JOURNAL OF CLINICAL MICROBIOLOGY, Aug. 2000, p. 2846–2852 Vol. 38, No. 8
0095-1137/00/$04.00+0

Evaluation of *recA* Sequences for Identification of *Mycobacterium* Species

KYM S. BLACKWOOD,[1] CHENG HE,[2] JAMES GUNTON,[1] CHRISTINE Y. TURENNE,[1]* JOYCE WOLFE,[1,2] AND AMIN M. KABANI[1,2]

National Reference Centre for Mycobacteriology, Bureau of Microbiology, Health Canada,[1] and Department of Clinical Microbiology, Health Sciences Centre,[2] Winnipeg, Manitoba, Canada

Received 3 April 2000/Returned for modification 1 May 2000/Accepted 29 May 2000

16S rRNA sequence data have been used to provide a molecular basis for an accurate system for identification of members of the genus *Mycobacterium*. Previous studies have shown that *Mycobacterium* species demonstrate high levels (>94%) of 16S rRNA sequence similarity and that this method cannot differentiate between all species, i.e., *M. gastri* and *M. kansasii*. In the present study, we have used the *recA* gene as an alternative sequencing target in order to complement 16S rRNA sequence-based genetic identification. The *recA* genes of 30 *Mycobacterium* species were amplified by PCR, sequenced, and compared with the published *recA* sequences of *M. tuberculosis*, *M. smegmatis*, and *M. leprae* available from GenBank. By *recA* sequencing the species showed a lower degree of interspecies similarity than they did by 16S rRNA gene sequence analysis, ranging from 96.2% between *M. gastri* and *M. kansasii* to 75.7% between *M. aurum* and *M. leprae*. Exceptions to this were members of the *M. tuberculosis* complex, which were identical. Two strains of each of 27 species were tested, and the intraspecies similarity ranged from 98.7 to 100%. In addition, we identified new *Mycobacterium* species that contain a protein intron in their *recA* genes, similar to *M. tuberculosis* and *M. leprae*. We propose that *recA* gene sequencing offers a complementary method to 16S rRNA gene sequencing for the accurate identification of the *Mycobacterium* species.

Mycobacteria are aerobic or microaerophilic rods that are characterized by acid-fast properties and high G+C contents (8). The conventional identification of *Mycobacterium* species is still based on biochemical analysis, which is both time-consuming and labor-intensive, since cultivation of many of the slowly growing *Mycobacterium* species may take weeks. In addition, the biochemical tests are further hindered by the increasing number of rare and newly discovered disease-causing *Mycobacterium* species. Alternative methods, such as cell wall lipid chromatography, require pure culture and significant amounts of bacteria that normally need at least 3 weeks of growth (9). The DNA-based diagnostic methods that are available such as methods that use the DNA probe technology developed by Gen-Probe are rapid and relatively sensitive, but their chief drawback is that they have limited numbers of species-specific probes (12).

One of the most accurate molecular identification methods is based on 16S rRNA gene sequences, with identification commonly being achieved by comparison of the variable regions in the 16S rRNA gene. Within the *Mycobacterium* genus, the interspecies percent similarity of the 16S rRNA gene sequences is relatively high, from 94.3% between *M. chelonae* and *M. xenopi* to 100% between *M. kansasii* and *M. gastri* (17). In this study, we chose to investigate the genetic relatedness of *Mycobacterium* species using the *recA* gene, which exists in all bacteria due to its important function in homologous DNA recombination, DNA damage repair, and induction of the SOS response (13). As part of the SOS response, *recA* coordinates the induction of over 20 genes involved in DNA repair, DNA synthesis, DNA recombination, and cell division (11, 13). Bacterial classification and identification derived from the results of *recA* gene sequence analysis have previously been studied, with a focus on gram-negative bacteria and a few gram-positive bacteria (4–6). Comparative phylogenetic analyses based on the *recA* and 16S rRNA gene sequences of various bacterial species have demonstrated highly similar branching patterns (6), indicating that the *recA* gene is a good choice for use in molecular systematic studies and species identification. However, the *recA* genes of *Mycobacterium* species have been far less studied. To date, only three sequences from *Mycobacterium* species are published or available from GenBank. The lengths of the *recA* sequences of *M. tuberculosis* and *M. leprae* are 2,373 and 2,136 bp, respectively. Both contain a protein intron, and these protein introns in the two species vary in size and location within the *recA* gene (2). The *recA* gene of *M. smegmatis* does not contain the protein intron and is 1,050 bp in length (15).

The primary purpose of this study was to determine the potential of *recA* gene sequencing for the identification of *Mycobacterium* species, of which more than 80 have been described (12), by using characterized reference strains as a foundation. The secondary purpose of the study was to determine the utility of *recA* gene sequencing in comparison with that of 16S rRNA gene sequencing, particularly among species for which the similarity is above 99%.

MATERIALS AND METHODS

Bacterial species and media. The reference strains used in this study are listed in Table 1. All strains, stocked at −20°C in skim milk, were inoculated into BACTEC 12B liquid medium (Becton-Dickinson, Oakville, Ontario, Canada) and were subcultured onto either Middlebrook 7H10 agar or Lowenstein-Jensen slant and grown under optimum conditions, depending on the species. A loopful of bacteria from a solid-medium culture was resuspended in 1 ml of sterile distilled H_2O containing 4- to 6-mm-diameter glass beads. The mixture was vortexed for 2 min for mechanical breakage, transferred to a 1.5-ml microcentrifuge tube, and boiled for 10 min. The resulting crude lysate was stored at −20°C until PCR.

