JOURNAL OF BACTERIOLOGY, May 1993, p. 2818–2825
0021-9193/93/102818-08$02.00/0

Vol. 175, No. 10

Sequence-Based Differentiation of Strains in the *Mycobacterium avium* Complex

RICHARD FROTHINGHAM AND KENNETH H. WILSON*

Infectious Disease Section, Durham Veterans Affairs Medical Center, Durham, North Carolina 27705, and Division of Infectious Disease, Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710

Received 17 December 1992/Accepted 3 March 1993

The complete 16S-23S rDNA internal transcribed spacer (ITS) was sequenced in 35 reference strains of the *Mycobacterium avium* complex. Twelve distinct ITS sequences were obtained, each of which defined a "sequevar"; a sequevar consists of the strain or strains which have a particular sequence. ITS sequences were identified which corresponded to *M. avium* (16 strains, four ITS sequevars) and *Mycobacterium intracellulare* (12 strains, one ITS sequevars). The other seven *M. avium* complex strains had ITS sequences which varied greatly from those of *M. avium* and *M. intracellulare* and from each other. The 16S-23S rDNA ITS was much more variable than 16S rDNA, which is widely used for genus and species identification. Phylogenetic trees based on the ITS were compatible with those based on 16S rDNA but were more detailed and had longer branches. The results of ITS sequencing were consistent with the results of hybridization with *M. avium* and *M. intracellulare* probes (Gen-Probe) for 30 of 31 strains tested. Serologic testing correlated poorly with ITS sequencing. Strains with the same sequence were different serovars, and those of the same serovar had different sequences. Sequencing of the 16S-23S rDNA ITS should be useful for species and strain differentiation for a wide variety of bacteria and should be applicable to studies of epidemiology, diagnosis, virulence, and taxonomy.

Strain differentiation within a particular species has many potential uses. Strain markers are useful for epidemiology, both for tracing the spread of particular strains among patients and for evaluating possible environmental sources of pathogens. A strain marker may also define groups of strains which differ in virulence or pathophysiology. For example, many enterohemorrhagic strains of *Escherichia coli* are identified on the basis of the distinctive serovars O157:H7 and O26:H11. In this case, serology provides a marker for strains which produce a distinct clinical entity (5). Lastly, strain markers may be helpful in studies of microbial ecology. Symbiotic *Frankia* strains, detected by oligonucleotide hybridization, differ in their competitiveness for nodulation in plant hosts (24).

There are many alternative methods for strain differentiation (e.g., serology, colony morphology, biochemical testing, antibiotic susceptibility, phage typing, restriction fragment length polymorphism [RFLP], and random amplified polymorphic DNA). An ideal strain differentiation method should be rapid, reproducible, and applicable to all strains.

The strains in the *Mycobacterium avium* complex (MAC) have gained importance in recent years because of the high prevalence of disseminated MAC infection in patients with AIDS. Serology has been the standard method for MAC strain differentiation, but it has several limitations. The assay is laborious and the sera are not generally available, so serologic testing of MAC strains is limited to a few laboratories. Many MAC strains are nontypeable (6, 34), either because of autoagglutination or because their antigens are not recognized by currently available sera. The results of serologic testing have been compared with those of RFLP analysis and multilocus enzyme electrophoresis (11, 17, 31). Serologic testing correlated incompletely with these genotypic methods.

The taxonomy of the MAC has been discussed in two recent reviews (10, 32). *M. avium* and *Mycobacterium intracellulare* were originally defined by virulence in chickens but more recently have been defined on the basis of serovar. *M. avium* is currently considered to include serovars 1 through 6, 8 through 11, and 21 (23, 32). *M. intracellulare* includes serovars 7, 12 through 17, 19, 20, and 25 (23, 32). The status of the other serovars is still uncertain (32). The availability of specific hybridization probes for *M. avium* and *M. intracellulare* (Gen-Probe, Inc., San Diego, Calif.) has provided another method for defining the two species. Strains belonging to the serovars described above have generally yielded hybridization results consistent with their serovar status. However, within serovars 18, 22, 24, 26, and 28, some strains have hybridized with the *M. intracellulare* probe, while others have not hybridized with either probe (23).

Many investigators have used 16S rDNA sequence to differentiate bacteria at the genus and species level. However, 16S rDNA sequences do not vary greatly within a species. The internal transcribed spacer (ITS) between 16S rDNA and 23S rDNA is more variable than 16S rDNA (28). It is a single-copy gene region of about 280 bases in slow-growing mycobacteria (2). We sequenced the 16S-23S rDNA ITS for 35 MAC reference strains and compared the results with those achieved through 16S rDNA sequencing, serology, and the use of Gen-Probe hybridization probes.

MATERIALS AND METHODS

Sources of strains. Thirty-three MAC reference strains were obtained as frozen broth cultures from a collection at the Colorado State University, Fort Collins, Colo. (formerly housed at the National Jewish Hospital). The *M. avium* type

* Corresponding author.

strain (ATCC 25291) and the *M. intracellulare* type strain (ATCC 13950) were obtained from the Durham Veterans Administration Medical Center, Durham, N.C. A second isolate of the *M. intracellulare* type strain was obtained as part of a survey from the College of American Pathologists, Northfield, Ill.

