*^{cap}, *pncA*⁵⁷, and *oxyR*²⁸⁵ point mutations present in *M. caprae* and/or *M. bovis* (Tables 1 to 3) (7, 38, 57, 72, 76). In addition, we found SNP markers in *gyrB*^Δ (*gyrB*^{oryx}) and TbD1¹⁷¹, as well as tandem SNPs in *PPE55*²¹⁶² and *PPE55*²¹⁶³, for the oryx

bacilli that differentiated them from the other MTC isolates in this study (Table 4; Fig. 2, note circle 20; Fig. S1A in the supplemental material). Note that screens for *gyrB*^{oryx} and *gyrB*^{can}, such as the novel PCR-RFLP strategies utilized herein, may be easily incorporated into other *gyrB*^Δ MTC species identification protocols (57, 66). A strain-specific nonsynonymous SNP in the CA-24 isolate was also found at *katG*⁴⁷⁰ (codon 470, position 1405G→C), suggesting that these bacilli were not immediately clonal. In contrast to the above, and as with *M. caprae* and *M. bovis*, the oryx bacillus was RD7, RD8, RD9, and RD10 deleted and possessed all the PGG1a-associated polymorphisms, as well as the *mmpL6*⁵⁵¹ SNP (Tables 1 to 3) (7). The oryx bacillus did not possess any intraspecific polymorphisms described herein for other MTC species (Tables 1 to 3). Similar observations with respect to RD loci were also made in a recent independent analysis of the gross genomic content of the oryx bacillus (52).

The combined data therefore indicate that the oryx bacillus is positioned higher than *M. caprae* and *M. bovis* but follows *M. africanum* subtype Ia and the dassie bacillus along the evolutionary continuum of the MTC. Of course, it is difficult to draw conclusions from just two strains. However, as a potentially important phylogenetic linker species, the oryx bacillus may well meet the criteria for unique MTC species status and probably should not be grouped with *M. bovis* simply as a matter of taxonomic convenience. Hopefully increased veterinary awareness of this pathogen will result in the identification of additional strains.

***M. microti*.** The natural hosts of *M. microti* are bank voles and other small rodents of European origin (23). However, rare cases have also been identified in a variety of mammalian species, including llamas and humans (86). The strains studied herein were of vole, human, llama, or ferret origin and included ancestral *M. microti* strains, as indicated by prior spoligotyping results (42, 86).

Consistent with previous reports and the available genome sequence, each *M. microti* isolate that was tested possessed the PGG1a-associated interspecific polymorphisms of MTC species, had allele AAC at *mmpL6*⁵⁵¹, had *gyrB*^{mic}, and contained the MiD3 LSP (Tables 1 to 3; Fig. 2, note circles 21 and 22) (6, 7, 38, 57). In this work we identified and evaluated additional *M. microti*-specific SNPs within the RD13 locus (RD13⁶⁷ and RD13³⁸⁰) and the 16S rRNA gene (16S rRNA¹²³⁴) (Table 4). With the exception of MiD3 and RD13³⁸⁰ in *M. pinnipedii*, none of the other MTC isolates shared any of the above *M. microti* polymorphisms (Tables 1 to 4).

***M. pinnipedii*.** Formerly called the seal bacillus, *M. pinnipedii* is the cause of endemic TB in various seal species of southern hemispheric origin (12, 93). Recognition of TB in captive seals dates to the early 20th century, and there is at least one case reported in the literature of transmission of *M. pinnipedii* to a seal trainer who worked with infected captive seals at a marine park (14). Prior to its proposal as a unique MTC species, the seal bacillus was sometimes considered a subtype of *M. bovis* despite evidence to the contrary (7, 42).

In our investigations, *M. pinnipedii* exhibited all the associated polymorphisms of PGG1a MTC organisms, possessed the *mmpL6*⁵⁵¹ mutation, and was positive for *gyrB*^{af}, but it did not bear any of the *M. bovis*-specific polymorphisms (or other MTC intraspecific mutations), consistent with previous studies

(Tables 1 to 4) (7, 51, 93). All *M. pinnipedii* strains tested also possessed an intraspecific SNP that was identified in the 16S rRNA gene (16S rRNA¹²⁴⁹) (12) as well as a novel intraspecific SNP within the RD13 locus (RD13²²⁸) (Table 4; Fig. 2, note circle 23). A second interspecific SNP in the RD13 locus (RD13³⁸⁰) was present in both *M. pinnipedii* and *M. microti* and was coincident with a failure to PCR amplify from the IS1561' and PPE55 coloci in each of these species (Tables 1, 3, and 4; Fig. 2, note circle 21). The breakpoints of this apparent *M. pinnipedii* LSP were determined and, surprisingly, were found to be identical to those of the previously defined *M. microti* MiD3 14,636-bp deletion (Table 5) (6). As a result, despite otherwise clear biological differences between *M. pinnipedii* and *M. microti*, the RD13³⁸⁰ SNP and MiD3 data are consistent with the hypothesis that these species share a common progenitor that had itself emerged at some point earlier in MTC evolutionary history. If so, then the other *M. pinnipedii*- or *M. microti*-specific genetic differences most likely arose as these species became established in their respective niches. It just so happens that *M. microti* isolates were intermingled with *M. pinnipedii* strains in the synonymous SNP study of Gutacker et al. (30). However, at this time it is not known if any of the other *M. microti* LSPs are present in *M. pinnipedii* as well and is an area worth investigating further; at least RD2^{seal} is known to be present in *M. pinnipedii* but not *M. microti* (7, 48). Lastly, the finding of two different 16S rRNA SNPs in *M. microti* and *M. pinnipedii* contradicts the long-held view that this gene is nonpolymorphic within the MTC. This position was often based upon sequence analysis of just the 5' end of the 16S rRNA gene (81), whereas the identified SNPs were found towards the 3' end.