* Corresponding author. Mailing address: National Reference Centre for Mycobacteriology, Canadian Science Centre for Human and Animal Health, 1015 Arlington St., Winnipeg, Manitoba, Canada, R3E 3R2. Phone: (204) 789-6081. Fax: (204) 789-2036. E-mail: cturenne@hc-sc.gc.ca.

TABLE 1. *Mycobacterium* species and strains used in the study[a]

Organism	Strains tested	% Similarity	% Divergence
Slow-growing species			
M. africanum	ATCC[b] 25420^T	NA[c]	
M. asiaticum	ATCC 25276^T, ATCC 25274	99.4	0.6
M. avium	ATCC 25291^T, ATCC 35717	99.3	0.7
M. bovis	ATCC 35720, ATCC 35726	100	0.0
M. gastri	ATCC 15754^T, EB 1609	100	0.0
M. gordonae	ATCC 14470^T, ATCC 35756	99.6	0.4
M. intracellulare	ATCC 13950^T, ATCC 25122	100	0.0
M. kansasii	ATCC 12478^T, ATCC 35755	100	0.0
M. leprae	(X73822)[d]	NA	
M. marinum	ATCC 927^T, ATCC 11564	99.6	0.4
M. microti	ATCC 19422^T, ATCC 11152	100	0.0
M. nonchromogenicum	ATCC 19531, ATCC 35783	99.6	0.4
M. scrofulaceum	ATCC 19981^T, ATCC 35788	98.7	1.3
M. shimodei	ATCC 27962^T	NA	
M. simiae	ATCC 25275^T, 8988/68	99.9	0.1
M. szulgai	ATCC 35799^T, NCTC[e] 10829	99.9	0.1
M. terrae	ATCC 15755^T, EB 1614	100	0.0
M. triviale	ATCC 23292^T, TMC 1543	100	0.0
M. tuberculosis	H37Rv, ATCC 27294^T (X58485), Canetti (AJ000012)	100	0.0
M. xenopi	ATCC 19250^T, EB 1362	99.7	0.3
Fast-growing species			
M. abscessus	ATCC 19977^T, ATCC 23003	100	0.0
M. album	ATCC 29676, ATCC 29677	99.2	0.5
M. aurum	ATCC 23366^T	NA	
M. chelonae	ATCC 19237, ATCC 35749	99.9	0.1
M. fortuitum	ATCC 6841^T, ATCC 6842	100	0.0
M. gadium	ATCC 27726^T, CASAL 1080	100	0.0
M. mucogenicum	ATCC 49650^T, ATCC 49651	98.7	0.9
M. peregrinum	ATCC 14467^T, ATCC 23015	96.2	3.9
M. phlei	ATCC 11758^T, ATCC 27206	99.8	0.2
M. porcinum	ATCC 33776^T, ATCC 33775	100	0.0
M. smegmatis	ATCC 19420^T, mc2 155 (X99208)	99.6	0.4

[a] The percent similarity and percent divergence are indicated for all species of which two strains each were tested.
[b] ATCC, American Type Culture Collection, Manassas, Va.
[c] NA, not applicable.
[d] Sequences obtained from GenBank; accession numbers are given in parentheses.
[e] NCTC, National Collection of Type Cultures, London, England.

Primer design and PCR. Four relatively conserved regions were identified by aligning the available *Mycobacterium recA* sequences found in GenBank, and these four regions were used to design four pairs of degenerate primers. All oligonucleotides were synthesized by the DNA Core Facility, Bureau of Microbiology, Health Canada. PCR was performed with the GeneAmp PCR system 9600 (PE Biosystems, Foster City, Calif.). The reaction mix consisted of 5 μl of crude DNA lysate, each deoxynucleoside triphosphate at a concentration of 200 μM, each primer at a concentration of 1 μM, 1× PCR buffer with 1.5 mM $MgCl_2$ (Qiagen Inc., Valencia, Calif.), and 1.25 U of *Taq* (Qiagen Inc.) with 1× Q solution (Qiagen Inc.) for a total volume of 50 μl. For the amplification of the first PCR product, fragment A (Fig. 1), a forward degenerate primer (recF1; 5′-GGT GTT CGN CTA NTG TGG TG-3′) was paired with a reverse degenerate primer (recR1; 5′-AGC TGG TTG ATG AAG ATY GC-3′). For those strains from which a product was not amplified, seminested PCR was performed with a 1/100 dilution of the recF1-recR1 PCR product by using forward degenerate primer recF2 (5′-GYG TCA CSG CCA ACC GAY C-3′) and recR1. The cycles used were 5 min at 94°C, followed by 30 cycles of 94°C for 1 min, 48°C for 1 min, and 72°C for 1 min, with a final extension at 72°C for 10 min. For the amplification of the second product, fragment B, forward degenerate primer recF3 (5′-GGC AAR GGY TCG GTS ATG C-3′) and reverse primer recR2 (5′-TTG ATC TTC TTC TCG ATC TC-3′) were used in a touchdown PCR protocol. Conditions were 94°C for 5 min; 5 cycles of 94°C for 1 min, 50°C for 1 min (with a decrease of 1°C each cycle), and 72°C for 1 min; 25 cycles of 94°C for 1 min, 45°C for 1 min, and 72°C for 1 min; and a final extension of 72°C for 10 min. For amplification of *recA* gene fragment B that contained an intein (a posttranslationally self-splicing protein intron), as well as for difficult templates with significant nonspecific amplification or weak or no amplification, a different forward primer, recF4, was used. The sequence of primer recF4 is complementary to that of primer recR1 and is used in a seminested PCR with a 1/100 dilution of the recF3-recR2 product. Conditions for this PCR are 94°C for 5 min; 3 cycles of 94°C for 1 min, 40°C for 1 min, 2 min of ramping to 72°C, and 72°C for 1 min; 27 cycles of 94°C for 1 min, 55°C for 1 min, 1 min of ramping to 72°C, and 72°C for 1 min; and a final extension of 72°C for 10 min. A schematic representation of the primer location along the gene can be found in Fig. 1.