Bacterial lysis. Bacterial lysis was done basically as previously described (8). The MAC strains were received frozen in broth culture. A volume of 500 μl of sterile TE (10 mM Tris, 1 mM EDTA [pH 8.0]) was added to a sterile 1.5-ml polypropylene microcentrifuge tube with a screw top and an O-ring seal (Starstedt, Nümbrecht, Germany). About 100 μl of acid-washed glass beads (25 to 50 μm in diameter) was added with a sterile spatula. The tube and its contents were irradiated on the floor of a UV Stratalinker 1800 (Stratagene, La Jolla, Calif.) for 20 min to inactivate any contaminating DNA. A volume of 100 μl of slightly turbid broth culture was transferred to the tube. The tube was closed tightly and chilled for 10 min on ice. The tube was agitated in a Mini-Beadbeater (Biospec Products, Bartlesville, Okla.) for 1 min to lyse the bacteria, releasing the DNA. The tube was again chilled on ice for 10 min and then centrifuged at 16,000 × *g* for 30 s in a microcentrifuge (Eppendorf 5415C; Eppendorf Gerätebau, Hamburg, Germany). The supernatant was divided into aliquots and stored at −20°C. A volume of 5 μl of the crude lysate was used in each 100-μl polymerase chain reaction (PCR). No precise quantification of the concentration of bacteria or DNA was attempted, as amplifications are successful over a wide range of DNA quantities.

DNA sequencing. The method used for DNA sequencing was an adaptation of gene amplification with transcript sequencing (8, 27). PCR primers incorporated promoters for phage RNA polymerases (T3 or T7), one promoter on each strand. The following primers were used to amplify the 16S-23S rDNA ITS: T3-Ec16S.1390p, 5′-GCGCAATTAACCCTCACTAAAGGGAATTGTACACACCGCCCGTCA-3′; and T7-Mb23S.44n, 5′-AATTTAATACGACTCACTATAGGGATCTCGATGCCAAGGCATCCACC-3′. The underlined portions of the two primers correspond to the T3 and T7 phage RNA polymerase promoters, respectively. The remainder of each primer was complementary to bacterial DNA. The primer nomenclature (20) includes the bacterial species, the name of a gene, the position in the gene, and the DNA strand homologous to the primer (positive or negative strand). Ec16S.1390p thus begins at position 1390 of the 16S rDNA gene of *E. coli* and is homologous to the sense strand. It is a modified inverse complement of primer C (13) and is highly conserved among bacteria. Mb23S.44n is homologous to the antisense strand of 23S rDNA for both *Mycobacterium bovis* and *Mycobacterium leprae* (16, 28). It is also well conserved among gram-positive bacteria. The locations of the PCR primers within the rRNA operon are shown in Fig. 1.

Each PCR mix contained 10 mM Tris (pH 8.3), 50 mM KCl, 0.01% gelatin, 1.2 mM $MgCl_2$, 100 nM each primer, 200 μM each of the four deoxynucleoside triphosphates (dNTPs), 2.5 U of *Taq* DNA polymerase (Amplitaq DNA polymerase; Perkin-Elmer Cetus, Norwalk, Conn.), and 5 μl of crude bacterial lysate in a total volume of 100 μl. All PCR components except the dNTPs and the sample were mixed and then irradiated for 2 min on the floor of a UV Stratalinker 1800 to inactivate any contaminating DNA. The efficiency of UV decontamination of PCR mixtures is greatly reduced when the UV-absorbent dNTPs are included (9). The dNTPs were added, the reaction mixtures were aliquoted, and samples were added. The mixture was layered with mineral oil and placed in a thermal cycler (Coy Laboratory Products, Ann Arbor, Mich.). The temperature was cycled to 92°C for 1 min (95°C for 2 min for the first two cycles; denaturation), then to 55°C for 2 min (annealing), and then to 75°C for 3 min (extension) for a total of 30 cycles. The PCR products were visualized on 2% agarose gels stained with ethidium bromide. Multiple interspersed negative controls were included whenever samples were amplified. Positive controls consisted of 100 ng of DNA from the *M. avium* type strain (ATCC 25291).

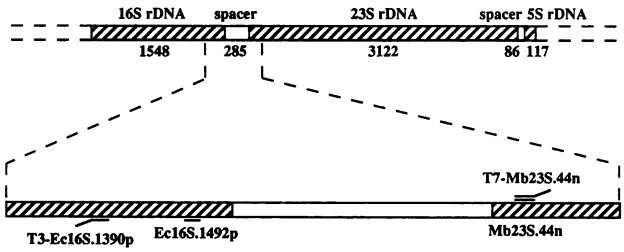

FIG. 1. Schematic map of the rRNA operon (*rrn*) of slowly growing mycobacteria. The distances are based on *M. leprae*, for which the entire operon has been sequenced. The locations of the PCR primers (T3-Ec16S.1390p and T7-Mb23S.44n) and the sequencing primers (Ec16S.1492p and Mb23S.44n) are indicated.

The PCR products incorporated the promoters for the T3 and T7 phage RNA polymerases which were present in the primers. Each PCR product was transcribed with each of the two RNA polymerases to produce two single-stranded RNA transcripts. In addition to providing single-stranded template, the transcription step amplified the number of copies by about 100. Each transcript was sequenced with reverse transcriptase and dideoxy chain termination. The sequencing primers included the complementary portion of one of the PCR primers, Mb23S.44n, and a conserved internal primer, Ec16S.1492p (Fig. 1). Ec16S.1492p (5′-AAGTCG TAACAAGGTA-3′) is the inverse complement of PC5, described previously (33). The details of the sequencing method were recently published (8).

Phylogenetic analysis. Sequences obtained from each strand were combined by using the ASSEMGEL program of PCGENE Release 6.5 (IntelliGenetics, Mountain View, Calif.). The sequences were aligned with the CLUSTAL program of PCGENE. Published sequences for the 16S-23S rDNA ITS of *M. bovis* BCG (28), *M. tuberculosis* H37Rv (12), and *M. leprae* Borstel strain (16) were included in the alignment. The alignment was adjusted manually. The phylogenetic trees were constructed by parsimony analysis with a branch-and-bound search with PAUP software (29). The branch-and-bound algorithm provides an efficient approach to finding all minimal-length trees. The trees were rooted by defining *M. bovis*, *M. tuberculosis*, and *M. leprae* as an outgroup. A majority-rule consensus was computed from all of the most parsimonious trees. The data were resampled with 100 bootstrap replications (7, 29). Each bootstrap replication represents a random resampling of the original data matrix. The trees shown represent the majority-rule consensus of these 100 bootstrap replications. In each case, they were identical to the majority-rule consensus of the most parsimonious trees. The percentage of bootstrap replications which yielded each grouping is indicated. This percentage represents a measure of statistical confidence. A grouping found on 95% of bootstrap replications is considered statistically significant (7).

