Extensive DNA Sequence Conservation throughout the Mycobacterium tuberculosis Complex

RICHARD FROTHINGHAM,^{1,2*} HAROLD G. HILLS,³[†] AND KENNETH H. WILSON^{1,2}

Infectious Diseases Section¹ and Research Service,³ Durham Veterans Affairs Medical Center, Durham, North Carolina 27705, and Division of Infectious Diseases, Department of Medicine, Duke University Medical Center, Durham, North Carolina 277102

Received 10 December 1993/Returned for modification 9 February 1994/Accepted 2 April 1994

The Mycobacterium tuberculosis complex includes the four species M. tuberculosis, M. bovis, M. africanum, and M. microti. We sequenced ¹³ M. tuberculosis complex strains in the 16S-to-23S rDNA internal transcribed spacer (ITS). The ITS has a high rate of nucleotide substitution. Previous reports found three nucleotide substitutions in the ITS between two M. tuberculosis complex strains. In contrast, we found the same ITS sequence in all 13 *M. tuberculosis* complex strains (including all four species and *M. bovis* BCG). This finding confirms the conservation of 16S rDNA sequence and the high DNA-DNA relatedness found in previous studies. By the usual criteria, the four species of the M. tuberculosis complex would be considered a single species. In a phylogenetic analysis based on the ITS sequence, the four species of the M. tuberculosis complex were distinct from nontuberculous mycobacteria. The ITS contains at least seven potential sites for oligonucleotide probes with specificity for the M. tuberculosis complex.

The *Mycobacterium tuberculosis* complex includes the four species M. tuberculosis, M. bovis, M. africanum, and M. microti. These species are genetically very similar, with 85 to 100% DNA-DNA relatedness (10), but they differ in their epidemiology. M. tuberculosis causes tuberculosis in humans and animals which have contact with humans. M. africanum causes tuberculosis in humans in Africa (22) . *M. bovis* is found in a wide range of wild and domestic animals, while M. microti causes tuberculosis in voles (22).

Slowly growing mycobacteria have ^a single rRNA operon (11, 13, 18). The operon is transcribed as a unit; a single transcript is then processed into mature 16S, 23S, and 5S rRNAs (11). The 16S rRNA gene (ribosomal DNA [rDNA]) is widely used for phylogenetic analysis at the genus and species levels. However, the internal transcribed spacer (ITS) between 16S rDNA and 23S rDNA is much more variable than 16S rDNA. In slowly growing mycobacteria, the ITS is about 280 bp long and does not contain the tRNA genes which are often found in other bacteria (18). The high nucleotide substitution rate of the ITS makes it a good site for phylogenetic analysis and a good target for diagnostic oligonucleotide probes directed at closely related species or strains (1, 8, 9, 13). The ITSs of 60 strains of eight species of slowly growing mycobacteria have been sequenced (8, 9, 11, 13, 18). Three ITS nucleotide substitutions were found between *M. tuberculosis* H37R v^T (the T indicates type strain) and M . bovis BCG Calmette (11, 18).

We sequenced the ITSs of 13 strains of the *M. tuberculosis* complex. The strains chosen for analysis included the type strain of each species as well as a strain of M. bovis BCG. Ten of the 13 strains were previously analyzed by restriction endonuclease digestion of chromosomal DNA and then Southern hybridization with the insertion element IS6110 (4). Two of the strains chosen had identical IS6110 fingerprints (TMC 410 and TMC 1024), two of the strains had fingerprints which

differed slightly (TMC ¹⁰² and TMC 201), and other strains had highly variable fingerprints (e.g., T2, T3, and T5). The strains chosen were thus expected to represent strains within the M. tuberculosis complex with various degrees of genetic conservation and variability.

(This work was presented in abstract form at the 31st Annual Meeting of the Infectious Diseases Society of America, New Orleans, La., 16 to ¹⁸ October 1993 [7].)