PCR product detection, purification, and quantitation. PCR products were visualized by UV detection of an ethidium bromide-stained 1.5% agarose gel following electrophoresis. The purification of the remaining PCR product was achieved with Microcon-100 microconcentrators (Millipore Corp., Bedford, Mass.) by following the manufacturer's instruction. On occasion, the PCR product was shown to contain a second nonspecific band on the gel, in which case the desired product was cut out of the gel and was purified with the QIAquick Gel Extraction Kit (Qiagen Inc.) by following the manufacturer's instructions. The concentration of the purified PCR product was determined spectrophotometrically by measuring the A_{260}.

DNA sequencing and sequencing analysis. The ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems) was used for the sequencing of the PCR product. The sequencing reaction required 4 μl of premix, 3.2 pmol of sequencing primer, and 150 ng of PCR product template in a total volume of 10 μl. PCR primers were used for sequencing, as was primer recG1 (5′-CTS GAR ATC GCC GAC ATG CTG-3′) (Fig. 1). The sequencing reaction and template preparation were performed in accordance with the instructions of the manufacturer (PE Biosystems). The sequencing product was purified with the Centricep columns (Princeton Separations, Adelphia, N.J.) recommended by PE Biosystems.

The sequencing output was analyzed by using the DNA Sequence Analyzer computer software (PE Biosystems). The Lasergene program, version 4.01 (DNASTAR, Inc., Madison, Wis.), was used for sequence assembly, sequence alignment, and phylogenetic analysis. Multiple sequence alignments were determined by using the Clustal method algorithm.

RESULTS

PCR with the designed primers yielded two products which have overlapping sequences corresponding to the first ~970 bp of the ~1-kb *recA* gene (Fig. 1). By using forward primers recF1 or recF2 in either a separate or a seminested reaction,

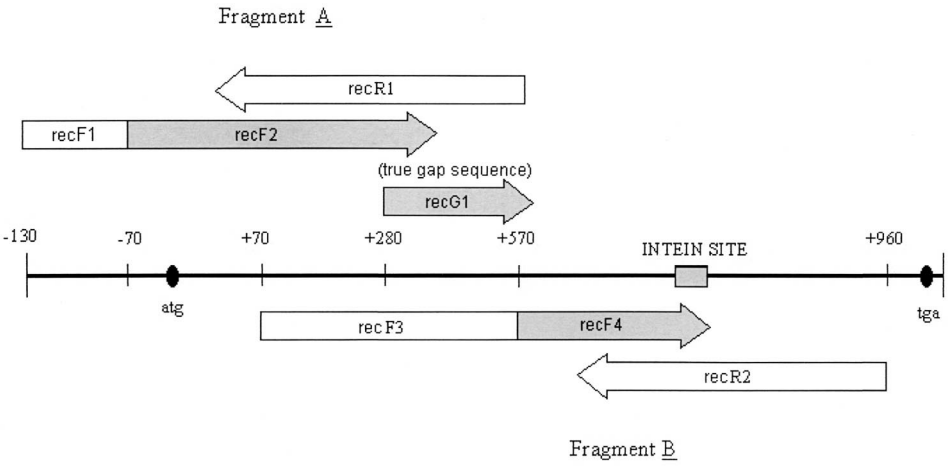

FIG. 1. Schematic illustration of the primer pairs and sites used in the amplification of the *recA* gene. The shaded segments correspond to primers applied for seminested PCR. Primer recG1 was used to determine the true sequence of the gap left between primers recR1 and recF4 when seminested PCR was performed.

fragment A was obtained from all species tested. Fragment A is homologous to the 5′ end of the *recA* gene and also contains a short stretch of DNA homologous to the sequence upstream of the *recA* gene. As a result, the exact size of this PCR product varies slightly among species (data not shown). When compared with the *recA* sequences of *M. tuberculosis* and *M. leprae* (2, 3), the fragment A sequence was determined to contain the region from 1 to 587 bp of the *recA* gene. Fragment B was obtained from most of the organisms tested by using primer set recF3 and recR2, and the sequencing results indicated that the size of this product was 907 bp, from 67 to 974 bp of the *recA* gene.

