Comparison of *Mycobacterium* 23S rRNA Sequences by High-Temperature Reverse Transcription and PCR

BENJAMIN B. STONE,* RAYMOND M. NIETUPSKI,† GARY L. BRETON, AND WILLIAM G. WEISBURG

GENE-TRAK, Inc., Framingham, Massachusetts 01701

We describe a modified rRNA sequence analysis method which we used to determine the phylogenetic relationships among 58 species belonging to the genus *Mycobacterium*. We combined the sensitivity of the reverse transcriptase PCR for amplifying nanogram amounts of template rRNA material with the elevated extension temperatures used for the thermostable DNA polymerase from *Thermus thermophilus*. A 70°C reverse transcription extension step permitted improved read-through of highly structured rRNA templates from members of the genus *Mycobacterium*, which have G+C contents of 66 to 71 mol%. The nucleic acid sequences of the amplified material were then determined by performing thermal cycle sequencing with α-32P-labeled primers, again with extension at 70°C. Nonspecifically terminated bands were chased by using terminal deoxynucleotidyl transferase. Our method had a template requirement of nanogram amounts or less of purified RNA or 2,000 CFU of intact cells and had sufficient sensitivity so that lyophils obtained from the American Type Culture Collection could be used as source material. Sequences from a 250-nucleotide stretch of the 23S rRNA were aligned, and phylogenetic trees were evaluated by using the De Soete distance treeing algorithm and *Rhodococcus bronchiatis* as the outgroup. Our 23S rRNA trees were compared with previously published 16S rRNA trees, including the comprehensive trees developed by the University of Illinois Ribosomal Database Project, and included 15 species not evaluated previously. Most of the groups were in general agreement and were consistent with relationships determined on the basis of biochemical characteristics, but some new relationships were also observed.

The genus *Mycobacterium* contained 56 validated species (18) as of 1991, and recently Shinnick and Good (34) listed 71 species. In addition, there are several subspecies and many trivially named strains. Members of this genus are frequently categorized as being slow growers (more than 7 days is required to see colonies on a plate) or fast growers (less than 7 days is required) (36). Several important human pathogens, including *Mycobacterium tuberculosis* and *Mycobacterium leprae*, are members of the slowly growing group, as are a number of opportunistic pathogens (Table 1) (15). In recent years there have been a number of reports describing various methods for improving the method that Lane et al. (23) used for rRNA sequence analysis. The improvements that have been described include separating the labeling step from the termination step (11), altering the concentrations of specific deoxynucleoside triphosphates (40), utilizing low numbers of cells (27, 39), sequencing DNA PCR amplicons (6) with an additional 70°C extension step (4), and coupling DNA amplification with thermal cycle sequencing (30, 33). None of the methods that have been described combined all of the characteristics for developing sequence information desired by us. We sought a method in which small amounts of RNA template material could be used in combination with high extension temperatures. In this paper we describe such a method and the phylogenetic relationships of 58 members of the genus *Mycobacterium* based on a 250-nucleotide, variable region of the 23S rRNA subunit. The organisms which we used included 17 species which have not been included yet on the complete 16S rRNA tree developed by the Ribosomal Database Project (RDP) at the University of Illinois; 15 of these species have not had their 16S rRNA sequences evaluated in a phylogenetic context yet.

**MATERIALS AND METHODS**

**Growth of bacterial strains.** Organisms (Table 1) were grown on Middlebrook 7H10 agar. Cell pellets or lyophils were processed by shaking them with zirconium beads and phenol as described previously (3).

**Nucleic acid purification.** RNA was purified from cell lysates (3) through a cesium trilauratoacetate acid gradient as described previously (3). When DNA was used as the template material, it was cloned from PCR-amplified ribosomal...
with EGTA [ethylene glycol-bis (6-aminoethyl ether.j-N,N,W,W-tetraacetic acid], the nucleic acid extractions were ethanol precipitated by using yeast tRNA as the carrier. The pellets were redissolved in 100 μl of 10 mM Tris (pH 8)-1 mM EDTA buffer, and 1 μl was used as template material for Tth RT amplification.