MATERIALS AND METHODS

ITS amplification by PCR. Heat-killed lysates of the type strains of M. africanum and M. microti were obtained from the Clinical Microbiology Laboratory of the Veterans Affairs Medical Center, Durham, N.C. DNAs from M. tuberculosis and M. bovis were obtained from the University of Arkansas for Medical Sciences, Little Rock (4). The ITS was amplified by using ³⁰ cycles of PCR as described previously (8). The primers (T3-Ecl6S.1390p and T7-Mb23S.44n) were complementary at their ³' ends to conserved regions of 16S and 23S rDNAs, which flank the ITS (8). Each primer also incorporated a T3 or T7 phage RNA polymerase promoter at the ⁵' end (not used in the present work). Either purified DNA (1 to ⁵⁰ ng) or ^a broth culture lysate (10 μ l) was used in each 100- μ l reaction mixture. Samples were mixed with PCR reagents in ^a room not used for PCR product analysis. Multiple interspersed negative controls (reagents only, no DNA) were included. Successful amplification was confirmed by electrophoresis on ethidium bromidestained 2% agarose gels. No amplification products were observed in the negative controls. Positive amplification controls were 100 ng of DNA from *M. avium* ATCC 25291^T.

Direct sequencing of PCR products. PCR products were purified by using the GENECLEAN II Kit (Bio-101, La Jolla, Calif.) and were sequenced by using the Taq Dye Deoxy Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, Calif.). The sequencing primers included the complementary portions of the PCR primers (Ecl6S.1390p and Mb23S.44n) and the internal primer Mtl6S.1501p (5'-CCGTA CCGGAAGGTGCGGCTG-3'). Each sequencing reaction included ¹⁰⁰ ng of PCR product and ¹⁰ pmol of sequencing

^{*} Corresponding author. Mailing address: Durham VA Medical Center, ⁵⁰⁸ Fulton St., Building 4, Durham, NC 27705. Phone: (919) 286-0411, ext. 6566. Fax: (919) 286-6824.

^t Present address: 1184 Molecular Biology Building, Iowa State University, Ames, Iowa 50011-3260.

primer in a total volume of 20 μ l. Each reaction underwent 25 cycles (94 \degree C for 30 s, 45 \degree C for 2 s, 60 \degree C for 4 min) in the GeneAmp PCR System 9600 thermal cycler (Perkin-Elmer, Norwalk, Conn.). The sequencing products were purified through Centri-Sep columns (Princeton Separations, Adelphia, N.J.) and were loaded onto 6% denaturing polyacrylamide gels in an automated sequencer (373A DNA Sequencer; Applied Biosystems). Sequences were combined by using the ASSEMGEL program of PCGENE release 6.5 (IntelliGenetics, Mountain View, Calif.). Discrepancies were resolved by examination of chromatographs from the automated sequencer.

Phylogenetic analysis. Sequences were aligned by using the CLUSTAL program of PCGENE. Previously published mycobacterial ITS sequences were included in the alignment (8, 9, 11, 13, 18). The alignment was adjusted manually. Only the ITS sequence (excluding the flanking portions of 16S and 23S rDNA sequences) was used for phylogenetic analysis. The phylogenetic tree reported here was constructed by using parsimony analysis with ^a branch-and-bound search by using PAUP version 3.1.1 (19). The data were resampled by using 100 bootstrap replications (19). Other phylogenetic trees were generated by using PHYLIP version 3.5c (5) and the following methods: maximum likelihood, Fitch-Margoliash, neighborjoining, and parsimony.

Nucleotide sequence accession numbers. The sequences of the type strains of M. tuberculosis (9), M. bovis, M. africanum, and M. microti and of M. bovis BCG Glaxo were submitted to GenBank. The accession numbers are listed in Table 1. GenBank accession numbers for other sequences are as follows: Mav-A, L07855; Mav-B, L07856; Mav-C, L07857; Mav-D, L07858; Min-A, L07859; MAC-A, L07847; MAC-B, L07848; MAC-C; L07849; MAC-D, L07851, MAC-E, L07852; MAC-F, L07853; Msc-B, L07854; MAC-H, L15620; Mce-A, L15621; Msc-A, L15622; Mxe-A, L15624; Mle-A, X56657.

RESULTS

Sequence results. Complete ITS sequences together with flanking partial 16S and 23S rDNA sequences were obtained for the 13 strains listed in Table 1. The sequences of all 13 M. tuberculosis complex strains were identical. The sequences of the four type strains and M. bovis BCG Glaxo are aligned in Fig. ¹ with the sequences of six nontuberculous mycobacteria (8, 9, 13).