***M. caprae* and *M. bovis*.** *M. bovis* and *M. caprae* are named for their preferential cow and goat hosts, respectively. However, these MTC species are also known to cause disease in a wide range of domesticated and wild animals as well as humans (25, 43, 61, 84). In many countries, TB in farm animals causes great agricultural loss, and feral reservoirs constitute important sources of infection that undercut control measures (25). Humans are most often infected following transmission of *M. caprae* or *M. bovis* from infected livestock (43, 84), and so this remains a serious public health concern in some countries. In addition, human-to-human transmission of *M. bovis* and the emergence of MDR *M. bovis* have also been documented, raising infection control concerns in health care settings (5, 26, 60).

In this work, we confirm that *M. bovis* and *M. caprae* share many genotypic features in common. These include the interspecific polymorphisms identified at the axis of *M. africanum* subtype Ia and Ib differentiation, the *mmpL6*⁵⁵¹ AAC codon, the *oxyR*²⁸⁵ point mutation, and coincident deletion of the RD12 and RD13 loci and joint *gyrB*^Δ SNPs (Tables 1 to 3) (7, 38, 76). We further found that *M. bovis* and *M. caprae* also share the N-RD25^{bovis/cap} LSP (Table 3) (68); sequence analysis of the N-RD25^{bovis/cap} bridge amplicon from *M. caprae* proved that the locus was identical to that of *M. bovis* strains, meaning this deletion most likely arose in a common ancestor (Fig. 2, note circle 24). On the other hand, it is now known that *M. caprae* can be genetically differentiated from *M. bovis* on the basis of positive amplification for the RD4 locus as well as SNP analysis of *gyrB*^Δ (*gyrB*^{cap} and *gyrB*^{bovis}, respectively) and

*pncA*¹⁶⁹ (Tables 1 and 2; Fig. 2, note circles 25 and 26) (38). Lastly, it should be mentioned that all *M. bovis* BCG strains in the collection were in possession of every typical *M. bovis* polymorphism that was evaluated, as would be expected, as well as the intraspecific RD1^{BCG} deletion (Fig. 2, note circle 27). These data were corroborated by a very recent publication by Mostowy et al. (52) that extensively evaluated the gross genomic content of *M. bovis* and phylogenetically related MTC strains.

DISCUSSION

In this study, we evaluated and correlated many of the most relevant phylogenetic LSPs and SNPs that are presently known, validated the significance of additional polymorphisms that had been previously noted but remained unproven, and identified new species- or lineage-restricted genetic markers, including some in certain MTC species for which specific confirmed SNPs had been lacking until now. Of note, we saw 100% consistency between all finalized deletion-based MTC species identifications and all SNP-derived MTC speciations. A combined integrated diagrammatic phylogenetic interpretation of the data is presented in Fig. 2. Highlights include (i) the first specific SNP markers reported for *M. africanum* subtype Ib, *M. africanum* subtype Ia, the dassie bacillus, and the oryx bacillus, as well as newfound additions to the few "*M. canettii*", *M. microti*-, and *M. pinnipedii*-restricted SNPs currently known, (ii) the LSP- or SNP-based intrasubspecies segregation of different "*M. canettii*", "modern" PGG1b *M. tuberculosis*, "modern" PGG2 *M. tuberculosis*, *M. africanum* subtype Ib, and *M. africanum* subtype Ia subgroups, and (iii) the genetic linkage of *M. africanum* subtype Ia and the dassie bacillus, as well as *M. microti* with *M. pinnipedii*, to shared common progenitors. Overall, our results confirm previous models of the evolution of the MTC (7, 51), further resolve certain phylogenetic interrelationships within the MTC, and expand the genetic definitions of many MTC species and/or their lineages. Of course, the distribution of the newly identified and proposed intraspecific polymorphisms must be validated in a greater array of isolates. Nonetheless, with the identification of widely distributed (if not restricted) mutations, phylogenetic links may be established between MTC species or otherwise distinct strains that are no longer obvious from fingerprint data and so may be used to create long-range temporal genealogies. Lastly, there are many reasons for the differentiation of MTC organisms to the subspecies level, such as the enabling of appropriate patient management and public health investigations, as well as for epidemiological purposes. In so being, our data also now provide more options for the creation of algorithms for the differentiation of MTC organisms by genetic criteria as well as offer clues to the molecular bases of pathobiological variations within the complex.

On this point, *M. bovis* is regarded as being less virulent for humans than *M. tuberculosis*. Nonetheless, it has been proposed that the extensive gene loss in *M. bovis* (and possibly the remaining non-*M. tuberculosis* MTC) actually represents a fine-tuning of its virulence as it evolved and expanded to new host niches (25). This may explain why *M. bovis* has a broader host range than any other MTC member and why it has much greater virulence than *M. tuberculosis* in the rabbit model (25).

However, despite the fact that LSPs are the most obvious genomic distinction among MTC members, nonsynonymous SNPs and intergenic polymorphisms have also been associated with certain unique phenotypic differences (25) and may be equally important determinants of variation. One notable polymorphism that we did not include in the present evaluation is the *narGHJI* gene cluster -215 C→T promoter polymorphism that is present in all "modern" *M. tuberculosis* strains that have been reported (40, 78). This mutation is associated with increased gene expression, elevated aerobic nitrate reductase activity, and heightened strain virulence in experimental models (79, 92). As such, it has been hypothesized that this SNP contributed to the evolutionary success of the "modern" *M. tuberculosis* strains that possess it (40). At present, it is not known how the point mutations uncovered in noncoding regions in this study affect the expression and regulation of the associated genes.