**Nucleic acid amplification.** For reverse transcription and subsequent PCR we followed the recommendations of Perkin-Elmer-Cetus, Norwalk, Conn., and used the Tth DNA polymerase and reagents in a GeneAmp thermocycleable Tth RT PCR kit (Perkin-Elmer-Cetus). Reverse primer 3284 (Table 2) was annealed with forward primer 3463 at 94°C for 20 s and annealing with forward primer 3463 and extension at 70°C for 20 s. The following three species required slightly less stringent annealing conditions (65°C): Mycobacterium gilvum, Mycobacterium smegmatis, and Mycobacterium rhodesiae. Amplifications were purified from buffer and primers by using Ultra Free-MC 30,000 mw filters (Millipore, Bedford, Mass.).

**Nucleic acid sequencing.** The primers used for amplification and sequencing (Table 2) were designed to hybridize primarily to high-G+C-content gram-positive organisms and to be stable for annealing at 70°C. Primer 3465 and reverse primer 3284 generated a 369-bp amplicon from M. tuberculosis. For thermal cycle sequencing of linear, double-stranded amplons we used α-32P-labeled primers, as well as the reagents and protocols described in an AmpliTaq cycle sequencing kit (Perkin-Elmer-Cetus). Most ambiguities were resolved by sequencing the opposite strand. Nondideoxy-terminated pause sites were alleviated with a terminal deoxynucleotidyl transferase chase at 37°C for 30 min as described previously (22). The sequences of DNA clones were determined by using a Sequenase kit (U.S. Biochemicals, Cleveland, Ohio).

**Nucleic acid sequence analysis.** Sequences (Escherichia coli positions 519 to 729) were manually aligned with VAX minicomputer by using the Olsen sequence editor (28) and were printed by using a MegAlign (version 3.02) package obtained from DNASTAR, Madison, Wis. Two hypervariable stem-loop