Comparison with previously published M. tuberculosis complex sequences. Other investigators have reported the sequences of the ITSs and flanking rDNAs for M. tuberculosis $\hat{H}37Rv^{T}$ and *M. bovis* BCG Calmette (11, 18). These two sequences are shown in Fig. 1. They differ from each other at 16S rDNA position 1444 and at ITS positions 12, 13, and 104, and both have an extra G after ITS position ¹⁷³ in comparison with the sequences of our *M. tuberculosis* complex strains. The discrepancies probably represent sequencing errors or cloning artifacts in the previous reports. We obtained the same sequence in all 13 M. tuberculosis complex strains, including one of the previously published strains $(M.$ tuberculosis H37Rv^T). The nucleotides that we identified at the discrepant ITS positions were present in most of the other mycobacterial species whose sequences are known (Fig. 1). Compressions in manual sequencing gels may have caused some of the discrepancies (e.g., GC versus CG, ⁵ G residues versus ⁴ G residues). We avoided the problem of cloning artifacts by direct sequencing (6).

Comparison of M. tuberculosis complex sequences with those of other mycobacteria. Sequences from six nontuberculous

TABLE 1. M. tuberculosis complex strains sequenced in the 16S-to-23S rDNA internal transcribed spacer

	Strain name	Accession no. ^{<i>a</i>}			
Species		TMC	ATCC	GenBank ^b	
M. tuberculosis	$H37Rv$ (type) H37Ra T1 ^c T2 ^c T3 ^c T4 ^c T5 ^c	102 201	27294 25177	L ₁₅₆₂₃	
M. bovis	Type Ravenel $AN-5$ BCG Glaxo	410 401 412 1024	19210 35720 35726 35741	L ₂₆₃₂₇ L ₂₆₃₂₈	
M. africanum	Type		25420	L ₂₆ 330	
M. microti	Type	1619	19422	L26329	

" Abbreviations: TMC, Trudeau Mycobacterial Culture Collection; ATCC, American Type Culture Collection.

 b GenBank accession numbers for the five sequences which were submitted are</sup> listed.

Recent clinical isolate (4).

mycobacterial species are included in the alignment in Fig. 1. The ³' ends of the 16S rDNAs were relatively well conserved among these species (M. tuberculosis 16S rDNA positions ¹⁴⁰¹ to 1536). In contrast, the ITSs had many nucleotide substitutions, insertions, and deletions. The highly variable sequence of the ITS makes it a good target for diagnostic oligonucleotide probes (1, 8, 13). Figure ¹ shows seven potential probe sites in the ITS for the M. tuberculosis complex. The sequences of more mycobacterial species are needed to confirm the specificities of these sites. Table 2 lists the numbers of nucleotide differences between pairs of ITS sequences from various species of slowly growing mycobacteria. The four species of the M. tuberculosis complex were distinct from the other species, differing by at least 32 nucleotides.

The phylogenetic tree in Fig. 2 displays the relationships among the 18 different ITS sequences obtained from 71 strains of slowly growing mycobacteria (8, 9, 13). This tree was selected on the basis of maximum parsimony (19). Four other ITS trees were generated by other methods and showed minor variations, especially in the placement of the M. scrofulaceum sequences. A recent phylogenetic analysis based on 16S rDNA included sequence data for all seven species listed in Table 2 (3). The branching order for those seven species was consistent in the 16S rDNA and ITS trees. This suggests that recombination does not occur frequently in the rRNA operon in slowly growing mycobacteria (8). The ITS tree has longer branch lengths than the 16S rDNA tree, reflecting the greater variability of the ITS sequence. ITS sequence analysis thus provides a potentially finer means of discriminating among closely related mycobacterial species or strains (8).