It was further interesting that overlapping deletions and/or an elevated nucleotide polymorphism status were often related to PE and PPE genes (e.g., N-RD25, MiD3, and *PPE55*) in the present evaluation. These data are consistent with previous evidence that these genes are hypervariable (9). With respect to N-RD25, the meaning of distinct overlapping deletions in two "modern" PGG1b *M. tuberculosis* sublineages, as well as in the dassie bacillus, *M. caprae*, and *M. bovis*, is not known but may reflect either the deletion of highly immunogenic proteins and/or the editing of nonessential genes out of the genome. Whether any of the nonsynonymous SNPs and genomic deletions identified in this study truly impact upon virulence and transmissibility remains to be determined through functional genomics studies.

In this paper, we noted that a seemingly elevated ratio of fixed interspecific polymorphisms appears to have arisen in the transition of *M. africanum* subtype Ib to *M. africanum* subtype Ia along the MTC phylogenetic tree (Fig. 2, note circle 13). It remains to be determined whether this observation is the result of the breakthrough transmission of a highly mutated clone or whether the genetic changes arose sequentially in evolutionary precursors of the *M. africanum* subtype Ia→*M. bovis* phylogeny and, in so being, may reflect a lengthened temporal distance between the *M. africanum* subtype Ia and Ib species. It should be noted that by all accounts RD7, RD8, and RD10 appear to follow an all-or-none rule in that no transitional MTC strains have yet been reported, supporting the first possibility. Since the specific LSPs and SNPs of *M. africanum* subtype Ib are not found in *M. africanum* subtype Ia and vice versa, the data also indicate that, although the *M. africanum* species share a common RD9-deleted ancestor (Fig. 2, note circle 10), *M. africanum* subtype Ia does not appear to be a direct descendant of *M. africanum* subtype Ib. Rather, the clonal speciation and expansion events of the current subtypes Ia and Ib of *M. africanum* appear more likely to have occurred from separate nodes along the MTC phylogenetic tree.

Overall, the genetic data from this and other studies support the hypothesis that an ancestral form of *M. caprae* gave rise to *M. bovis* (7, 51). Given that goat domestication predates the farming of cattle (44), perhaps the evolution of *M. bovis* from *M. caprae* is linked to human history. As an additional point of interest, *M. bovis* and its BCG vaccine derivative are innately resistant to the important anti-TB drug pyrazinamide (PZA) as

a result of the *pncA*¹⁶⁹ SNP, and PZA resistance is a phenotypic feature often used to differentiate *M. bovis* isolates from the other classical MTC members (43, 72). Prior to the introduction of molecular tools for MTC species identification and the recognition of *M. caprae* as a unique MTC species, numerical analysis of biochemical properties may have resulted in PZA-sensitive *M. caprae* being identified as *M. africanum*. Indeed, this may go to explain sporadic published reports of “*M. africanum*” infections of livestock and wild animals (1, 71). Interestingly, examples in the literature also exist for PZA-resistant *M. africanum* whose non-*M. bovis* species identity was clarified only after genotypic testing was performed (26, 32, 87; Andrea Gibson, personal communication).

All of the MTC species are known to infect humans (excepting the dassie and oryx bacilli), albeit rarely in some cases (14, 38, 86). However, how the various MTC species jumped species barriers to become firmly established in such diverse hosts as humans, hyraxes, oryxes, seals, voles, goats, and cattle remains an open question. *M. tuberculosis* is clearly infectious to a range of mammals, given the elevated rates of zoonotic transmission of this microbe from humans to animals in captivity (14, 49, 59). But presuming humans were minimally the preliminary source of tubercle bacillus infection, what conditions for transmission and infection were required in order for non-*M. tuberculosis* MTC clones to emerge and to show such a range of distinct host preferences? It is very attractive to speculate that cofarming of goats and cattle facilitated the zoonosis of *M. caprae* and subsequent emergence of *M. bovis*, but the origins of the other MTC species remain far more enigmatic. For instance, if truly antedated by a common progenitor, how did *M. microti* come to establish infection in European voles and *M. pinnipedii* to become endemic in southern hemispheric seals? It is also quite remarkable that infection took hold in exotic species such as the hyrax but not others, such as the dog, which has been humankind’s closest companion species for millennia and in which, despite their intimate and olfaction-centric behaviors, canine TB is rarely observed (16; Marcel Behr, personal communication). Do dogs possess an innate resistance to tubercle bacillus infection?

Although spoligotype data was not collected in this study, data suggest that some phylogenetically relevant changes have occurred in the DR locus. For instance, the spoligotypes of “modern” PGG2 and PGG3 *M. tuberculosis* strains typically lack spacers 33 to 36 (83; Natalia Kurepina, personal communication), harkening to a genetic event subsequent to both the TbD1 deletion and the acquisition of the -215 *narGHJI* SNP. Similarly, *M. bovis* and *M. caprae* strains are known to lack spacers 3, 9, and 16, thereby associating their loss with the RD12, RD13, N-RD25^{bovis/cap}, *gyrB*^A (756G→A), and *oxyR*²⁸⁵ polymorphisms. All *M. microti*, *M. pinnipedii*, *M. bovis*, and *M. caprae* strains, but not the oryx bacilli, are known to lack each of spacers 40 to 43 in their spoligotypes (42, 43, 55), and so these would appear to have been lost secondary to acquisition of the *mmpL6*⁵⁵¹ SNP. In our review of the spoligotyping literature, we also noted that all MTC organisms along the *M. africanum*→*M. bovis* evolutionary track lack spacers 9 and 39, potentially placing their deletion (if due to separate single events for each) at the RD9 node (bottleneck) of the MTC evolutionary tree (Fig. 2). We have not yet seen mention of spacers 9 and 39 in print as potential markers for the differ-

entiation of *M. tuberculosis* from the remaining MTC species by spoligotyping, although their absence has been noted for *M. africanum* subtype I (87).