TABLE 1. Results of 16S-23S rDNA ITS sequencing for 35 MAC reference strains[a]

Reference strain name	TMC no.	ATCC no.	Serovar[b] (species)	Gen-Probe hybridization[c]	16S-23S rDNA ITS sequevar
M. avium type strain	724	25291	2 (MA)	Not done	Mav-A
11907-300			1 (MA)	MA	Mav-A
B-92			1 (MA)	MA	Mav-A
14141-1395	715	35716	2 (MA)	Not done	Mav-A
6195			3 (MA)	MA	Mav-A
Sparrow 185	1463	35767	4 (MA)	MA	Mav-A[d]
34540-Wales			6 (MA)	MA	Mav-A
14186-1424	1462	35766	11 (MA)	MA	Mav-A[d]
13528-1079			4 (MA)	MA	Mav-B
25546-759			5 (MA)	MA	Mav-B
SJB #2			8 (MA)	MA	Mav-B
Borne	1461	35765	10 (MA)	MA	Mav-B[d]
1602-1965			10 (MA)	MA	Mav-B
2993			21 (MA)	MA	Mav-B
6194			2 (MA)	MA	Mav-C
17584-286	1479	35774	9 (MA)	MA	Mav-D[d]
W-552			19 (MI)	MI	MAC-A
M. intracellulare type strain	1406	13950	16 (MI)	Not done	Min-A[d]
157 Manten			7 (MI)	MI	Min-A
P-42	1405	35762	12 (MI)	MI	Min-A[d]
5509-Borstel		25122	13 (MI)	MI	Min-A
Edgar Boone	1403	35761	14 (MI)	MI	Min-A[d]
Dent	1473	35848	15 (MI)	MI	Min-A[d]
Yandle			16 (MI)	MI	Min-A
P-54	1411	35763	17 (MI)	MI	Min-A[d]
P-40	1419	35764	20 (MI)	MI	Min-A[d]
72-888			25 (MI)	MI	Min-A
1244 Hillberry			26 (?)	MI	Min-A
6845			28 (?)	MI	Min-A
12645			24 (?)	Neither	MAC-B
23393			23 (?)	Neither	MAC-C
Melnick	1467	35770	18 (?)	Both	MAC-D[d]
P-49	1476	35847	7 (MI)	Not done	MAC-E[d]
5154-O'Connor			22 (?)	Neither	MAC-F
Lane 3081			27 (?)	Neither	MAC-G

[a] Accession numbers are listed for strains belonging to the Trudeau Mycobacterial Culture Collection (TMC) and to the American Type Culture Collection (ATCC).

[b] Serovars are from references 18, 23, and 30. Species as defined by serovar (23, 32) to be *M. avium* (MA; serovars 1 through 6, 8 through 11, and 21), *M. intracellulare* (MI; serovars 7, 12 through 17, 19, 20, and 25), or presently undefined (?; serovars 18, 21 through 24, and 26 through 28).

[c] Hybridization with Gen-Probe probe for *M. avium* (MA) or *M. intracellulare* (MI), as determined by Saito et al. (23).

[d] Strains for which partial 16S rDNA sequence is also available (3).

Nucleotide sequence accession numbers. The sequences were submitted to GenBank and assigned accession numbers L07847 through L07849 and L07851 through L07859.

RESULTS

The 16S-23S rDNA ITS was amplified and sequenced for 35 MAC reference strains. The strains are listed in Table 1 together with their serovars (18, 23, 30), their species assignments on the basis of serovar (23, 32), the results of Gen-Probe hybridization (23), and the results of our ITS sequence analysis (below).

DNA from all 35 MAC reference strains was successfully amplified on the first attempt. No amplification products from any of the interspersed negative controls were observed. All amplification products were about 480 bases long, with no size differences detectable on the 2% agarose gels. A complete 16S-23S rDNA ITS sequence was obtained from all 35 MAC strains (Fig. 2) and was of high quality (ambiguity rate, <1%). The lengths of the 16S-23S rDNA ITS sequences from the MAC strains ranged from 278 to 281 bases. Small amounts of adjacent 16S rDNA and 23S rDNA sequence were also determined but were not sufficiently variable to be useful and were not included in the analysis.

From 35 MAC references strains, 12 distinct 16S-23S rDNA ITS sequences were obtained. Each distinct ITS sequence was used to define an infrasubspecific taxon, called a sequevar. The new word sequevar was chosen to correspond to other infrasubspecific terms (biovar, chemovar, morphovar, pathovar, phagovar, serovar, etc.), as described in the Bacteriological Code (14). Just as a serovar consists of the strain or strains which react with a certain serum, a sequevar was defined as the strain or strains which contain a given sequence. The suffix -var was used instead of -type