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain*</th>
<th>Growth rate*</th>
<th>Template*</th>
<th>GenBank accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strict pathogens</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium africanum*</td>
<td>ATCC 25420T</td>
<td>S</td>
<td>R/D</td>
<td>U24502</td>
</tr>
<tr>
<td>Mycobacterium asiaticum</td>
<td>ATCC 25276T</td>
<td>S</td>
<td>R</td>
<td>U24505</td>
</tr>
<tr>
<td>Mycobacterium bovis</td>
<td>ATCC 35745</td>
<td>S</td>
<td>D</td>
<td>U24509</td>
</tr>
<tr>
<td>Mycobacterium haemophilum*</td>
<td>GT3804</td>
<td>S</td>
<td>R</td>
<td>U24526</td>
</tr>
<tr>
<td>Mycobacterium leprae</td>
<td>S</td>
<td>X66567</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium malmoense</td>
<td>ATCC 29571T</td>
<td>S</td>
<td>R</td>
<td>U24531</td>
</tr>
<tr>
<td>Mycobacterium smolenf*</td>
<td>ATCC 27692T</td>
<td>S</td>
<td>R</td>
<td>U24547</td>
</tr>
<tr>
<td>Mycobacterium simiae</td>
<td>ATCC 25275T</td>
<td>S</td>
<td>R</td>
<td>U24548</td>
</tr>
<tr>
<td>Mycobacterium tuberculosis</td>
<td>H37Ra (= ATCC 25177)</td>
<td>S</td>
<td>R/D</td>
<td>U24555</td>
</tr>
<tr>
<td>Mycobacterium ulcerans*</td>
<td>GT4380</td>
<td>S</td>
<td>R</td>
<td>U24556</td>
</tr>
<tr>
<td><strong>Opportunistic pathogens</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium avium</td>
<td>GT3246</td>
<td>S</td>
<td>D</td>
<td>U24508</td>
</tr>
<tr>
<td>Mycobacterium celatum*</td>
<td>ATCC 51131T</td>
<td>S</td>
<td>R/D</td>
<td>U24510</td>
</tr>
<tr>
<td>Mycobacterium chelonae</td>
<td>ATCC 36245</td>
<td>F</td>
<td>R</td>
<td>U24511</td>
</tr>
<tr>
<td>Mycobacterium farcinogenes</td>
<td>ATCC 35753T</td>
<td>S</td>
<td>R</td>
<td>U24517</td>
</tr>
<tr>
<td>Mycobacterium fortuitum</td>
<td>ATCC 6841T</td>
<td>F</td>
<td>R/D</td>
<td>U24519</td>
</tr>
<tr>
<td>Mycobacterium Gardneri</td>
<td>ATCC 56251</td>
<td>F</td>
<td>R</td>
<td>U24523</td>
</tr>
<tr>
<td>Mycobacterium intracellulari</td>
<td>ATCC 13590T</td>
<td>S</td>
<td>D</td>
<td>U24527</td>
</tr>
<tr>
<td>Mycobacterium kansasii</td>
<td>ATCC 12478T</td>
<td>S</td>
<td>D</td>
<td>U24526</td>
</tr>
<tr>
<td>Mycobacterium marinum</td>
<td>ATCC 927T</td>
<td>S</td>
<td>R</td>
<td>U24532</td>
</tr>
<tr>
<td>Mycobacterium microti</td>
<td>ATCC 11152</td>
<td>S</td>
<td>R</td>
<td>U24533</td>
</tr>
<tr>
<td>Mycobacterium paratuberculosis*</td>
<td>ATCC 19689</td>
<td>S</td>
<td>R</td>
<td>U24539</td>
</tr>
<tr>
<td>Mycobacterium porcinum*</td>
<td>ATCC 33776T</td>
<td>F</td>
<td>R</td>
<td>U24541</td>
</tr>
<tr>
<td>Mycobacterium scrofulaceum</td>
<td>ATCC 19981T</td>
<td>S</td>
<td>D</td>
<td>U24545</td>
</tr>
<tr>
<td>Mycobacterium smegmatis</td>
<td>ATCC 17406</td>
<td>F</td>
<td>R</td>
<td>U24549</td>
</tr>
<tr>
<td>Mycobacterium sputum</td>
<td>ATCC 35799T</td>
<td>S</td>
<td>R</td>
<td>U24551</td>
</tr>
<tr>
<td>Mycobacterium xenopi</td>
<td>ATCC 19250T</td>
<td>S</td>
<td>R</td>
<td>U24558</td>
</tr>
<tr>
<td><strong>Rare, potential pathogens</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium flavescens</td>
<td>ATCC 14474T</td>
<td>F</td>
<td>R</td>
<td>U24518</td>
</tr>
<tr>
<td>Mycobacterium gordoniae</td>
<td>ATCC 35756</td>
<td>S</td>
<td>R/D</td>
<td>U24525</td>
</tr>
<tr>
<td>&quot;Mycobacterium lactis*&quot;</td>
<td>ATCC 27356</td>
<td>S</td>
<td>R</td>
<td>U24530</td>
</tr>
<tr>
<td>Mycobacterium lepraemurium*</td>
<td>ATCC 35779</td>
<td>S</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>Mycobacterium neauromai</td>
<td>ATCC 27595T</td>
<td>F</td>
<td>R</td>
<td>U24553</td>
</tr>
<tr>
<td>Mycobacterium neoaurum</td>
<td>ATCC 17468</td>
<td>F</td>
<td>R</td>
<td>U24554</td>
</tr>
<tr>
<td>Mycobacterium thermoresistibile</td>
<td>ATCC 19527T</td>
<td>F</td>
<td>R</td>
<td>U24553</td>
</tr>
<tr>
<td><strong>Nonpathogenic environmental mycobacteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium agri*</td>
<td>ATCC 27406T</td>
<td>F</td>
<td>R</td>
<td>U24503</td>
</tr>
<tr>
<td>Mycobacterium achiense</td>
<td>ATCC 27280T</td>
<td>F</td>
<td>R</td>
<td>U24504</td>
</tr>
<tr>
<td>Mycobacterium aurum</td>
<td>ATCC 23266T</td>
<td>F</td>
<td>R</td>
<td>U24506</td>
</tr>
<tr>
<td>Mycobacterium astroaffricum*</td>
<td>ATCC 33464T</td>
<td>F</td>
<td>R</td>
<td>U24507</td>
</tr>
<tr>
<td>Mycobacterium chitae</td>
<td>ATCC 19627T</td>
<td>F</td>
<td>D</td>
<td>U24512</td>
</tr>
<tr>
<td>Mycobacterium chubuense</td>
<td>ATCC 27278T</td>
<td>F</td>
<td>R</td>
<td>U24513</td>
</tr>
<tr>
<td>Mycobacterium dienhoferi</td>
<td>ATCC 19543</td>
<td>F</td>
<td>R</td>
<td>U24514</td>
</tr>
<tr>
<td>Mycobacterium doriae*</td>
<td>ATCC 43910T</td>
<td>F</td>
<td>R</td>
<td>U24515</td>
</tr>
<tr>
<td>Mycobacterium fallax</td>
<td>ATCC 35219T</td>
<td>F</td>
<td>R</td>
<td>U24516</td>
</tr>
<tr>
<td>Mycobacterium gadi</td>
<td>ATCC 27726T</td>
<td>F</td>
<td>R</td>
<td>U24520</td>
</tr>
<tr>
<td>&quot;Mycobacterium gallinarum*&quot;</td>
<td>ATCC 17910</td>
<td>R</td>
<td></td>
<td>U24521</td>
</tr>
<tr>
<td>Mycobacterium gastri</td>
<td>ATCC 15754T</td>
<td>S</td>
<td>R/D</td>
<td>U24522</td>
</tr>
<tr>
<td>Mycobacterium gilvum</td>
<td>ATCC 43909T</td>
<td>F</td>
<td>R</td>
<td>U24524</td>
</tr>
</tbody>
</table>