If the observation regarding spacers 9 and 39 is true, then it is very interesting that the spoligotype derived from mycobacterial DNA isolated from a 17,000-year-old North American bison skeleton possessed spacers 9 and 39, as well as spacers 33, 40, 41, and 43, based upon the provided spoligotyping hex code (67; Jeffrey Driscoll, personal communication). In so being, these data indicate that the bison microbe was of a more primordial MTC genotype, such as an “ancestral” *M. tuberculosis*, rather than an *M. africanum* isolate, as was reported (67). Moreover, given evidence suggestive of TB in pre-Columbian societies of central South America (69), it is unclear whether *M. tuberculosis* crossed the Bering Strait into North America with humans (12,000 to 14,000 years ago) or whether these people acquired *M. tuberculosis* from the infected bovids that had preceded them. In a recent genome-wide comparative reevaluation of the *M. tuberculosis* synonymous nucleotide substitution rate-based molecular clock of strains CDC1551 (PGG2) and H37Rv (PGG3), Hughes et al. estimated that the most recent common ancestor of these “modern” *M. tuberculosis* strains occurred approximately 35,000 years ago (39). Moreover, recent projections based upon DNA sequence diversity in “prototuberculosis” strains place the origin of the MTC precursor species at 3 million years ago (31). Combined, these calculations greatly extend the probable duration throughout which TB has been a human disease. Clearly, as with the data presented herein, closer inspection of the genomes of MTC organisms will continue to contribute new and unexpected insights to our understanding of MTC biology and evolutionary history.

ACKNOWLEDGMENTS

We thank Christine Pourcel for technical assistance as well as W. Ray Butler, Susan Massarella, Walter Haas, Davise Larone, and Barry Kreiswirth for providing bacterial isolates. We are also grateful to Gilles Vergnaud, José Roberto Lapa e Silva, and Warren D. Johnson, Jr., for suggestions, support, and encouragement.

Funding was provided by NIH grants R0-1 AI39606, R0-1 HL61960, R21 AI063147, and R21 AI063147 (to J.L.H.), NIH Fogarty International Center Training grant (FICTG) (D43 TW00018), a grant from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (Ministry of Education-Brazil), and a grant from the Laura Cook Hull Trust Fund (LCHTF) (Warren D. Johnson Jr., principal investigator). R.C.H. was supported by the LCHTF, and L.C.O.L. was an FICTG and CAPES trainee. D.V.S. received funding from EC grant 200-0630 for the Molecular Epidemiology of Tuberculosis.

REFERENCES

1. Alfredsen, S., and F. Saxegaard. 1992. An outbreak of tuberculosis in pigs and cattle caused by *Mycobacterium africanum*. *Vet. Rec.* **131**:51–53.
2. Aranaz, A., D. Cousins, A. Mateos, and L. Dominguez. 2003. Elevation of *Mycobacterium tuberculosis* subsp. *caprae* Aranaz et al. 1999 to species rank as *Mycobacterium caprae* comb. nov., sp. nov. *Int. J. Syst. Evol. Microbiol.* **53**:1785–1789.
3. Barnes, P. F., and M. D. Cave. 2003. Molecular epidemiology of tuberculosis. *N. Engl. J. Med.* **349**:1149–1156.
4. Behr, M. A., M. A. Wilson, W. P. Gill, H. Salamon, G. K. Schoolnik, S. Rane, and P. M. Small. 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**:1520–1523.
5. Blazquez, J., L. E. Espinosa de los Monteros, S. Samper, C. Martin, A. Guerrero, J. Cobo, J. van Embden, F. Baquero, and E. Gomez-Mampaso. 1997. Genetic characterization of multidrug-resistant *Mycobacterium bovis* strains from a hospital outbreak involving human immunodeficiency virus-positive patients. *J. Clin. Microbiol.* **35**:1390–1393.
6. Brodin, P., K. Eglmeier, M. Marmiesse, A. Billault, T. Garnier, S. Niemann,