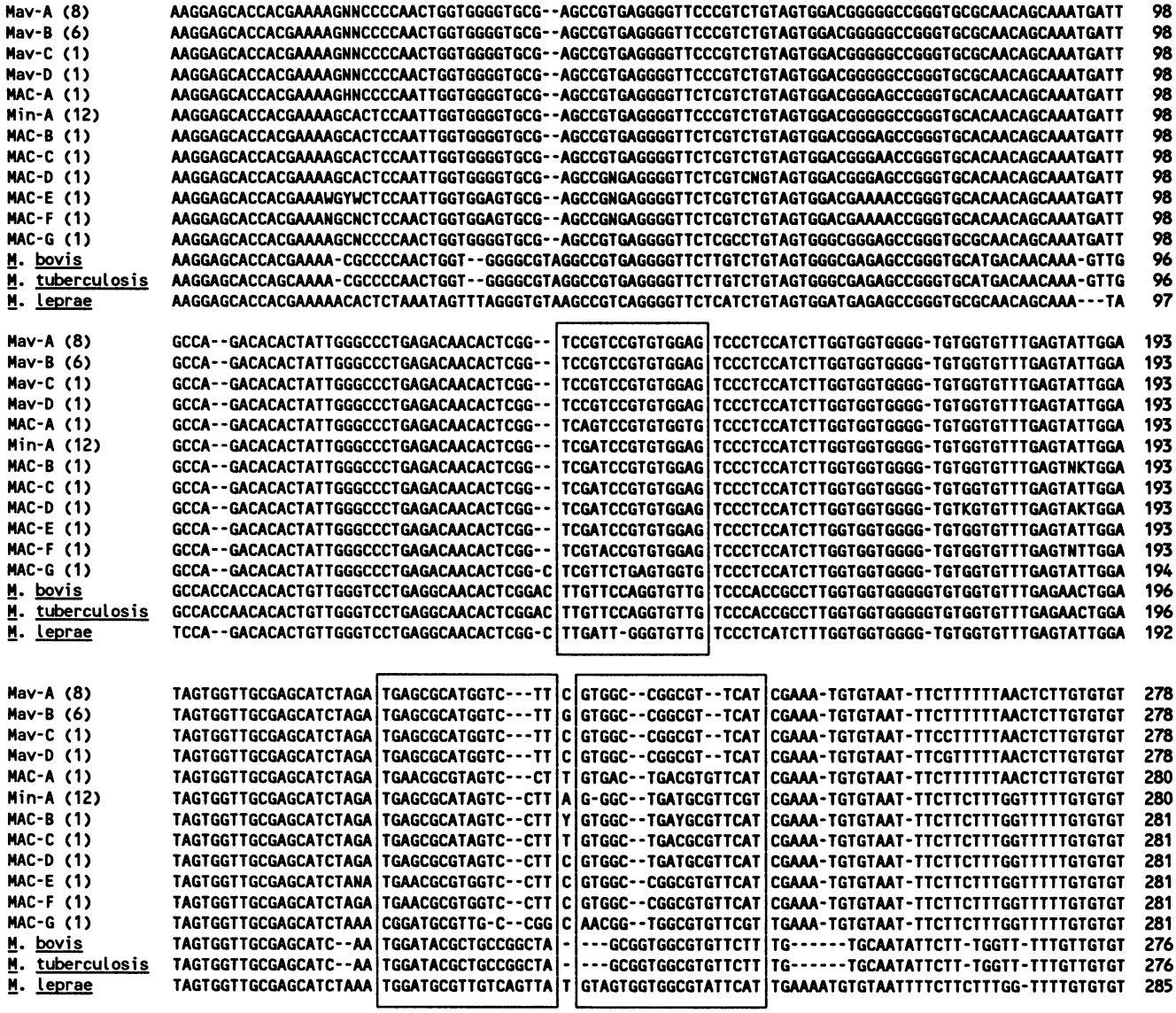

FIG. 2. Alignment of 16S-23S rDNA ITS sequences. The alignment includes 15 distinct sequences (each defining a sequevar) derived from the 35 MAC reference strains and three other mycobacterial species. The number of strains belonging to each sequevar is indicated in parentheses. Ambiguous positions represent sequencing uncertainties, not variations among strains within each sequevar. Potential probe-primer sites for *M. avium* are boxed. Each of these sites is conserved for all four *M. avium* sequevars (Mav-A to Mav-D) but has two or more mismatches to all other known ITS sequences.

(e.g., serotype) in accord with the recommendation of the Bacteriological Code that type be used strictly for nomenclatural type (14).

The 16S-23S rDNA ITS sequevar is listed for each strain in Table 1. The numbers of nucleotide differences between pairs of ITS sequences are listed in Table 2. The tree in Fig. 3 displays the phylogenetic relationships among the ITS sequevars. The tree should not be interpreted as a precise representation of evolutionary history. It displays phylogenetic groupings which are supported with various degrees of statistical confidence. ITS sequences from three other mycobacterial species were used to root the tree. Relative to these species, the MAC sequevars formed a well-defined group, which was found on 98% of bootstrap replications (Fig. 3).

***M. avium* reference strains.** The first 16 strains in Table 1 would be considered *M. avium* on the basis of hybridization with the Gen-Probe *M. avium* probe and on the basis of belonging to serovars 1 through 6, 8 through 11, or 21. These 16 strains fell into four 16S-23S rDNA ITS sequevars (Mav-A, Mav-B, Mav-C, and Mav-D). The ITS sequences of these four sequevars were very similar, differing by only one or two nucleotides in pairwise comparisons (Table 2). These four sequevars formed a well-defined group, confirmed on 99% of bootstrap replications (Fig. 3). Their ITS sequences differed by 14 or more nucleotides from those of all other sequevars (Table 2). Thus, the species *M. avium* as defined by serology and Gen-Probe hybridization was also well defined by sequence analysis.