Continued on following page
were performed with these same samples to demonstrate that various sources, including \textit{M. leprae} tinely detected at least 1 ng in each experiment when we used the template material. We detected no products by ethidium DNA PCR in which we used AmpliTaq DNA polymerase represented between 10^6 and 10^7 initial rRNA template molecules. We were concerned about the relatively poor sensitivity in our experiments compared with the data of Meyers and Gel- fand (26), who observed levels of sensitivity between 100 and 1,000 copies with transcribed RNA. Synthetic transcripts of a clone of the 5' one-half of the 23S rRNA gene from \textit{M. tuberculosis} H37Ra (avirulent) were prepared by using T7 RNA polymerase (Promega, Madison, Wis.), and the amounts of these transcripts were determined by measuring the \textit{A}_{260}. Next, serial dilutions were prepared, and 1-\mu l aliquots were used as template material. We found that the level of sensitivity obtained with the synthetic RNA was at least 3 orders of magnitude higher than the level of sensitivity obtained with the natural material (data not shown). Since the transcript included only about 1,600 nucleotides, 1 fg represented 10^4 molecules. These results more closely approached the results of Meyers and Gelfand (26). There are several possible reasons for the difference in sensitivity between the natural and synthetic rRNAs. Posttranscrip- tional modifications of the natural rRNA, such as methylation, could have inhibited efficient copying by the polymerase. It is also possible that the secondary and/or tertiary structure of the full-length natural rRNA was more stable, which could have further impeded copying. The synthetic transcript included only the first one-half of the 23S rRNA molecule.