- S. T. Cole, and R. Brosch. 2002. Bacterial artificial chromosome-based comparative genomic analysis identifies *Mycobacterium microti* as a natural ESAT-6 deletion mutant. *Infect. Immun.* **70**:5568–5578.
7. 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.
 8. Brudey, K., M. C. Gutierrez, V. Vincent, L. M. Parsons, M. Salfinger, N. Rastogi, and C. Sola. 2004. *Mycobacterium africanum* genotyping using novel spacer oligonucleotides in the direct repeat locus. *J. Clin. Microbiol.* **42**: 5053–5057.
 9. Cole, S. T. 2002. Comparative and functional genomics of the *Mycobacterium tuberculosis* complex. *Microbiology* **148**:2919–2928.
 10. Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry III, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentile, N. Hamlin, S. Holroyd, T. Hornsby, K. Jegels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537–544.
 11. Constant, P., E. Perez, W. Malaga, M. A. Laneelle, O. Saurel, M. Daffe, and C. Guilhot. 2002. Role of the *pks15/1* gene in the biosynthesis of phenolglycolipids in the *Mycobacterium tuberculosis* complex: evidence that all strains synthesize glycosylated *p*-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the *pks15/1* gene. *J. Biol. Chem.* **277**:38148–38158.
 12. Cousins, D. V., R. Bastida, A. Cataldi, V. Quse, S. Redrobe, S. Dow, P. Duignan, A. Murray, C. Dupont, N. Ahmed, D. M. Collins, W. R. Butler, D. Dawson, D. Rodriguez, J. Loureiro, M. I. Romano, A. Alito, M. Zumarraga, and A. Bernardelli. 2003. Tuberculosis in seals caused by a novel member of the *Mycobacterium tuberculosis* complex: *Mycobacterium pinnipedii* sp. nov. *Int. J. Syst. Evol. Microbiol.* **53**:1305–1314.
 13. Cousins, D. V., R. L. Peet, W. T. Gaynor, S. N. Williams, and B. L. Gow. 1994. Tuberculosis in imported hyrax (*Procavia capensis*) caused by an unusual variant belonging to the *Mycobacterium tuberculosis* complex. *Vet. Microbiol.* **42**:135–145.
 14. Cousins, D. V., S. N. Williams, R. Reuter, D. Forshaw, B. Chadwick, D. Coughran, P. Collins, and N. Gales. 1993. Tuberculosis in wild seals and characterization of the seal bacillus. *Aust. Vet. J.* **70**:92–97.
 15. Desmond, E., A. T. Ahmed, W. S. Probert, J. Ely, Y. Jang, C. A. Sanders, S. Y. Lin, and J. Flood. 2004. *Mycobacterium africanum* cases, California. *Emerg. Infect. Dis.* **10**:921–923.
 16. Erwin, P. C., D. A. Bemis, S. B. McCombs, L. L. Sheeler, I. M. Himelright, S. K. Halford, L. Diem, B. Metchock, T. F. Jones, M. G. Schilling, and B. V. Thomsen. 2004. *Mycobacterium tuberculosis* transmission from human to canine. *Emerg. Infect. Dis.* **10**:2258–2260.
 17. Fabre, M., J. L. Koeck, P. Le Fleche, F. Simon, V. Herve, G. Vergnaud, and C. Pourcel. 2004. 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*.” *J. Clin. Microbiol.* **42**:3248–3255.
 18. Fang, Z., C. Doig, D. T. Kenna, N. Smittipat, P. Palittapongarnpim, B. Watt, and K. J. Forbes. 1999. IS6110-mediated deletions of wild-type chromosomes of *Mycobacterium tuberculosis*. *J. Bacteriol.* **181**:1014–1020.
 19. Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, 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. Drobniowski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, G. Kallenius, E. Kassa-Kelembho, T. Koivula, H. M. Ly, A. Makrathathis, C. Mammina, G. Martin, P. Mostrom, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasoloflo-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. 2002. Global distribution of *Mycobacterium tuberculosis* spoligotypes. *Emerg. Infect. Dis.* **8**:1347–1349.
 20. Filliol, I., A. S. Motiwal, M. Cavatore, W. Qi, M. Hernandez Hazbon, M. Bobadilla Del Valle, J. Fyfe, L. Garcia-Garcia, N. Rastogi, C. Sola, T. Zozio, M. I. Guerrero, C. I. Leon, J. Crabtree, S. Angiuoli, K. D. Eisenach, R. Durmaz, M. L. Joloba, A. Rendon, J. Sifuentes-Osorio, A. Ponce de Leon, M. D. Cave, R. Fleischmann, T. S. Whittam, and D. Alland. 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J. Bacteriol.* **188**:759–772.
 21. Fleischmann, R. D., D. Alland, J. A. Eisen, L. Carpenter, O. White, J. Peterson, R. DeBoy, R. Dodson, M. Gwinn, D. Haft, E. Hickey, J. F. Kolonay, W. C. Nelson, L. A. Umayam, M. Ermolaeva, S. L. Salzberg, A. Delcher, T. Utterback, J. Weidman, H. Khouri, J. Gill, A. Mikula, W. Bishai, W. R. Jacobs, Jr., J. C. Venter, and C. M. Fraser. 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J. Bacteriol.* **184**:5479–5490.
 22. Friedman, C. R., G. C. Quinn, B. N. Kreiswirth, D. C. Perlman, N. Salomon, N. Schluger, M. Luffey, J. Berger, N. Poltoratskaia, and L. W. Riley. 1997. Widespread dissemination of a drug-susceptible strain of *Mycobacterium tuberculosis*. *J. Infect. Dis.* **176**:478–484.
 23. Frota, C. C., D. M. Hunt, R. S. Buxton, L. Rickman, J. Hinds, K. Kremer, D. van Soolingen, and M. J. Colston. 2004. Genome structure in the vole bacillus, *Mycobacterium microti*, a member of the *Mycobacterium tuberculosis* complex with a low virulence for humans. *Microbiology* **150**:1519–1527.
 24. Frothingham, R., P. L. Strickland, G. Bretzel, S. Ramaswamy, J. M. Musser, and D. L. Williams. 1999. Phenotypic and genotypic characterization of *Mycobacterium africanum* isolates from West Africa. *J. Clin. Microbiol.* **37**: 1921–1926.
 25. Garnier, T., K. Eiglmeier, J. C. Camus, N. Medina, H. Mansoor, M. Pryor, S. Duthoy, S. Grondin, C. Lacroix, C. Monsempe, S. Simon, B. Harris, R. Atkin, J. Doggett, R. Mayes, L. Keating, P. R. Wheeler, J. Parkhill, B. G. Barrell, S. T. Cole, S. V. Gordon, and R. G. Hewinson. 2003. The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl. Acad. Sci. USA* **100**: 7877–7882.
 26. Gibson, A. L., G. Hewinson, T. Goodchild, B. Watt, A. Story, J. Inwald, and F. A. Drobniowski. 2004. Molecular epidemiology of disease due to *Mycobacterium bovis* in humans in the United Kingdom. *J. Clin. Microbiol.* **42**: 431–434.
 27. 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.
 28. Greth, A., J. R. Flamand, and A. Delhomme. 1994. An outbreak of tuberculosis in a captive herd of Arabian oryx (*Oryx leucorox*): management. *Vet. Rec.* **134**:165–167.
 29. 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.
 30. Gutacker, M. M., J. C. Smoot, C. A. Migliaccio, S. M. Ricklefs, S. Hua, D. V. Cousins, E. A. Graviss, E. Shashkina, B. N. Kreiswirth, and J. M. Musser. 2002. Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics* **162**: 1533–1543.
 31. Gutierrez, M. C., S. Brisse, R. Brosch, M. Fabre, B. Omais, M. Marmiesse, P. Supply, and V. Vincent. 2005. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog.* **1**:e5.
 32. Gutierrez, M. C., J. C. Galan, J. Blazquez, E. Bouvet, and V. Vincent. 1999. Molecular markers demonstrate that the first described multi-drug resistant *Mycobacterium bovis* outbreak was due to *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **37**:971–975.
 33. Haas, W. H., G. Bretzel, B. Amthor, K. Schilke, G. Krommes, S. Rusch-Gerdes, V. Sticht-Groh, and H. J. Bremer. 1997. Comparison of DNA fingerprint patterns of isolates of *Mycobacterium africanum* from east and west Africa. *J. Clin. Microbiol.* **35**:663–666.
 34. Ho, T. B., B. D. Robertson, G. M. Taylor, R. J. Shaw, and D. B. Young. 2000. Comparison of *Mycobacterium tuberculosis* genomes reveals frequent deletions in a 20 kb variable region in clinical isolates. *Yeast* **17**:272–282.
 35. Hoffner, S. E., S. B. Svenson, R. Norberg, F. Dias, S. Ghebremichael, and G. Kallenius. 1993. Biochemical heterogeneity of *Mycobacterium tuberculosis* complex isolates in Guinea-Bissau. *J. Clin. Microbiol.* **31**:2215–2217.
 36. Howard, S. T., T. F. Byrd, and C. R. Lyons. 2002. A polymorphic region in *Mycobacterium abscessus* contains a novel insertion sequence element. *Microbiology* **148**:2987–2996.
 37. Huard, R. C., S. Chitale, M. Leung, L. C. de Oliveira Lazzarini, H. Zhu, E. Shashkima, S. Laal, M. Conde, A. L. Kritski, J. T. Belisle, B. N. Kreiswirth, J. R. Lapa e Silva, and J. L. Ho. 2003. The *Mycobacterium tuberculosis* complex-restricted gene *cfp32* encodes an expressed protein that is detectable in tuberculosis patients and is positively correlated with pulmonary interleukin-10. *Infect. Immun.* **71**:6871–6883.
 38. Huard, R. C., L. C. de Oliveira Lazzarini, W. R. Butler, D. van Soolingen, and J. L. Ho. 2003. A PCR-based method to differentiate the subspecies of the *Mycobacterium tuberculosis* complex on the basis of genomic deletions. *J. Clin. Microbiol.* **41**:1637–1650.
 39. Hughes, A. L., R. Friedman, and M. Murray. 2002. Genomewide pattern of synonymous nucleotide substitution in two complete genomes of *Mycobacterium tuberculosis*. *Emerg. Infect. Dis.* **8**:1342–1346.
 40. Khye, S. G., N. Rastogi, M. Berchel, R. C. Huard, and C. Sola. 2005. Molecular evolutionary history of tubercle bacilli assessed by study of the polymorphic nucleotide within the nitrate reductase (*narGHJ*) operon promoter. *J. Clin. Microbiol.* **43**:4010–4014.
 41. Koivula, T., M. Ekman, T. Leitner, S. Lofdahl, S. Ghebremichael, S. Mostowy, M. A. Behr, S. B. Svenson, and G. Kallenius. 2004. Genetic