DISCUSSION

All 13 strains of the four species of the M. tuberculosis complex that we tested had identical ITS sequences, suggesting a very close evolutionary relationship. These four species also have identical 16S rDNA sequences (12) and identical partial dnaJ gene sequences (20). Together, these sequence results support the idea that the M. tuberculosis complex represents a

FIG. 1. Sequence alignment, including alignment of the sequences of the complete 16S-to-23S rDNA ITS and flanking partial 16S and 23S rDNAs. The sequences of only 5 of the 13 M. tuberculosis complex strains sequenced are displayed. The sequence of the M. tuberculosis type strain is given in full. The dots represent identity with that sequence. The dashes represent deletions. Potential oligonucleotide probe sites with specificity for the *M. tuberculosis* complex are underlined. Previously published sequences which differ from the sequences obtained in the present study are bracketed. Abbreviations corresponding to the phylogenetic tree in Fig. 2 follow each sequence line (see Fig. 2 legend for definitions of
abbreviations). ^T, type strain; BCG, bacillus Calmette-Guérin; Gl, Glaxo; Ca, Calm C or G.

TABLE 2. Nucleotide differences between pairs of 16S-to-23S rDNA ITS sequences

	Species	No. of nucleotide differences from species:						
		M. tuber- culosis	М.	avium cellulare	M. intra- M. scrofu- laceum	М. leprae	М. celatum	
	M. tuberculosis a,b							
	$M.$ avium ^b	51						
	M. intracellulare ^b	48	17					
	M. scrofulaceum ^b	32	28	26				
	M. leprae	41	50	44	38			
	M. celatum	53	53	51	44	59		
	$M.$ xenopi b	58	66	67	60	64		

 a There were no nucleotide differences between M . tuberculosis, M . bovis, M . africanum, and M. microti.

Type strain.

single species. Generally, a bacterial species includes strains which have 70% or greater DNA-DNA relatedness (21). The four species of the M. tuberculosis complex have DNA-DNA relatedness of 85 to 100% (10), and so they would be considered a single species by this criterion. The division of the M. tuberculosis complex into four species is probably an artifact of the great historical interest in this pathogen.

Even when the members of the M. tuberculosis complex are considered a single species, their sequences are more conserved than those of some other mycobacterial species. Figure 2 contrasts the sequence conservation of the M. tuberculosis complex with the sequence variability of the M. avium complex (the 12 ITS sequences from Mav-A to Mav-E). The M. avium complex includes the species M. avium, M. intracellulare, and

FIG. 2. Phylogenetic tree of ¹⁸ complete 16S-to-23S rDNA ITS sequences obtained from 71 mycobacterial strains. The numbers in parentheses indicate the number of strains with each sequence. Multiple isolates with the same sequence associated with the same patient are counted as one strain (9). The following abbreviations are used: Mav, M. avium; MAC, M. avium complex; Min, M. intracellulare; Msc, M. scrofulaceum; Mle, M. leprae; Mtb, M. tuberculosis; Mbo, M. bovis; Maf, M. africanum; Mmi, M. microti; Mce, M. celatum; Mxe, M. xenopi. Horizontal lengths represent genetic distances; bar, 10 nucleotide (nt) differences. Vertical lengths are not meaningful. Sequences grouped on the same branch are hypothesized to share a common ancestry. The percentage of bootstrap replications which yielded each grouping is indicated. This percentage is a measure of how well the sequence data support the grouping.

strains not assigned to either species. Strains of the M. avium complex differ in their ITS sequences by up to 21 nucleotides. Also, two strains of M. scrofulaceum differ in their ITS sequences by five nucleotides (Msc-A and Msc-B in Fig. 2 [8, 9]). Strains of M. intracellulare, M. kansasii, M. gordonae, and M. celatum have substantial sequence variability even in the more conserved 16S rDNA (3, 12, 15).

Several alternate explanations may account for the high degree of sequence conservation among M. tuberculosis complex strains. First, the M. tuberculosis complex may have evolved relatively recently. Second, the prolonged clinical latency of tuberculosis may lead to a slow evolutionary clock. Third, members of the *M. tuberculosis* complex are obligate parasites. In contrast, other mycobacterial species are widely distributed in the environment. This may result in a smaller total number of M. tuberculosis complex organisms and fewer opportunities for mutations. Fourth, bacteria in the M. tuberculosis complex normally live inside host macrophages. This limits the opportunity for recombination with other species or strains and may result in genetic homogeneity. Fifth, the M. tuberculosis complex is presumably well adapted to the relatively constant environment of the host. Most mutations probably provide no selective advantage, and thus are not preserved. In contrast, other mycobacteria propagate in various environments, allowing selection of various mutations.