***M. intracellulare* reference strains.** The *M. intracellulare* type strain (two isolates from different sources) had the same 16S-23S rDNA ITS sequence as 11 other reference strains

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

Sequevar or species	No. of nucleotide differences from sequevar or species:													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Mav-A	—													
2 Mav-B	1	—												
3 Mav-C	1	2	—											
4 Mav-D	1	2	1	—										
5 MAC-A	14	14	15	15	—									
6 Min-A	17	17	18	18	19	—								
7 MAC-B	16	17	17	17	14	4	—							
8 MAC-C	18	18	19	19	15	6	1	—						
9 MAC-D	18	19	19	19	15	5	1	4	—					
10 MAC-E	18	19	19	19	19	15	10	10	10	—				
11 MAC-F	18	19	19	19	21	18	13	13	13	3	—			
12 MAC-G	30	31	31	31	29	28	26	28	26	26	26	—		
13 *M. bovis*	51	51	52	52	48	53	50	51	50	48	47	40	—	
14 *M. tuberculosis*	52	52	53	53	49	54	51	52	51	49	48	41	3	—
15 *M. leprae*	48	48	49	49	46	47	42	43	43	39	42	41	46	47

(sequevar Min-A, Table 1). All of these strains hybridized strongly with the *M. intracellulare* probe (23). All of them belong to serovars which are considered *M. intracellulare* or which are still undefined (23, 32). The ITS sequence of sequevar Min-A differed by four or more nucleotides from the ITS sequences of all other sequevars (Table 2).

Other MAC reference strains. The other seven reference strains all yielded different 16S-23S rDNA ITS sequences (MAC-A to MAC-G), which varied considerably from one another (Fig. 3). Two of these strains belong to serovars considered to be *M. intracellulare* (W-552 and P-49), and one of these (W-552) also hybridized strongly with the *M. intracellulare* probe (Table 1). However, their sequences (MAC-A and MAC-E) were markedly different from each other and from the single sequence (Min-A) obtained from 12 other *M. intracellulare* strains (Fig. 3). The existence of MAC strains which do not fit well into either *M. avium* or *M. intracellulare* is well recognized (23), and the possibility of a third species has been raised (32). From the sequence diversity observed among these MAC reference strains, it is possible that several additional species exist within the *M. avium* complex.

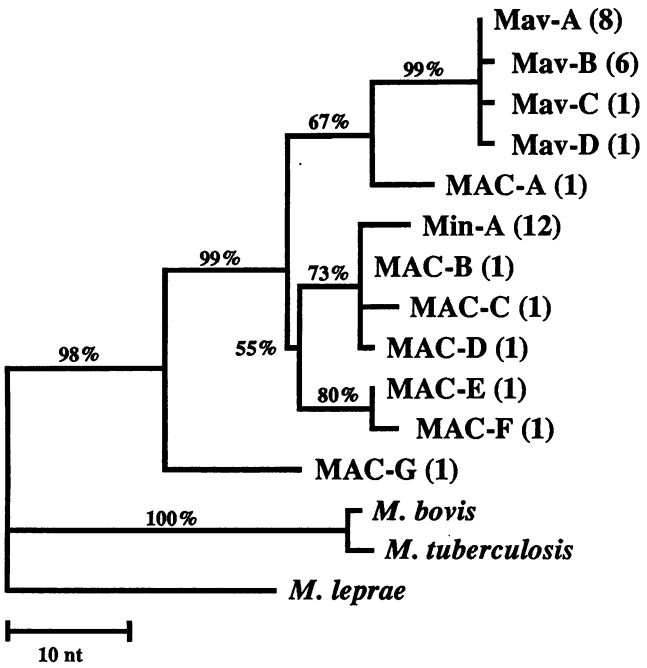

FIG. 3. Rooted phylogenetic tree of 15 16S-23S rDNA ITS sequevars obtained from 35 MAC reference strains and three other mycobacterial species. The number of strains belonging to each sequevar is indicated in parentheses. Horizontal lengths represent genetic distances; the bar represents 10-nucleotide differences. Vertical lengths are not meaningful. The percentage of bootstrap replications which yielded each grouping is indicated.

Comparison with 16S rDNA sequence. If mutation were the only cause of the sequence variability demonstrated, one would expect phylogenetic trees generated from different genetic regions to be similar. On the other hand, if recombination plays a significant role, it would be possible for a single organism to have genetic material derived from more than one ancestral genotype (4). In this case, the position of such an organism on a phylogenetic tree would vary depending on the genomic region analyzed. Such recombination has been demonstrated in phylogenetic studies of *E. coli* and poliovirus (4, 21). The sequence of the 16S rDNA has been widely used to define taxonomic relationships at the kingdom, genus, and species levels. We thus compared the results of 16S-23S rDNA ITS sequencing with those of 16S rDNA sequencing.

Of the 35 MAC reference strains sequenced, the 16S rDNA sequence has been reported (3) for 12 strains (ATCC 35767, 35766, 35765, 35774, 13950, 35762, 35761, 35848, 35763, 35764, 35770, and 35847) for 611 base positions (*E. coli* 16S rDNA positions 111 to 320 and 1110 to 1503). These base positions include the gene regions which are most variable in mycobacteria and account for most of the variability found in the entire 16S rDNA gene (3). The complete sequence of 16S rDNA is also known for *M. bovis* BCG (28), *M. tuberculosis* H37Rv (12), and *M. leprae* (15).