- characterization of the Guinea-Bissau family of *Mycobacterium tuberculosis* complex strains. *Microbes Infect.* 6:272–278.
42. Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. P. Blikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. van Embden. 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. *J. Clin. Microbiol.* 37:2607–2618.
 43. Kubica, T., S. Rüscher-Gerdes, and S. Niemann. 2003. *Mycobacterium bovis* subsp. *caprae* caused one-third of human *M. bovis*-associated tuberculosis cases reported in Germany between 1999 and 2001. *J. Clin. Microbiol.* 41:3070–3077.
 44. Lawler, A. 2003. International collaboration. Iran reopens its past. *Science* 302:970–973.
 45. Lee, H., H.-J. Park, S.-N. Cho, G.-H. Bai, and S.-J. Kim. 2000. Species identification of mycobacteria by PCR-restriction fragment length polymorphism of the *rpoB* gene. *J. Clin. Microbiol.* 38:2966–2971.
 46. Le Flèche, P., M. Fabre, F. Denoëud, 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.* 2:37–48.
 47. Lomme, J. R., C. O. Thoen, E. M. Himes, J. W. Vinson, and R. E. King. 1976. *Mycobacterium tuberculosis* infection in two East African oxryes. *J. Am. Vet. Med. Assoc.* 169:912–914.
 48. 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.
 49. Michalak, K., C. Austin, S. Diesel, M. J. Bacon, P. Zimmerman, and J. N. Maslow. 1998. *Mycobacterium tuberculosis* infection as a zoonotic disease: transmission between humans and elephants. *Emerg. Infect. Dis.* 4:283–287.
 50. Mostowy, S., D. Cousins, and M. A. Behr. 2004. Genomic interrogation of the dassie bacillus reveals it as a unique RD1 mutant within the *Mycobacterium tuberculosis* complex. *J. Bacteriol.* 186:104–109.
 51. Mostowy, S., D. Cousins, J. Brinkman, A. Aranaz, and M. A. Behr. 2002. Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. *J. Infect. Dis.* 186:74–80.
 52. Mostowy, S., J. Inwald, S. Gordon, C. Martin, R. Warren, K. Kremer, D. Cousins, and M. A. Behr. 2005. Revisiting the evolution of *Mycobacterium bovis*. *J. Bacteriol.* 187:6386–6395.
 53. Mostowy, S., A. Onipede, S. Gagneux, S. Niemann, K. Kremer, E. P. Desmond, M. Kato-Maeda, and M. Behr. 2004. Genomic analysis distinguishes *Mycobacterium africanum*. *J. Clin. Microbiol.* 42:3594–3599.
 54. Mostowy, S., A. G. Tsolaki, P. M. Small, and M. A. Behr. 2003. The in vitro evolution of BCG vaccines. *Vaccine* 21:4270–4274.
 55. Mostrom, P., M. Gordon, C. Sola, M. Ridell, and N. Rastogi. 2002. Methods used in the molecular epidemiology of tuberculosis. *Clin. Microbiol. Infect.* 8:694–704.
 56. Musser, J. M., A. Amin, and S. Ramaswamy. 2000. Negligible genetic diversity of *Mycobacterium tuberculosis* host immune system protein targets: evidence of limited selective pressure. *Genetics* 155:7–16.
 57. Niemann, S., D. Harmsen, S. Rusch-Gerdes, and E. Richter. 2000. Differentiation of clinical *Mycobacterium tuberculosis* complex isolates by *gyrB* DNA sequence polymorphism analysis. *J. Clin. Microbiol.* 38:3231–3234.
 58. Niemann, S., S. Rusch-Gerdes, M. L. Joloba, C. C. Whalen, D. Guwatudde, J. J. Ellner, K. Eisenach, N. Fumokong, J. L. Johnson, T. Aisu, R. D. Mugerwa, A. Okwera, and S. K. Schwander. 2002. *Mycobacterium africanum* subtype II is associated with two distinct genotypes and is a major cause of human tuberculosis in Kampala, Uganda. *J. Clin. Microbiol.* 40:3398–3405.
 59. Oh, P., R. Granich, J. Scott, B. Sun, M. Joseph, C. Stringfield, S. Thisdell, J. Staley, D. Workman-Malcolm, L. Borenstein, E. Lehnkering, P. Ryan, J. Soukup, A. Nitta, and J. Flood. 2002. Human exposure following *Mycobacterium tuberculosis* infection of multiple animal species in a metropolitan zoo. *Emerg. Infect. Dis.* 8:1290–1293.
 60. Palenque, E., V. Villena, M. J. Rebollo, M. S. Jimenez, and S. Samper. 1998. Transmission of multidrug-resistant *Mycobacterium bovis* to an immunocompetent patient. *Clin. Infect. Dis.* 26:995–996.
 61. Proding, W. M., A. Eigentler, F. Allerberger, M. Schonbauer, and W. Glawischnig. 2002. Infection of red deer, cattle, and humans with *Mycobacterium bovis* subsp. *caprae* in western Austria. *J. Clin. Microbiol.* 40:2270–2272.
 62. Pym, A. S., P. Brodin, L. Majlessi, R. Brosch, C. Demangel, A. Williams, K. E. Griffiths, G. Marchal, C. Leclerc, and S. T. Cole. 2003. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nat. Med.* 9:533–539.
 63. Rajakumar, K., J. Shafi, R. J. Smith, R. A. Stabler, P. W. Andrew, D. Modha, G. Bryant, P. Monk, J. Hinds, P. D. Butcher, and M. R. Barer. 2004. Use of genome level-informed PCR as a new investigational approach for analysis of outbreak-associated *Mycobacterium tuberculosis* isolates. *J. Clin. Microbiol.* 42:1890–1896.