The sequences of PCR fragments A and B were combined to yield a usable nucleotide sequence of a minimum of 915 bp up to a maximum of ~970 bp of the *recA* gene (excluding the protein intron region found in certain species). Thus, the first 915 bp of all species was used for sequence alignment and analysis of the *recA* gene, which revealed the presence of a large number of nucleotide substitutions among the species tested, with interspecies similarities ranging from 75.7% between *M. leprae* and *M. aurum* or *M. mucogenicum* to 96.2% between *M. gastri* and *M. kansasii* (Table 2). Six of the species tested, *M. fortuitum*, *M. peregrinum*, *M. album*, *M. porcinum*, *M. aurum*, and *M. mucogenicum*, contain an extra glutamine near the N terminus at amino acid position 4, while the rest of the species studied do not have the insertion.

The intraspecies variability has been determined with 2 reference strains of each of 27 of the *Mycobacterium* species tested in this study (Table 1). No intraspecies variability was present in 13 of the species, whereas 13 species demonstrated an intraspecies variability that ranged from 98.7 to 99.9%. Comparison of two strains of *M. peregrinum*, however, resulted in a 96.2% similarity. The 16S rRNA gene sequences of these two strains were also determined in our laboratory (data not shown) and indicated a 100% similarity. A phylogenetic tree of the 31 species, including those species tested as well as *M. leprae* and *M. tuberculosis*, was generated with the type strain of each species set when possible as well as with both *M. peregrinum* strains (Fig. 2).

Amplification of fragment B from *M. xenopi*, *M. asiaticum*, *M. shimoidei*, and members of the *M. tuberculosis* clade resulted in a PCR product ~1 kb larger than expected (data not shown), suggesting the presence of a DNA insertion. Sequencing results for *M. xenopi* indicated that the PCR product was indeed the desired fragment of the *recA* gene in which 1,092 bp of extraneous DNA was inserted. In these cases, another forward primer, recF4, approximately 500 bp downstream from the *recA* gene start codon, was used instead of recF3 in a seminested reaction. In addition, the complete protein intron of *M. xenopi* ATCC 19250 was sequenced by using a pair of primers specifically designed from the sequenced *recA* region that flanked the intein and was compared to the intein of *M. leprae* (3). The size of the insertion fragment in *M. xenopi* is 1 amino acid residue short of the size of the protein intron in *M. leprae*. The missing amino acid residue was identified as L140. More importantly, the sites of insertion within the *recA* genes of these two species are identical. Despite the similarity in size and location, the two protein introns show only 86% similarity at the protein level and 77.8% similarity at the DNA level (data not shown). Likewise, *M. shimoidei* and *M. asiaticum* were also found to contain insertions that resemble the intein of *M. leprae* (partial sequences were determined).

DISCUSSION

Despite its role in DNA recombination and repair, the *recA* genes of *Mycobacterium* species have not been studied extensively. The only known *recA* sequences available prior to this study were those of *M. tuberculosis*, *M. leprae*, and *M. smegmatis*.

Sequence alignment and analysis of the *recA* genes that belong to 31 species of mycobacteria revealed the presence of a large number of nucleotide substitutions among the species tested. Unlike the 16S rRNA gene, in which variability is confined to certain areas of the gene, the sequence similarities of the *recA* genes of *Mycobacterium* species are significantly lower among species (≤96.2%) and variability occurs throughout the *recA* gene. The majority of substitutions were found to be confined to the third position of a codon, also known as the wobble position, allowing a conserved amino acid sequence across the genus, thereby preserving the important functions of the RecA protein. This pattern of sequence divergence is analogous to the more extensively studied *recA* sequences of *Escherichia coli* (13).

Previous studies have found that the *recA* gene of *M. smegmatis* (15) contains an extra glutamine residue near the N terminus, at amino acid position 4 (nucleotide positions 10 to 12), whereas *M. tuberculosis* and *M. leprae* do not have this

TABLE 2. Interspecies similarity of partial *recA* gene sequences (~915 bp)[a]