Phylogenetic trees were derived for these 12 MAC reference strains, again with the three other mycobacterial species considered an outgroup. Figure 4 shows the trees obtained by using complete 16S-23S rDNA ITS sequence, by using partial 16S rDNA sequences, and by combining both sequences in a single analysis. Every strain belonging to each 16S-23S rDNA ITS sequevar had an identical partial

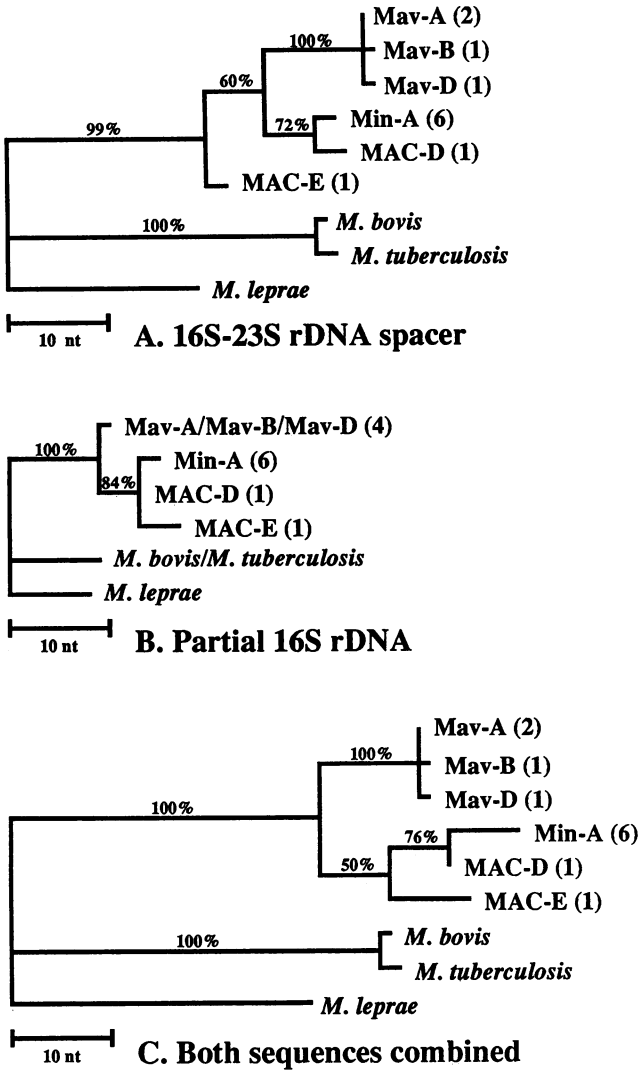

FIG. 4. Rooted phylogenetic trees of nine sequevars derived from 12 MAC reference strains and three other mycobacterial species. The trees are based on complete 16S-23S rDNA ITS sequence (A), partial 16S rDNA sequence (B), and both sequences combined (C). The number of strains belonging to each sequevar is indicated in parentheses. Horizontal lengths represent genetic distances; the bar represents 10-nucleotide differences. Vertical lengths are not meaningful. The percentage of bootstrap replications which yielded each grouping is indicated.

16S rDNA sequence, so the ITS sequevar names were used in all three trees.

The division of MAC strains by the two sequence-based methods was consistent. The four *M. avium* strains yielded three closely related ITS sequevars which grouped together in 100% of bootstrap replications (Mav-A, Mav-B, and Mav-D, Fig. 4A). All four strains had identical partial 16S rDNA sequences (Fig. 4B). The six *M. intracellulare* strains in ITS sequevar Min-A all had identical partial 16S rDNA sequences. *M. bovis* and *M. tuberculosis* differed by three nucleotides in the 16S-23S rDNA ITS but grouped together in 100% of bootstrap replications (Fig. 4A). They had no nucleotide differences in their partial 16S rDNA sequences (Fig. 4B).

The branching orders differed slightly among the three trees (Fig. 4). However, when only groupings found in 95% of bootstrap replications were retained (7), the three trees were identical. The branches on the 16S-23S rDNA ITS tree (Fig. 4A) were longer than those on the partial 16S rDNA tree (Fig. 4B), reflecting the greater numbers of nucleotide differences found in the ITS. For example, the strains in sequevar Mav-A differed from those in sequevar Min-A by 17 nucleotides of ITS sequence and by 7 nucleotides of partial 16S rDNA sequence. Also, *M. bovis* differs from *M. leprae* by 46 nucleotides of ITS sequence, by 17 nucleotides of partial 16S rDNA sequence (as used for Fig. 4B), and by 25 nucleotides of complete 16S rDNA sequence (15, 28).

The consistency of the phylogenetic trees generated by using partial 16S rDNA sequences and by using 16S-23S rDNA ITS sequences (Fig. 4A and 4B) contrasts with the results of Dykhuizen and Green (4). These authors generated phylogenetic trees for eight wild strains of *E. coli* by using three different genes. The three gene trees differed significantly, which the authors explained on the basis of recombination (i.e., horizontal gene transfer). Our results may indicate that members of the *M. avium* complex are less prone to recombination than *E. coli*. Alternatively, our results may reflect the greater stability of the rRNA operon as a phylogenetic marker or our choice of two adjacent gene regions within the same operon for comparison. Sequencing of other genes of the same MAC strains should clarify these questions.

DISCUSSION

The taxonomy and phylogenetic relationships within the MAC have not been resolved despite the application of currently available methods used to classify bacteria. We have sequenced the 16S-23S rDNA ITS from a broad sample of MAC strains, including all 28 currently recognized serovars. This method gave sequence-based definitions for the species *M. avium* (sequevars Mav-A to Mav-D) and *M. intracellulare* (sequevar Min-A) which may be more useful than current definitions. It also revealed a wide range of genetic diversity among MAC reference strains which are neither *M. avium* nor *M. intracellulare*, suggesting the presence of several as yet undefined species. The study was limited to serotypeable reference strains, and the diversity seen might have been even greater had nontypeable MAC strains been included.

Advantages of strain differentiation by DNA sequence data. Sequence-based strain differentiation has been successfully applied to viruses (21, 25). The results of this study suggested advantages to the use of DNA sequence data for bacterial strain differentiation. Indeed, some of the advantages seem self-evident. Assuming that all strains share the chosen gene locus, every strain can be characterized by sequence and placed phylogenetically in relation to the other strains. This contrasts to serologic methods, phage typing, and bacteriocin typing, all of which may yield nontypeable strains. The nontypeable strains may be the most interesting taxonomically, as they may be the most divergent. Sequences generated by different laboratories can be easily compared by using a computerized data base. In contrast, RFLP patterns from different laboratories are more difficult to compare.