 64. Ramaswamy, S. V., R. Reich, S. J. Dou, L. Jasperse, X. Pan, A. Wanger, T. Quitugua, and E. A. Graviss. 2003. Single nucleotide polymorphisms in genes associated with isoniazid resistance in *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* 47:1241–1250.
 65. Rastogi, N., K. S. Goh, M. Berchell, and A. Bryskier. 2000. Activity of rifapentine and its metabolite 25-O-desacetyl-rifapentine compared with rifampicin and rifabutin against *Mycobacterium tuberculosis*, *Mycobacterium africanum*, *Mycobacterium bovis* and *M. bovis* BCG. *J. Antimicrob. Chemother.* 46:565–570.
 66. Richter, E., M. Weizenegger, S. Rusch-Gerdes, and S. Niemann. 2003. Evaluation of genotype MTBC assay for differentiation of clinical *Mycobacterium tuberculosis* complex isolates. *J. Clin. Microbiol.* 41:2672–2675.
 67. Rothschild, B. M., L. D. Martin, G. Lev, H. Bercovier, G. K. Bar-Gal, C. Greenblatt, H. Donoghue, M. Spigelman, and D. Brittain. 2001. *Mycobacterium tuberculosis* complex DNA from an extinct bison dated 17,000 years before the present. *Clin. Infect. Dis.* 33:305–311.
 68. Salamon, H., M. Kato-Maeda, P. M. Small, J. Drenkon, and T. R. Gingeras. 2000. Detection of deleted genomic DNA using a semiautomated computational analysis of GeneChip data. *Genome Res.* 10:2044–2054.
 69. Salo, W. L., A. C. Auferheide, J. Buikstra, and T. A. Holcomb. 2004. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proc. Natl. Acad. Sci. USA* 91:2091–2094.
 70. Sasseti, C. M., D. H. Boyd, and E. J. Rubin. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* 48:77–84.
 71. Scorpio, A., D. Collins, D. Whipple, D. Cave, J. Bates, and Y. Zhang. 1997. Rapid differentiation of bovine and human tubercle bacilli based on a characteristic mutation in the bovine pyrazinamidase gene. *J. Clin. Microbiol.* 35:106–110.
 72. Scorpio, A., and Y. Zhang. 1996. Mutations in *pncA*, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. *Nat. Med.* 2:662–667.
 73. Semret, M., G. Zhai, S. Mostowy, C. Cleto, D. Alexander, G. Cangelosi, D. Cousins, D. M. Collins, D. van Soolingen, and M. A. Behr. 2004. Extensive genomic polymorphism within *Mycobacterium avium*. *J. Bacteriol.* 186:6332–6334.
 74. Smith, N. 1960. The dassie bacillus. *Tubercle (London)* 41:203–212.
 75. Sola, C., N. Rastogi, M. C. Gutierrez, V. Vincent, R. Brosch, L. Parsons, S. Niemann, S. Rüscher-Gerdes, and S. K. Schwander. 2003. Is *Mycobacterium africanum* subtype II (Uganda I and Uganda II) a genetically well-defined subspecies of the *Mycobacterium tuberculosis* complex? *J. Clin. Microbiol.* 41:1345–1348.
 76. Sreevatsan, S., P. Escalante, X. Pan, D. A. Gillies, Jr., S. Siddiqui, C. N. Khalaf, B. N. Kreiswirth, P. Bifani, L. G. Adams, T. Ficht, V. S. Perumaalla, M. D. Cave, J. D. van Embden, and J. M. Musser. 1996. Identification of a polymorphic nucleotide in *oxyR* specific for *Mycobacterium bovis*. *J. Clin. Microbiol.* 34:2007–2010.
 77. 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.
 78. Stermann, M., A. Bohrsen, C. Diephaus, S. Maass, and F. C. Bange. 2003. Polymorphic nucleotide within the promoter of nitrate reductase (*narGHJ*) is specific for *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 41:3252–3259.
 79. Stermann, M., L. Sedlacek, S. Maass, and F. C. Bange. 2004. A promoter mutation causes differential nitrate reductase activity of *Mycobacterium tuberculosis* and *Mycobacterium bovis*. *J. Bacteriol.* 186:2856–2861.
 80. Telenti, A., F. Marchesi, M. Balz, F. Bally, E. C. Bottger, and T. Bodmer. 1993. Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. *J. Clin. Microbiol.* 31:175–178.
 81. Tortoli, E. 2003. Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. *Clin. Microbiol. Rev.* 16:319–354.
 82. Tsolaki, A. G., A. E. Hirsh, K. DeRiemer, J. A. Enciso, M. Z. Wong, M. Hannan, Y. O. Goguet de la Salmoniere, K. Aman, M. Kato-Maeda, and P. M. Small. 2004. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proc. Natl. Acad. Sci. USA* 101:4865–4870.
 83. 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.
 84. van Soolingen, D., P. E. de Haas, J. Haagsma, T. Eger, P. W. Hermans, V. Ritacco, A. Alito, and J. D. van Embden. 1994. Use of various genetic markers in differentiation of *Mycobacterium bovis* strains from animals and humans and for studying epidemiology of bovine tuberculosis. *J. Clin. Microbiol.* 32:2425–2433.
 85. 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. A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. *Int. J. Syst. Bacteriol.* 47:1236–1245.
 86. van Soolingen, D., A. G. van der Zanden, P. E. de Haas, G. T. Noordhoek, A.