Species	% Similarity																													
	M. africanum	*M. bovis*	*M. microti*	*M. abscessus*	*M. chelonae*	*M. xenopi*	*M. szulgai*	*M. intracellulare*	*M. avium*	*M. gastri*	*M. kansasii*	*M. gordonae*	*M. asiaticum*	*M. simiae*	*M. scrofulaceum*	*M. gadium*	*M. phlei*	*M. shimoidei*	*M. terrae*	*M. nonchromogenicum*	*M. triviale*	*M. marinum*	*M. leprae*	*M. smegmatis*	*M. mucogenicum*	*M. fortuitum*	*M. peregrinum* ATCC 14467	*M. peregrinum* ATCC 23015	*M. porcinum*	*M. album*	*M. aurum*
M. tuberculosis	100	100	100	81.5	81.7	86.3	89.4	89.7	90.0	91.4	90.7	88.9	89.8	89.8	89.7	85.9	86.4	86.7	86.7	86.1	84.9	89.5	83.3	85.5	84.3	85.1	84.5	84.9	86.1	84.6	81.5
M. africanum		100	100	81.5	81.7	86.3	89.4	89.7	90.0	91.4	90.7	88.9	89.8	89.8	89.7	85.9	86.4	86.7	86.7	86.1	84.9	89.5	83.3	85.5	84.3	85.1	84.5	84.9	86.1	84.6	81.5
M. bovis			100	81.5	81.7	86.3	89.4	89.7	90.0	91.4	90.7	88.9	89.8	89.9	89.7	85.9	86.4	86.7	86.7	86.1	84.9	89.5	83.3	85.5	84.3	85.1	84.5	84.9	86.1	84.6	81.5
M. microti				81.5	81.7	86.3	89.4	89.7	90.0	91.4	90.7	88.9	89.8	89.9	89.7	85.9	86.4	86.7	86.7	86.1	84.9	89.5	83.3	85.5	84.3	85.1	84.5	84.9	86.1	84.6	81.5
M. abscessus					90.4	81.1	83.0	85.5	83.6	83.0	82.9	83.6	82.7	84.5	84.2	85.0	84.9	82.9	84.4	85.0	84.2	82.0	78.9	87.4	88.4	85.5	85.2	84.5	85.9	85.5	82.3
M. chelonae						80.0	82.9	85.1	84.2	83.2	82.9	81.7	82.2	85.1	84.5	85.2	84.2	82.6	84.4	85.0	84.0	81.1	78.4	85.5	87.9	85.0	86.0	85.7	86.2	85.0	82.3
M. xenopi							85.3	86.6	85.6	85.7	85.4	84.5	84.8	87.5	86.3	83.0	82.9	86.6	84.9	84.6	83.8	84.2	80.8	82.9	81.5	82.7	82.3	82.2	83.4	83.2	78.5
M. szulgai								88.7	90.0	90.8	90.6	89.1	90.1	89.7	89.4	85.2	86.8	86.5	86.5	85.7	85.4	90.5	83.0	85.0	84.4	84.8	84.2	84.1	85.2	85.0	80.9
M. intracellulare									95.9	91.1	90.2	90.2	90.0	95.0	95.4	87.7	89.7	88.5	90.8	90.7	88.8	88.9	84.6	87.6	87.6	87.6	87.5	87.8	89.4	88.0	84.3
M. avium										91.1	90.4	89.1	90.0	94.3	95.1	87.5	89.9	88.4	91.3	89.9	89.1	88.8	84.1	87.6	87.4	87.2	87.8	87.6	88.8	86.8	84.4
M. gastri											96.2	89.9	90.8	90.2	90.4	86.3	87.7	87.2	87.4	87.2	85.5	90.9	86.0	85.2	84.3	85.5	85.2	85.4	86.4	85.5	82.0
M. kansasii												89.4	89.9	90.1	90.0	86.5	87.6	86.4	87.2	86.7	85.0	91.0	86.1	85.1	84.3	85.3	85.1	84.9	86.5	84.4	81.2
M. gordonae													90.1	89.1	89.1	87.1	87.8	86.6	87.8	87.6	86.0	88.9	82.1	85.5	83.4	85.0	84.5	84.3	86.0	85.2	80.9
M. asiaticum														90.1	89.6	86.8	87.8	86.3	88.5	87.8	86.6	89.4	84.6	86.1	85.5	86.3	85.5	85.3	86.4	85.7	82.4
M. simiae															94.5	86.1	87.9	87.8	90.4	89.8	88.6	88.9	84.5	87.3	87.2	86.6	87.1	87.5	88.8	87.0	83.9
M. scrofulaceum																87.0	88.8	87.8	89.9	89.5	89.4	88.7	84.6	87.3	86.7	87.2	86.8	87.3	88.7	85.7	83.2
M. gadium																	94.8	87.3	86.7	87.6	87.3	85.1	79.6	87.5	86.6	87.5	87.3	87.5	88.0	87.2	84.9
M. phlei																		87.1	88.5	88.7	88.5	86.6	79.9	89.3	88.2	88.2	88.2	88.4	88.9	88.8	85.9
M. shimoidei																			88.5	88.3	86.3	85.6	81.8	84.6	83.8	84.9	85.7	85.7	85.9	85.9	83.8
M. terrae																				93.2	89.0	86.2	81.9	88.9	87.1	87.8	88.7	88.8	89.1	87.7	84.9
M. nonchromogenicum																					88.9	86.0	82.5	88.4	87.5	88.9	89.4	89.5	89.5	87.8	84.5
M. triviale																						85.2	79.7	87.2	87.0	86.4	86.4	86.4	86.7	86.0	84.7
M. marinum																							83.2	84.8	84.3	84.6	83.7	83.7	85.4	83.7	80.6
M. leprae																								78.5	78.6	79.7	79.4	79.3	80.3	79.6	75.7
M. smegmatis																									89.8	92.2	92.2	92.0	93.2	90.9	86.2
M. mucogenicum																										89.0	89.6	89.6	90.1	89.3	85.4
M. fortuitum																											94.3	94.0	94.4	90.9	84.7
M. peregrinum ATCC 14467																												96.2	95.4	90.7	85.0
M. peregrinum ATCC 23015																													95.7	90.6	85.4
M. porcinum																														91.8	85.8
M. album																															85.1

[a] The sequences used were those of the type strains representative of each species, when available (refer to Table 1), *M. leprae* (GenBank accession number X73822), *M. album* ATCC 29676, *M. bovis* ATCC 35720, and *M. nonchromogenicum* ATCC 19531.