\begin{table}
\centering
\caption{The GenBank accession numbers of the 23S rRNA sequences shown in Fig. 1 are given in Table 1.}
\begin{tabular}{lllll}
\hline
Species & Strain & Growth rate & Template & GenBank accession no. \\
\hline
\textit{Mycobacterium genosense} & ATCC 3520^T & F & R & U24529 \\
\textit{Mycobacterium maritima}^e & ATCC 43059^T & F & R & U24534 \\
\textit{Mycobacterium nonchromogenicum} & ATCC 19530 & S & R & U24536 \\
\textit{Mycobacterium olivarius} & ATCC 27023T & F & R & U24537 \\
\textit{Mycobacterium parafortuitum}^e & ATCC 19686T & F & R & U24538 \\
\textit{Mycobacterium phlei} & ATCC 11758^T & F & R & U24540 \\
\textit{Mycobacterium porcicola}^e & ATCC 35087^T & F & R & U24542 \\
\textit{Mycobacterium pulvers}^e & ATCC 35154^T & F & R & U24543 \\
\textit{Mycobacterium rhodense} & ATCC 27024 & F & R & U24544 \\
\textit{Mycobacterium rhodesiae} & ATCC 33027 & F & R & U24550 \\
\textit{Mycobacterium terrae} & ATCC 23292 & S & R & U24557 \\
\textit{Mycobacterium tokaiense}^e & ATCC 27282* & F & R & U24554 \\
\textit{Mycobacterium triviale} & ATCC 15483 & S & R & U24559 \\
\textit{Mycobacterium vaccae} & ATCC 13032 & R & R & U24559 \\
\textit{Nonmycobacteria} & & & & \\
\textit{Corynebacterium glutamicum} & ATCC 381 & R & R & U24559 \\
\textit{Micrococcus luteus} & ATCC 381 & R & R & U24559 \\
\textit{Rhodococcus bronchialis} & ATCC 25592 & S & R & U24559 \\
\hline
\end{tabular}
\end{table}

\begin{table}
\centering
\caption{ Amplification and sequencing primers targeted to high-G+C-content gram-positive eubacteria}
\begin{tabular}{llll}
\hline
Primer & No. of nucleotides & G+C content (mol\%) & \textit{E. coli} positions & Sequence (5' to 3') \\
\hline
3415 & 32 & 53 & 457-488 & AGTACCGTGA GGGAAGGTG AAAAGTACCC CG \\
3359 & 27 & 48 & 553-579 & GTGATGGCGT GCCTTTTGWA GAATGAG \\
3429 & 27 & 56 & 751-725 & GTGATGGCGT GCCTTTTGWA GAATGAG \\
3284 & 22 & 59 & 791-770 & GTGATGGCGT GCCTTTTGWA GAATGAG \\
\hline
\end{tabular}
\end{table}
We wanted to demonstrate the utility of our technique for amplifying rRNA fragments from very low yields of whole cells. We used *Rhodococcus equi* as a model instead of mycobacteria to eliminate the need for BL-3 containment during processing. Our results revealed a level of sensitivity of about 2,000 CFU (data not shown). Assuming that there are approximately 1,000 23s rRNA targets per CFU, this finding is consistent with the presence of $10^9$ to $10^{10}$ molecules of purified natural rRNA. These values do not account for possible losses during the extraction procedure and therefore are probably underestimates of the true level of sensitivity. Nevertheless, these values are useful for demonstrating the limits for potential screening programs. Single colonies should provide adequate material for amplification. We were able to amplify material from lyophils of *Mycobacterium paratuberculosis*.

**Sequence analysis.** The mycobacterial 23S rRNA regions homologous to *E. coli* positions 519 to 729 are shown in Fig. 1. These regions were used to develop the phylogenetic trees shown in Fig. 2. In the detailed overview analysis (Fig. 2A) we included 17 species that have not been included yet on the 16S rRNA tree available from the RDP at the University of Illinois. Two of these species, *Mycobacterium celatum* (34) and *Mycobacterium africanum* (13), have been placed phylogenetically by other workers. Figure 2B shows the positions of a selection of species extracted from the whole group and reanalyzed to clarify the main groups, as well as improve the general reliability of the results. For the most part the groups of species on these trees are similar to the groups on the rRNA 16S rRNA tree. Some specific differences are discussed below.