Members of the \overline{M} . tuberculosis complex are commonly identified by using a commercial chemiluminescent oligonucleotide probe (AccuProbe; Gen-Probe, Inc., San Diego, Calif.). The target of this probe is the hypervariable region of 16S $rRNA$ (2). Several recent reports have identified strains of M . c elatum and M . terrae which give false-positive results with the M. tuberculosis complex AccuProbe and which have sequences similar to that of the M . tuberculosis complex in that region $(2, 1)$ 16, 17, 23). The seven ITS probe sites identified in Fig. ¹ may be more specific for the M. tuberculosis complex. The ITS of one M . celatum strain which gives a positive M . tuberculosis complex AccuProbe result has been sequenced. It differs substantially from the M . tuberculosis complex at all of the potential ITS probe sites (Fig. 1). The ITS may be a useful target for the simultaneous identification of multiple mycobacterial species. PCR could be performed by using primers complementary to flanking regions of 16S and 23S rDNAs with specificity for the genus Mycobacterium. The PCR product could then be identified at the species level by hybridization with a panel of oligonucleotide probes, direct sequencing (12), or possibly, restriction endonuclease digestion (14).

ACKNOWLEDGMENTS

The work was supported by NIH grants AI07392 and A135230, the Durham VA Medical Center's Research Center on AIDS and HIV Infection, and the U.S. Department of Veterans Affairs.

We thank M. Donald Cave and Kathleen Eisenach of the University of Arkansas for Medical Sciences and Kathy H. McDonald, Nancy Myers, and Peter Zwadyk, Jr., of the Durham VA Medical Center for providing DNA or culture lysates from M. tuberculosis complex strains. We thank Murray Bennett for assistance in drawing Fig. 1.

REFERENCES

- 1. Barry, T., G. Colleran, M. Glennon, L. K. Dunican, and F. Gannon. 1991. The 16S/23S ribosomal spacer region as a target for DNA probes to identify eubacteria. PCR Methods Applications 1:51-56.
- 2. Butler, W. R., S. P. O'Connor, M. A. Yakrus, and W. M. Gross. 1994. Cross-reactivity of genetic probe for detection of Mycobacterium tuberculosis with newly described species Mycobacterium celatum. J. Clin. Microbiol. 32:536-538.
- 3. Butler, W. R., S. P. O'Connor, M. A. Yakrus, R. W. Smithwick,

B. B. Plikaytis, C. W. Moss, M. M. Floyd, C. L. Woodley, J. 0. Kilburn, F. S. Vadney, and W. M. Gross. 1993. Mycobacterium celatum sp. nov. Int. J. Syst. Bacteriol. 43:539-548.