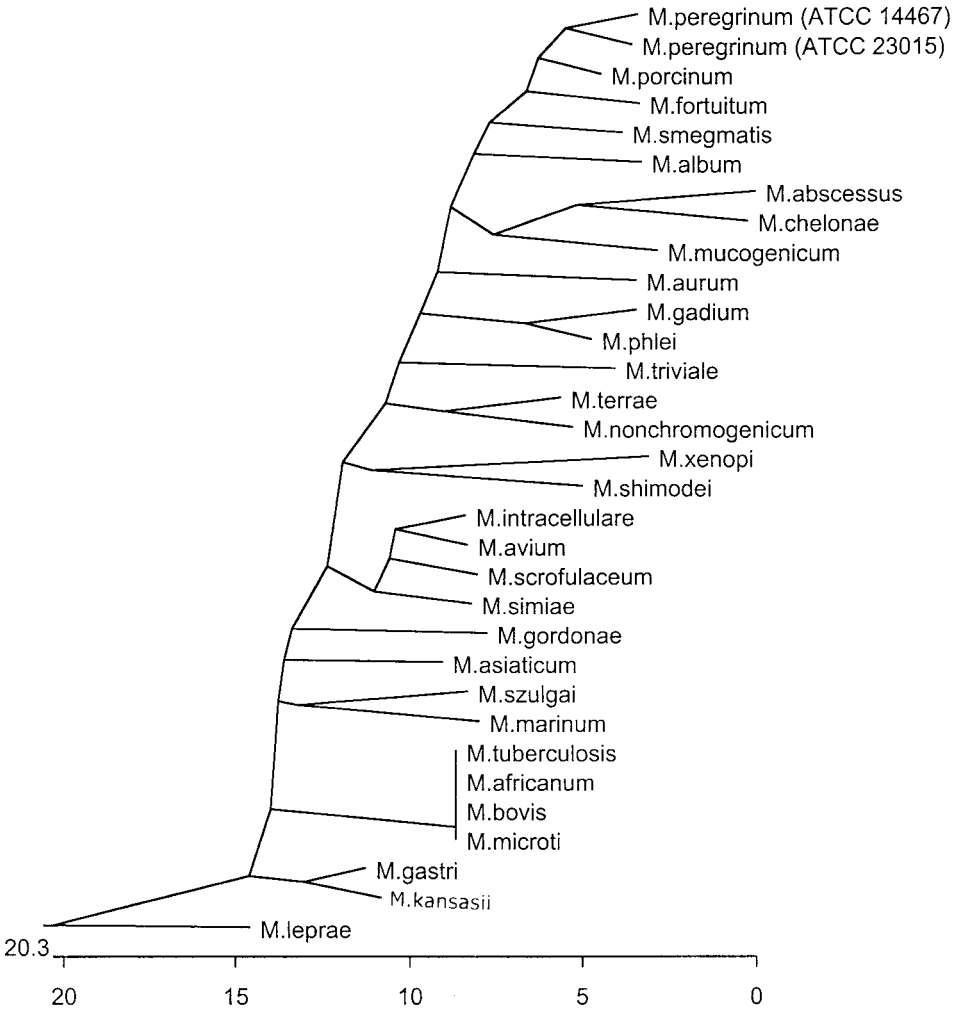

FIG. 2. Phylogenetic relationships of 31 *Mycobacterium* species derived from the sequence of the *recA* gene (excluding the intein regions). The tree was generated from the alignment obtained with the Megalign tool of Lasergene (DNAstar Inc.) with the Clustal method algorithm.

extra glutamine (2, 3). In this study, we have also found that *M. fortuitum*, *M. peregrinum*, *M. album*, *M. porcinum*, *M. aurum*, and *M. mucogenicum* also contain the extra glutamine at the exact same location. These are all nonchromogenic rapid growers (8). The sequence alignment of all species tested revealed that these species share a very high degree of *recA* sequence similarity and are clustered together on the phylogenetic tree, apart from all slow growers.

Comparison of the results of *recA* gene-based sequence analysis and the results of 16S rRNA gene-based sequence analysis revealed a general likeness in the relative position of each species within the tree, with a division between the rapid growers and slow growers (Fig. 2) (17). Excluding members of the *M. tuberculosis* complex, there is greater variability between the species by *recA* gene sequence analysis (from 75.7 to 96.2%) than by 16S rRNA gene sequence analysis (94.3 to 100%) (17). 16S rRNA sequence analysis is unable to distinguish *M. kansasii* from *M. gastri* due to their identical sequences, whereas our preliminary findings demonstrate that these species can be differentiated by the *recA* gene sequence analysis, which indicated 96.2% similarity. While the two species can be differentiated on the basis of their photochromogenicities, *recA* gene sequence analysis can prove to be useful in cases in which identification may rely exclusively on molecular methods when growth is unsuccessful. Since *M. kansasii* is considered clinically significant, whereas *M. gastri* is not, the distinction between the two is essential (18). Furthermore, *M. bovis* and *M. marinum*, which are 99.4% similar by 16S rRNA gene sequence analysis (17), are 89.5% similar by *recA* gene sequence analysis (Table 2). Alternatively, upon sequencing of the species in the *M. tuberculosis* complex, it was found that the sequences of *M. tuberculosis*, *M. bovis*, *M. microti*, and *M. africanum* were identical and that *recA* gene sequencing would therefore not serve as a method for differentiation between the members of this complex.