*Mycobacterium chitae* ATCC1962T (T = type strain) clustered in the genus *Mycobacterium* near *Mycobacterium komosense*, as found on the 16S rRNA tree obtained from the RDP and as described in a reexamination of this strain by Pitulle et al. (29). This finding is not consistent with the data of Stahl and Urbance (36), who found that *M. chitae* is less closely related to the genus *Mycobacterium* than *Rhodococcus equi* is. We found that *Mycobacterium fallax* and *Mycobacterium poriferae* were separated from the main group of the genus *Mycobacterium*, whereas on the 16S rRNA trees *M. fallax* clusters with *M. chitae* and *Mycobacterium gadium*.

Figure 2C shows the results of an analysis of the fast-growing species, in which *M. tuberculosis* was used as the outgroup. The large cluster from *M. gadium* was quite similar to a large group on the RDP 16S rRNA tree. *Mycobacterium farcinogenes* was excluded from this analysis because of its slow growth rate, but both on the tree shown in Fig. 2A and on the 16S rRNA trees this organism was placed in this cluster. We found that *M. gadium* grouped with *Mycobacterium dierhoferi* and *Mycobacterium neoaurum* and is closely related to *M. chitae*. As noted above, on the RDP 16S rRNA tree *M. fallax* clusters close to these species, whereas on our tree it clustered separately, along with *M. poriferae*.

Figure 2D shows the results of an analysis of the slowly growing species, in which *M. smegmatis* was used as the out-

![Table 1](image-url)
FIG. 2. Phylogenetic relationships based on partial 23S rRNA sequences. (A). Overview of 58 species of the genus Mycobacterium. Note that “Mycobacterium lactis” and “Mycobacterium gallinarum” are not officially recognized species. R. bronchialis was used as the outgroup. (B). Selected overview of the genus Mycobacterium. Reducing the number of species in the tree should have improved the general reliability of the relationships. (C). Selected subgroup of fast growers. M. tuberculosis was used as the outgroup. (D). Selected subgroup of slow growers. M. smegmatis was used as the outgroup.
group. Mycobacterium marinum clustered with Mycobacterium gordonae and close to Mycobacterium asiaticum and Mycobacterium szulgai. On the 16S rRNA trees M. marinum is closely related to M. tuberculosis. The recently described organism Mycobacterium genavense is very closely related to Mycobacterium simiae, as found on the RDP 16S rRNA tree and as determined by a mycologic acid profile analysis (9). M. celatum is identical to Mycobacterium kansasi, a finding which is inconsistent with the phenotypic placement of these organisms as a slowly growing nonphototroph and a slowly growing phototroph, respectively (34). Mycobacterium haemophilum, whose phylogenetic position has not been described, exhibited a high level of 23S rRNA sequence similarity to M. leprae despite many dissimilar phenetic traits (34).

**DISCUSSION**

**Method.** The method described above was a synthesis of the following two established techniques for molecular analysis of RNA sequences: (i) reverse transcription coupled to PCR amplification in which the recombinant DNA polymerase from Thermus thermophilus is used, and (ii) thermal cycle sequencing of the resulting amplicons to avoid many of the problems associated with sequencing linear, double-stranded templates. In this method 33P-end-labeled sequencing primers are used, and amplicons are rapidly separated from amplification primers by using microcentrifuge spin filters. Thus, there is no requirement for gel purification (2), exonuclease digestion (19, 32), biotin capture (25), cold temperature “snap” annealing (8), or separate asymmetric amplification of each strand (17). Our method has several distinct advantages which we used for phylogenetic analysis of the vast majority of species in the genus Mycobacterium (currently 69 type strains are listed by the American Type Culture Collection).