- Kiers, N. A. Foudraine, F. Portaels, A. H. Kolk, K. Kremer, and J. D. van Embden. 1998. Diagnosis of *Mycobacterium microti* infections among humans by using novel genetic markers. *J. Clin. Microbiol.* **36**:1840–1845.
87. Viana-Niero, C., C. Gutierrez, C. Sola, I. Filliol, F. Boulahbal, V. Vincent, and N. Rastogi. 2001. Genetic diversity of *Mycobacterium africanum* clinical isolates based on IS6110-restriction fragment length polymorphism analysis, spoligotyping, and variable number of tandem DNA repeats. *J. Clin. Microbiol.* **39**:57–65.
88. Vijaya Bhanu, N., D. van Soolingen, J. D. van Embden, L. Dar, R. M. Pandey, and P. Seth. 2002. Predominance of a novel *Mycobacterium tuberculosis* genotype in the Delhi region of India. *Tuberculosis* **82**:105–112.
89. Wagner, J. C., and V. Bokkenheuser. 1961. The *Mycobacterium* isolated from the dassie *Procavia capensis* (Pallas). *Tubercle* **42**:47–56.
90. Wagner, J. C., G. Buchanan, V. Bokkenheuser, and S. Levisseur. 1958. An acid-fast bacillus isolated from the lungs of the Cape hyrax, *Procavia capensis* (Pallas). *Nature* **181**:284–285.
91. 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.
92. Weber, I., C. Fritz, S. Ruttkowski, A. Kreft, and F. C. Bange. 2000. Anaerobic nitrate reductase (*narGHII*) activity of *Mycobacterium bovis* BCG in vitro and its contribution to virulence in immunodeficient mice. *Mol. Microbiol.* **35**:1017–1025.
93. Zumarraga, M. J., A. Bernardelli, R. Bastida, V. Quse, J. Loureiro, A. Cataldi, F. Bigi, A. Alito, M. Castro Ramos, S. Samper, I. Otal, C. Martin, and M. I. Romano. 1999. Molecular characterization of mycobacteria isolated from seals. *Microbiology* **145**:2519–2526.