The other focus of this study was evaluation of the degree of intraspecies variation of the *recA* gene. The *recA* gene sequence was available for two reference strains of each of 27 of the 31 species tested (Table 1). The sequencing results revealed that overall the intraspecies sequence variation is insignificant: the greatest variation was observed for *M. scrofulaceum* (ATCC 19981 and ATCC 35788) and *M. mucogenicum* (ATCC 49650 and ATCC 49651), with 1.3% variations for each species. All other species have 0.9 to 0% divergence. The slight divergence seen within the same species could demonstrate the presence of *recA* alleles. Several species of *Mycobac-*

terium have been found to exhibit a number of 16S rRNA and/or *rpoB* alleles (7, 10, 14, 16). As suggested previously for these two genes, the use of only one reference strain for each species may not provide a large enough wealth of information for the accurate identification of species and subspecies (1, 7). Although these findings are generated from a small data set, the *recA* sequences derived from American Type Culture Collection reference strains, including, when possible, the type strain of each species, allude to the viability of this method. These sequences can be used as a foundation on which to base a database system, analogous to the 16S rRNA gene system, with reference type strains serving as standards for identification.

We have found in this study that, in general, the interspecies deviation of the *recA* gene sequence was more than 3.8%, while the intraspecies variation was less than 1.3%, with one exception: *M. peregrinum*. Strains ATCC 14467^T and ATCC 23015 of *M. peregrinum* show 3.9% divergence, the same divergence that exists between *M. gastri* and *M. kansasii*, which cannot be differentiated by 16S rRNA gene sequence analysis. Comparison of the 16S rRNA gene sequences of *M. peregrinum* ATCC 14467^T (GenBank accession number AF130308) and ATCC 23015 performed in our laboratory showed that they had 100% similarity. In addition, the biochemical profile determined in our laboratory indicated that the profiles of these two strains follow that of *M. peregrinum*: 3-day arylsulfatase positivity, positivity for iron uptake and nitrate, tolerance of 5% NaCl, no growth at 42°C, and fermentation of mannitol (9). However, ATCC 23015 has a positive 10-day Tween test result and does not develop a dark pink color, unlike the type strain, when growing on MacConkey agar without crystal violet. The percent similarity of the *recA* genes between these two strains would suggest that they may perhaps, like *M. kansasii* and *M. gastri*, be two closely related species or subspecies, as they do occupy similar positions in the *M. fortuitum* cluster of the phylogenetic tree shown in Fig. 2. This does not, however, eliminate the possibility that the discrepancy may be due to the presence of different copies of *recA* in the genome.

Previous studies have indicated the presence of an extraneous DNA sequence known as a posttranslationally self-splicing protein intron (or intein) in the *recA* genes of *M. tuberculosis* and *M. leprae* (2, 3). As of yet, the function of this element is unknown, but it has been hypothesized that the intein contributes to a novel regulatory mechanism needed for *recA* expression rather than being just a selfish element (2). After first reporting the existence of an insertion element in the *M. tuberculosis recA* gene in 1991 (3), Davis and coworkers (2) discovered the presence of a protein intron in *M. leprae*, another highly pathogenic *Mycobacterium* species. Their study indicated that the insertion elements in the *recA* genes of these two species differ in size, sequence, as well as location with respect to the gene, which led them to propose that the acquisition of the inteins in these two species occurred through independent pathways. Furthermore, they concluded that the intein is confined to *M. tuberculosis* and *M. leprae* only on the basis of the fact that none of the other *Mycobacterium* species that they studied contains the insertion. However, we report on the presence of an intein in other *Mycobacterium* species, including *M. xenopi*, *M. shimoidei*, and *M. asiaticum*, none of which were included in the study of Davis et al. (2). Predictably, all members of the *M. tuberculosis* complex tested in this study were found to contain the intein. The insertion site of the intein of *M. xenopi* was identical to that of *M. leprae*, and the sizes differed by one amino acid residue. Despite their similarities in size and location, the sequences of the protein introns of these two species were quite different (86% protein sequence homology). By determination of partial sequences, the sequences of the inteins of *M. shimodei* and *M. asiaticum* also appeared to resemble that of the intein of *M. leprae*. On the basis of these results, we conclude, first, that *M. leprae*, *M. xenopi*, *M. asiaticum*, and *M. shimoidei* acquired the protein intron through a common pathway that differs from the mechanism used by *M. tuberculosis*. Second, the sequence variation provides little information on whether these species acquired the intein simultaneously from a common source or from other intron-containing species at various stages of their evolution. In any case, the high degree of sequence variation among inteins of the same origin suggests that the acquisition of the intein may have occurred early in the evolution of these species (2) and that significant sequence variation is the accumulated product of spontaneous mutations over the generations.

The discovery of the intein in the *recA* genes of the two most pathogenic species of mycobacteria had led Davis and coworkers (2) to propose a selection hypothesis. They hypothesized that the presence of the intein in two major human pathogens raises the possibility that inteins play a role in intracellular survival or pathogenesis (2). Both *M. tuberculosis* and *M. leprae* are pathogenic slow growers. This is also true of *M. xenopi*, *M. shimoidei*, and *M. asiaticum*, which seems to support the notion of selection among pathogenic species. We predict that a more extensive study of *recA* may reveal the existence of more intron-containing *Mycobacterium* species.