Very small amounts of template material are required for the method described in this paper. Routinely, less than 1 ng of total RNA (as little as 10 pg was used in this study) or fewer than 104 cells are needed to generate amplicons, compared with the 4 to 10 µg of RNA needed for avian myeloblastosis virus-based RT sequencing (23). This is an improvement in sensitivity of at least 3 to 4 (up to 6) orders of magnitude. Compared with the sensitivity of PCR amplification of rDNA genes (5, 6), this is not a step forward. However, we initially targeted RNA (instead of the coding rDNA) because we assumed that the natural copy number (at least 1,000 copies per CFU for M. tuberculosis [1]) would improve the sensitivity of the amplification technique. This assumption was investigated experimentally as described above. Furthermore, we were slightly concerned about sequence variation in chromosomal rDNA genes in those species (presumably including the fast growers) that contain more than one copy. Using the rRNA molecules directly provided sequences from the functional final gene products.

A high temperature during the annealing and extension step permitted improved read-through of stable secondary structures, whereas avian myeloblastosis virus RT is used at 37°C. This is particularly important for high-G+C-content organisms. Our experience with avian myeloblastosis virus RT showed that our technique resulted in significant improvement in the ability to read sequence information. Despite this, some secondary structures were still a problem during sequencing extension with members of the genus Mycobacterium, even at 70°C. We determined that these structures were not gel artifacts since they could be chased away with terminal deoxynucleotidyl transferase.

Sequencing two strands results in improved resolution in denaturing gels because of different specific secondary structures that form with different primers. Avian myeloblastosis virus RT sequencing offers fewer options for checking sequence information. However, it is important to note that the sequence information originates from a single strand, compared with material cloned from PCR amplification preparations.

While the latter approach is an option, it is more time consuming. In addition, because at least 1,000 template molecules are required, the method described in this paper shares the advantage of direct PCR amplification and subsequent thermal cycle sequencing of rDNA genes (6) in that single misincorporations should not detectably affect the sequence information. This can be a problem with material cloned from PCR amplification preparations.

**Phylogenetic relationships.** The data presented in this paper place for the first time 15 members of the genus Mycobacterium in a phylogenetic context based on rRNA sequence information. Several of these organisms, including M. haemophilum, Mycobacterium shimoidei, and Mycobacterium ulcerans, are considered pathogenic for humans. We compared our results with the comprehensive overview provided by the 16S rRNA phylogenetic tree of the RDP, as well as with the results of specific analyses described previously (13, 29, 31, 36). For the most part, our results corroborate the 16S rRNA tree data, although there are some exceptions.

It is important to note that a true phylogeny probably will not be derived from a single character, such as a fragment of the 23S rRNA sequence. A polyphasic approach (35) will help even out the deficiencies of using specific traits (e.g., molecular size, stability, “evolutionary clock”). Therefore, our 23S rRNA data are not intended to be used to rewrite the phylogeny of the genus Mycobacterium, but rather complement the data obtained previously. In the case of the 15 species which we studied whose 16S rRNA sequences have not been aligned yet, phenetic traits still should be given significant weight.

Stahl and Urbanicz (36) and Rogall et al. (31) found that the growth rates of species (slowly growing versus fast growing) appeared to be phylogenetically conserved on the basis of 16S rRNA sequence data. Figures 2A and B show that all of the organisms except two can be clustered on the basis of growth rates. Initially, M. chelonae (a fast grower) was grouped with the slowly growing species (Fig. 2A). However, when we examined a subset of the members of the genus (Fig. 2B), M. chelonae was placed in the fast-growing species cluster. This change in position probably resulted from the limited sequence examined (29).