By using sequence-based strain differentiation, it is possible to evaluate the statistical certainty of the phylogenetic relationships derived. For example, all 16 *M. avium* strains (sequevars Mav-A to Mav-D) had very similar ITS se-

quences and grouped together on 99% of bootstrap replications, confirming the hypothesis that they have a common ancestor. Serologic testing defines groups of strains by a single characteristic, the ability to react with a particular serum. In many cases, these groupings have been useful, but there is no way to predict with statistical confidence that groupings based on a single characteristic will have broad significance (7). This study and others (11, 17, 31) have shown that MAC strains within the same serovar are genetically diverse. For instance, the two strains in this study which belonged to serovar 7 had very dissimilar sequences. Hybridization assays also define groups by only a single characteristic, the ability to hybridize to a particular probe. While 11 of the strains which hybridized with the *M. intracellulare* probe (Gen-Probe) belonged to the same ITS sequevar (Min-A), one such strain belonged to a different sequevar (MAC-A) with a highly divergent ITS sequence (Table 1 and Fig. 3).

Statistical analysis can also be applied to RFLP and multilocus enzyme electrophoresis data, since each of these strain differentiation methods measures multiple characteristics. However, these methods measure sequence differences indirectly. Also, both are now relatively more cumbersome, because PCR has greatly increased the ease and rapidity of DNA sequencing. When the same region is sequenced repetitively, time is saved by processing many samples together. Also, automated sequencing is now available, resulting in further time savings.

The 16S-23S rDNA ITS is a promising locus for sequence-based strain differentiation. Located centrally in an operon which is widely used for molecular taxonomy, it yields more nucleotide differences than 16S rDNA (Fig. 4) and is easier to sequence because of its smaller size. It should be generally useful for mapping phylogenetic relationships among closely related species. For example, three phylogenetic trees for the genus *Mycobacterium* based on 16S rDNA sequences have recently been published (19, 22, 26). The uncertainty of some of the branching points on these trees has been acknowledged (19, 26). The greater variability of the ITS sequence may allow better definition of these species relationships.

The high sequence variability of the 16S-23S rDNA ITS is also advantageous for the development of specific oligonucleotide probes or primers (1). Figure 2 displays three potential probe-primer sites for *M. avium*. Each of these sites was completely conserved in the four *M. avium* sequevars, which included all 16 *M. avium* reference strains, but had two or more mismatches to all other known 16S-23S rDNA ITS sequences. Similarly, there were two potential probe-primer sites for the Min-A sequevar, which included the *M. intracellulare* type strain and 11 other strains (data not shown). Genus- or species-specific amplification of the ITS could be performed, followed by hybridization with various probes, each specific for a different species or sequevar (24).

Sequence-based strain differentiation can be used to study bacteria without the biases inherent in cultivation. By using PCR with hybridization probes, it would be possible to determine the composition of a complex microbial flora at the strain level or to detect a specific strain directly in a clinical specimen. For example, it should be possible to identify environmental reservoirs of genetically distinct MAC strains which infect AIDS patients (11). Alternately, a PCR-hybridization assay might allow a particular epidemic strain of multidrug-resistant *M. tuberculosis* to be diagnosed rapidly.

Limitations of 16S-23S rDNA ITS sequencing. Clearly, the 16S-23S rDNA ITS will not be the ideal locus for all applications of sequence-based strain differentiation. It is well suited to slow-growing mycobacteria, which have only one copy of the rRNA operon. ITS sequencing will be more complex for species which have more than one rRNA operon because of ITS sequence variation between operons. Furthermore, the variability of the ITS will not be optimal for all applications. Although the ITS demonstrated substantial sequence variability across the *M. avium* complex, not all strains could be distinguished. The 16 *M. avium* strains were divided into four sequevars (Mav-A to Mav-D), but half of the strains belonged to a single sequevar. A region of greater sequence variability would be needed to differentiate all strains. This finding is not surprising. Although it does not code for a final product, the 16S-23S rDNA ITS is transcribed and has a processing function, so there is presumably some selective pressure for its conservation. A spacer region between two operons which is not transcribed would be expected to be even more variable and might allow more detailed sequence-based strain differentiation.

ACKNOWLEDGMENTS

This work was supported by the Durham VA Medical Center's Research Center on AIDS and HIV Infection, the Department of Veterans Affairs, and Public Health Service training grant AI07392 from the National Institutes of Health.

We thank Anna Y. Tsang, J. Michael Harter, and Patrick J. Brennan from Colorado State University for providing the reference MAC strains. We thank Kathy H. McDonald and Peter Zwadyk, Jr., of the Durham VA Medical Center for providing the *M. avium* and *M. intracellulare* type strains.