A target gene must be sufficiently conserved among strains of species for use for genotypic identification. We conclude that the *recA* gene of *Mycobacterium* species is less conserved at the nucleic acid level than the 16S rRNA gene and is thus potentially useful for identification. Phylogenetic analysis based on *recA* gene sequence analysis grouped the *Mycobacterium* species generally in the same way as that based on phylogenetic studies by 16S rRNA gene sequence analysis. Therefore, we believe that *recA* sequencing can be used in conjunction with 16S rRNA-based species identification, particularly when 16S rRNA sequences share high similarity values.

REFERENCES

1. **Clayton, R. A., G. Sutton, P. S. Hinkle, C. Bult, and C. Fields.** 1995. Intraspecific variation in small-subunit rRNA sequences in GenBank: why single sequences may not adequately represent prokaryotic taxa. Int. J. Syst. Bacteriol. **45:**595–599.
2. **Davis, E. O., H. S. Thangaraj, P. C. Brooks, and M. J. Colston.** 1994. Evidence of selection for protein introns in the RecAs of pathogenic mycobacteria. EMBO J. **13:**699–703.
3. **Davis, E. O., S. G. Sedgwick, and M. J. Colston.** 1991. Novel structure of the *recA* locus of *Mycobacterium tuberculosis* implies processing of the gene product. J. Bacteriol. **173:**5653–5662.
4. **Duwat, P., S. D. Ehrlich, and A. Gruss.** 1992. A general method for cloning *recA* genes of gram-positive bacteria by polymerase chain reaction. J. Bacteriol. **174:**5171–5175.
5. **Dybvig, K., S. K. Hollingshead, D. G. Heath, D. B. Clewell, F. Sun, and A. Woodard.** 1992. Degenerate oligonucleotide primers for enzymatic amplification of *recA* sequences from gram-positive bacteria and mycoplasmas. J. Bacteriol. **174:**2729–2732.
6. **Eisen, J. A.** 1995. The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. J. Mol. Evol. **41:**1105–1123.
7. **Gingeras, T. R., G. Ghandour, E. Wang, A. Berno, P. E. Small, F. Drobniewski, D. Alland, E. Desmond, M. Holodniy, and J. Drenkow.** 1998. Simultaneous genotyping and species identification using hybridization pattern recognition analysis of generic *Mycobacterium* DNA arrays. Gene Res. **8:**435–448.
8. **Goodfellow, M., and J. G. Magee.** 1998. Taxonomy of mycobacteria, p. 1–49. *In* P. R. J. Gangadharam and P. A. Jenkins (ed.), Mycobacteria I. Basic aspects. Chapman & Hall, New York, N.Y.
9. **Heifets, L. B., and P. A. Jenkins.** 1998. Speciation of mycobacteria in clinical laboratories, p. 308–350. *In* P. R. J. Gangadharam and P. A. Jenkins (ed.), Mycobacteria I. Basic aspects. Chapman & Hall, New York, N.Y.
10. **Kim, B.-J., S.-H. Lee, M.-A. Lyu, S.-J. Kim, G.-H. Bai, S.-J. Kim, G.-T. Chae, E.-C. Kim, C.-Y. Cha, and Y.-H. Kook.** 1999. Identification of mycobacterial

species by comparative sequence analysis of the RNA polymerase gene (*rpoB*). J. Clin. Microbiol. **37:**1714–1720.
11. **Kowalczykowski, S. C., D. A. Dixon, A. K. Eggleston, S. S. Lauder, and W. M. Rehrauer.** 1994. Biochemistry of homologous recombination in Escherichia coli. Microbiol. Rev. **58:**401–465.
12. **Metchock, B. G., F. S. Nolte, and R. J. Wallace, Jr.** 1999. Mycobacterium, p. 399–437. *In* P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover, and R. H. Yolken (ed.), Manual of clinical microbiology, 7th ed. American Society for Microbiology, Washington, D.C.
13. **Miller, R. V., and T. A. Kokjohn.** 1990. General microbiology of *recA*: environmental and evolutionary significance. Annu. Rev. Microbiol. **44:**365–394.
14. **Ninet, B., M. Monod, S. Embler, J. Pawlowski, C. Metral, P. Rohner, R. Auckenthaler, and B. Hirschel.** 1996. Two different 16S rRNA genes in a mycobacterial strain. J. Clin. Microbiol. **34:**2531–2536.
15. **Papavinasasundaram, K. G., F. Movahedzadeh, J. T. Keer, N. G. Stoker, M. J. Colston, and E. O. Davis.** 1997. Mycobacterial *recA* is cotranscribed with a potential regulatory gene called *recX*. Mol. Microbiol. **24:**141–153.
16. **Reischl, U., K. Feldmann, L. Naumann, B. J. M. Gaugler, B. Ninet, B. Hirschel, and S. Emler.** 1998. 16S rRNA sequence diversity in *Mycobacterium celatum* strains caused by presence of two different copies of 16S rRNA gene. J. Clin. Microbiol. **36:**1761–1764.
17. **Rogall, T., J. Wolters, T. Flohr, and E. C. Bottger.** 1990. Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*. Int. J. Syst. Bacteriol. **40:**323–330.
18. **Wayne, L. G., and H. A. Sramek.** 1992. Agents of newly recognized or encountered mycobacterial diseases. Clin. Microbiol. Rev. **5:**1–25.