M. farcinogenes, a slowly growing species, clustered with the fast growers, close to Mycobacterium fortuitum (Fig. 2A) and in a subcluster adjacent to Mycobacterium senegalense. This position is very similar to the position determined on the basis of 16S rRNA sequences (29). Although the taxonomic position of this species has not been determined by using other phenetic traits, the results of DNA hybridization studies place it near M. senegalense (34). Assuming that the extracted and realigned positions of M. chelonae and its fellow fast-growing species in Fig. 2B are correct, this is the only example of a contradiction with the slow-fast phylogenetic split (31, 36). The results of several independent approaches (16S rRNA, 23S rRNA, DNA hybridization) have placed M. farcinogenes with the fast growers, and slow growth appears to be a recently acquired trait (29). It is possible that the loss of rapid growth potential in M. farcinogenes has been acquired independently.

The position of M. chitae ATCC19627T which we determined in this study is similar to the position of this organism on the RDP 16S rRNA tree and the position determined by Pitulle et al. (29). Pitulle et al. reexamined the 16S rRNA
sequence of this strain and obtained dramatically different results than Stahl and Urbance obtained (36). Both groups of authors recommended a comprehensive reexamination of the authenticity of this strain. Since the phenetic traits and mycolic acid composition of *M. chitae* still place it in the genus *Mycobacterium* (18), there may have been a problem with the specific culture used by Stahl and Urbance (36).

The identical 23S rRNA sequences of *M. celatum* and *M. kansasii*, like the identical 16S rRNA sequences of *M. gastri* and *M. kansasii*, provide another example of the limitations of relying on a single trait to place species in groups. While both *M. celatum* and *M. kansasii* are opportunistic pathogens in immunocompromised hosts, they belong to different Runyon groups (groups III [nonphotochromogens] and I [photochromogens], respectively) and have different biochemical test profiles (34). Along the same lines, in this study *M. haemophilum* and *M. leprae* clustered closely, although they clearly are quite different bacteria (34). *M. leprae*, an obligately intracellular pathogen, clusters close to (RDP) or between (34) the *M. tuberculosis* complex and the *Mycobacterium avium* complex. Although *M. haemophilum*, a nonphotochromogen, is similar to the *M. avium* complex in several biochemical characteristics, it has a distinctive mycolic acid profile; furthermore, it requires complexed iron for growth and can be cultured in vitro. Thus, the 23S rRNA sequence similarities belie important phenotypic differences, which should again promote caution in the development of a true phylogeny.

Another difference is the branching of *M. fallax* and *M. poriferae* away from the main line of the genus. While the 23S rRNA sequences exhibited the deepest split within the genus in this study, the 16S rRNA trees (29, 34) also exhibit a reasonably deep branch point for *M. fallax* (until now *M. poriferae* has not been placed taxonomically). The true evolutionary significance of these differences may be muted by some of the limitations inherent in using this data set.

First, the sequences are not totally nonbiased for selection. Note the differences in the *M. tuberculosis* complex sequences, the *M. avium* complex sequences, and the sequences of other species. This 23S rRNA region was chosen because of probe development considerations, as well as method development and phylogenetic evaluation considerations. Therefore, this region was not randomly chosen on the basis of conserved structures but rather was chosen on the basis of hypervariable regions. Second, as noted above, only about 250 nucleotides were evaluated, and fewer positions in an alignment are never as informative as more positions. Third, the phylogenetic trees which we obtained are not completely robust. Future changes in groups and membership are likely when more sequence information becomes available.

Finally, short branch lengths can be questionable, as discussed by Pitulke et al. (29). Short internodal distances are less predictive than long internodal distances. This could explain the variable position of *M. chelonae* depending upon the subset of species with which it is analyzed. Of course, efforts to neatly categorize the species highlight the anthropocentric nature of the fast-versus-slow division, and it may be more appropriate to consider some species phylogenetically intermediate between the two groups.

**ACKNOWLEDGMENTS**

We thank Debra Buxton, Annette Hendricks, Joseph Murphy, and William O'Brien for culturing strains and Seth Cohen and David Lane for critically reviewing the manuscript.

**REFERENCES**

24. Maisak, B. Personal communication.
VOL. 45, 1995

MYCOBACTERIUM 23S rRNA SEQUENCES 819