- 4. Cave, M. D., K. D. Eisenach, P. F. McDermott, J. H. Bates, and J. T. Crawford. 1991. IS6110: conservation of sequence in the Mycobacterium tuberculosis complex and its utilization in DNA fingerprinting. Mol. Cell. Probes 5:73-80.
- 5. Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Computer program distributed by the University of Washington, Seattle.
- 6. Frothingham, R. 1991. The agent of bacillary angiomatosis. N. Engl. J. Med. 324:1512. (Letter.)
- 7. Frothingham, R., H. G. Hills, and K. H. Wilson. 1993. Genetic homogeneity among the four species of the Mycobacterium tuberculosis (Mtb) complex, abstr. 34, p. 8-A. Infect. Dis. Soc. Am. 31st Annu. Meet. Final Program Abstr. 1993. Infectious Diseases Society of America, Washington, D.C.
- 8. Frothingham, R., and K. H. Wilson. 1993. Sequence-based differentiation of strains in the Mycobacterium avium complex. J. Bacteriol. 175:2818-2825.
- 9. Frothingham, R., and K. H. Wilson. 1994. Molecular phylogeny of the Mycobacterium avium complex demonstrates clinically meaningful divisions. J. Infect. Dis. 169:305-312.
- 10. Imaeda, T. 1985. Deoxyribonucleic acid relatedness among selected strains of Mycobacterium tuberculosis, Mycobacterium bovis, Mycobactenium bovis BCG, Mycobactenum microti, and Mycobacterium africanum. Int. J. Syst. Bacteriol. 35:147-150.
- 11. Kempsell, K. E., Y. Ji, I. C. E. Estrada-G, M. J. Colston, and R. A. Cox. 1992. The nucleotide sequence of the promoter, 16S rRNA and spacer region of the ribosomal RNA operon of Mycobacterium tuberculosis and comparison with Mycobacterium leprae precursor rRNA. J. Gen. Microbiol. 138:1717-1727.
- 12. Kirschner, P., B. Springer, U. Vogel, A. Meier, A. Wrede, M. Kiekenbeck, F.-C. Bange, and E. C. Bottger. 1993. Genotypic identification of mycobacteria by nucleic acid sequence determination: report of a 2-year experience in a clinical laboratory. J. Clin. Microbiol. 31:2882-2889.
- 13. Liesack, W., S. Sela, H. Bercovier, C. Pitulle, and E. Stackebrandt. 1991. Complete nucleotide sequence of the Mycobacterium leprae 23S and 5S rRNA genes plus flanking regions and their potential in designing diagnostic oligonucleotide probes. FEBS Lett. 281: 114-118.
- 14. Plikaytis, B. B., B. D. Plikaytis, M. A. Yakrus, W. R. Butler, C. L. Woodley, V. A. Silcox, and T. M. Shinnick. 1992. Differentiation of slowly growing Mycobacterium species including Mycobacterium tuberculosis, by gene amplification and restriction fragment length polymorphism analysis. J. Clin. Microbiol. 30:1815-1822.
- 15. Ross, B. C., K. Jackson, M. Yang, A. Sievers, and B. Dwyer. 1992. Identification of a genetically distinct subspecies of Mycobacterium kansasii. J. Clin. Microbiol. 30:2930-2933.
- 16. Saubolle, M. A., G. D. Roberts, N. L. Goodman, T. Davis, and V. Jonas. 1993. Isolates of nontuberculous mycobacteria identified as members of tuberculosis complex by commercial genetic probes, abstr. U-93, p. 185. Abstr. 93rd Gen. Meet. Am. Soc. Microbiol. 1993. American Society for Microbiology, Washington, D.C.
- 17. Stockman, L., B. Springer, E. C. Böttger, and G. D. Roberts. 1993. Mycobactenium tuberculosis nucleic acid probes for rapid diagnosis. Lancet 341:1486. (Letter.)
- 18. Suzuki, Y., A. Nagata, Y. Ono, and T. Yamada. 1988. Complete nucleotide sequence of the 16S rRNA gene of Mycobacterium bovis BCG. J. Bacteriol. 170:2886-2889.
- 19. Swofford, D. L. 1993. PAUP: phylogenetic analysis using parsimony, version 3.1.1. Computer program distributed by the Illinois Natural History Survey, Champaign.
- 20. Takewaki, S.-I., K. Okuzumi, I. Manabe, M. Tanimura, K. Miyamura, K.-I. Nakahara, Y. Yazaki, A. Ohkubo, and R. Nagai. 1994. Nucleotide sequence comparison of the mycobacterial dnaJ gene and PCR-restriction fragment length polymorphism analysis for identification of mycobacterial species. Int. J. Syst. Bacteriol. 44:159-166.
- 21. Wayne, L G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, 0. Kandler, M. I. Krichevsky, L H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Trüper. 1987. Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. Int. J. Syst. Bacteriol. 37:463-464.
- 22. Wayne, L. G., and G. P. Kubica. 1986. The mycobacteria, p. 1435-1457. In P. H. A. Sneath and J. G. Holt (ed.), Bergey's manual of systemic bacteriology, vol. 2. The Williams & Williams Co., Baltimore.
- 23. Young, K. K. Y., P. Hewitt, Y. Jang, E. Ford, and E. Desmond. 1993. Sequence analysis of mycobacterial isolates with discrepant species identification, abstr. U-86, p. 184. Abstr. 93rd Gen. Meet. Am. Soc. Microbiol. 1993. American Society for Microbiology, Washington, D.C.