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. **Bercovier, H., O. Kafri, and S. Sela.** 1986. Mycobacteria possess a surprisingly small number of ribosomal RNA genes in relation to the size of their genomes. Biochem. Biophys. Res. Commun. **136**:1136–1141.
3. **Böddinghaus, B., J. Wolters, W. Heikens, and E. C. Böttger.** 1990. Phylogenetic analysis and identification of different serovars of *Mycobacterium intracellulare* at the molecular level. FEMS Microbiol. Lett. **70**:197–204.
4. **Dykhuizen, D. E., and L. Green.** 1991. Recombination in *Escherichia coli* and the definition of biological species. J. Bacteriol. **173**:7257–7268.
5. **Eisenstein, B. I.** 1990. New molecular techniques for microbial epidemiology and the diagnosis of infectious diseases. J. Infect. Dis. **161**:595–602.
6. **Falkinham, J. O., III, B. C. Parker, and H. Gruft.** 1980. Epidemiology of infection by nontuberculous mycobacteria. I. Geographical distribution in the eastern United States. Am. Rev. Respir. Dis. **121**:931–937.
7. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.
8. **Frothingham, R., R. L. Allen, and K. H. Wilson.** 1991. Rapid 16S ribosomal DNA sequencing from a single colony without DNA extraction or purification. BioTechniques **11**:40–44.
9. **Frothingham, R., R. B. Blitchington, D. H. Lee, R. C. Greene, and K. H. Wilson.** 1992. UV absorption complicates PCR decontamination. BioTechniques **13**:208–210.
10. **Grange, J. M., M. D. Yates, and E. Boughton.** 1990. The avian tubercle bacillus and its relatives. J. Appl. Bacteriol. **68**:411–431.
11. **Hampson, S. J., F. Portaels, J. Thompson, E. P. Green, M. T. Moss, J. Hermon-Taylor, and J. J. McFadden.** 1989. DNA probes demonstrate a single highly conserved strain of *Myco-*

bacterium avium infecting AIDS patients. Lancet **i**:65–68.
12. **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.
13. **Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace.** 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. USA **82**:6955–6959.
14. **Lapage, S. P., P. H. A. Sneath, E. F. Lessel, V. B. D. Skerman, H. P. R. Seeliger, and W. A. Clark (ed.).** 1992. International code of nomenclature of bacteria, p. 133–135. American Society for Microbiology, Washington, D.C.
15. **Liesack, W., C. Pitulle, S. Sela, and E. Stackebrandt.** 1990. Nucleotide sequence of the 16S rRNA from *Mycobacterium leprae*. Nucleic Acids Res. **18**:5558–5559.
16. **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.
17. **McFadden, J. J., P. D. Butcher, J. Thompson, R. Chiodini, and J. Hermon-Taylor.** 1987. The use of DNA probes identifying restriction-fragment-length polymorphisms to examine the *Mycobacterium avium* complex. Mol. Microbiol. **1**:283–291.
18. **National Institutes of Health.** 1980. Mycobacterial culture collection. National Institutes of Health, Bethesda, Md.
19. **Pitulle, C., M. Dorsch, J. Kazda, J. Wolters, and E. Stackebrandt.** 1992. Phylogeny of rapidly growing members of the genus *Mycobacterium*. Int. J. Syst. Bacteriol. **42**:337–343.
20. **Regnery, R. L., C. L. Spruill, and B. D. Plikaytis.** 1991. Genotypic identification of rickettsiae and estimation of intraspecies sequence divergence for portions of two rickettsial genes. J. Bacteriol. **173**:1576–1589.
21. **Rico-Hesse, R., M. A. Pallansch, B. K. Nottay, and O. M. Kew.** 1987. Geographic distribution of wild poliovirus type 1 genotypes. Virology **160**:311–322.
22. **Rogall, T., J. Wolters, T. Flohr, and E. C. Böttger.** 1990. Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*. Int. J. Syst. Bacteriol. **40**:323–330.
23. **Saito, H., H. Tomioka, K. Sato, H. Tasaka, and D. J. Dawson.** 1990. Identification of various serovar strains of *Mycobacterium avium* complex by using DNA probes specific for *Mycobacterium avium* and *Mycobacterium intracellulare*. J. Clin. Microbiol. **28**:1694–1697.
24. **Simonet, P., P. Normand, A. Moiroud, and R. Bardin.** 1990. Identification of *Frankia* strains in nodules by hybridization of polymerase chain reaction products with strain-specific oligonucleotide probes. Arch. Microbiol. **153**:235–240.
25. **Smith, J. S., L. A. Orciari, P. A. Yager, H. D. Seidel, and C. K. Warner.** 1992. Epidemiologic and historical relationships among 87 rabies virus isolates as determined by limited sequence analysis. J. Infect. Dis. **166**:296–307.
26. **Stahl, D., and J. W. Urbance.** 1990. The division between fast- and slow-growing species corresponds to natural relationships among the mycobacteria. J. Bacteriol. **172**:116–124.
27. **Stoflett, E. S., D. D. Koeberl, G. Sarkar, and S. S. Sommer.** 1988. Genomic amplification with transcript sequencing. Science **239**:491–494.
28. **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.
29. **Swofford, D. L.** 1991. PAUP: phylogenetic analysis using parsimony, version 3.0s. Computer program distributed by the Illinois Natural History Survey, Champaign, Ill.
30. **Tsang, A. Y., I. Drupa, M. Goldberg, J. K. McClatchy, and P. J. Brennan.** 1983. Use of serology and thin-layer chromatography for the assembly of an authenticated collection of serovars within the *Mycobacterium avium-Mycobacterium intracellulare-Mycobacterium scrofulaceum* complex. Int. J. Syst. Bacteriol. **33**:285–292.
31. **Wasem, C. F., C. M. McCarthy, and L. W. Murray.** 1991. Multilocus enzyme electrophoresis analysis of the *Mycobacterium avium* complex and other mycobacteria. J. Clin. Microbiol. **29**:264–271.
32. **Wayne, L. G., and H. A. Sramek.** 1992. Agents of newly recognized or infrequently encountered mycobacterial diseases. Clin. Microbiol. Rev. **5**:1–25.
33. **Wilson, K. H., R. B. Blitchington, and R. C. Greene.** 1990. Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction. J. Clin. Microbiol. **28**:1942–1946.
34. **Yakrus, M. A., and R. C. Good.** 1990. Geographic distribution, frequency, and specimen source of *Mycobacterium avium* serotypes isolated from patients with acquired immunodeficiency syndrome. J. Clin. Microbiol. **28**:926–929.