Differentiation of *Mycobacterium* species by direct sequencing of amplified DNA

TILL ROGALL, THOMAS FLOHR and ERIK C. BÖTTGER

Institut für Medizinische Mikrobiologie, Medizinische Hochschule Hannover, Konstanty-Gutschow-Str. 8, 3000 Hannover 61, FRG

(Received 5 February 1990; revised 11 May 1990; accepted 17 May 1990)

Nucleotide sequences specific for a range of *Mycobacterium* species were defined by computer-assisted sequence comparisons of small subunit ribosomal RNA. A polymerase chain reaction-based sequencing strategy was used to demonstrate that the 16S rRNA sequence can be used for the rapid identification of mycobacterial isolates. Identification at the species level can be obtained within 2 d, requiring <10 000 bacteria. This procedure reliably differentiates *Mycobacterium* spp. which are difficult to identify by classical methods, such as *M. malmoense*, *M. szulcgit* and *M. flavescens*.

Introduction

Mycobacteria are aerobic, acid-fast, slow-growing bacteria, which are widespread in nature and include some important human pathogens, e.g. *Mycobacterium tuberculosis* and *M. avium* (Kubica, 1984). A major descriptive division of mycobacteria is related to growth rate and pigmentation, and four groups are recognized. Group I consists of the photochromogenic species of slow growers, members of group II are scotochromogenic slow growers, group III contains the nonphotochromogenic slow growers and group IV consists of rapid growers, defined as maturing in less than 7 d (Wayne & Kubica, 1986).

Current methods for identification of mycobacteria rely primarily on phenotypic, chemotaxonomic and serotaxonomic tests performed on cultures. Identification of mycobacterial isolates based on phenotypic characters is a time-consuming procedure, requiring an additional 2-4 week period after the initial isolation. Many different features requiring complex and specialized tests are necessary to identify mycobacteria at the species level (Wayne & Kubica, 1986; Goodfellow & Cross, 1983). Numerous investigations have been undertaken to define reliable features and test assays for differentiating between species (Wayne et al., 1974, 1976, 1983, 1989). Attempts to subdivide mycobacterial species on the basis of immunological methods, DNA homology and related analyses (Baess, 1979, 1983; McFadden et al., 1987; Gross & Wayne, 1970; Imaeda et al., 1988; Wayne & Kubica, 1986) have proved useful in determining taxonomic relationships, but the reliable identification of mycobacteria remains problematical.

An alternative approach to the identification of mycobacterial isolates is the definition of species-specific nucleotide sequences, combined with a convenient and rapid procedure for obtaining these sequences. Small subunit ribosomal RNAs (rRNA) are an attractive target for this purpose. rRNA is an essential constituent of bacterial and eukaryotic ribosomes (Fox et al., 1980; Woese et al., 1983), and the 16S rRNA molecule is functionally constrained such that it is highly conserved, with rare sequence changes in certain positions. However, the location of these changes is specific to the group or species in which they occur (Gray et al., 1984; Stackebrandt & Woese, 1981; Woese, 1987; Dams et al., 1988). This specificity of the 16S rRNA sequence, and the information content of the molecule, is sufficient to allow both statistically valid phylogenetic analysis and species identification, once the sequence has been determined (Woese, 1987).

We have recently described a method for the rapid and reliable determination of bacterial rRNA sequences by direct sequencing of DNA (Böttger, 1989; Edwards et al., 1989) amplified by the polymerase chain reaction (PCR) (Saiki et al., 1985). In this paper, we define stretches of nucleotides in the structure of small subunit rRNA, based upon comparative 16S rRNA sequencing within a wide range of *Mycobacterium* species, which are highly specific for the various *Mycobacterium* species investigated. This information offers a new approach for the rapid and definitive identification of mycobacterial isolates at the species level.
Methods

Bacterial strains. The strains whose 16S rRNA sequences were determined in this study are listed in Table 1. Clinical isolates of the M. tuberculosis complex including M. bovis and M. africanum were kindly provided by K. Schröder, Forschungsinstitut Borstel.

Determination of sequences. Cultures grown on Löwenstein-Jensen agar slants (M. paratuberculosis required mycobactin) were harvested and washed in 10 mM-Tris, 1 mM-EDTA, 0.1 TE-saturated aqueous phase was transferred to another tube and the DNA precipitated by adding 0.1 vol. 3 M-sodium acetate pH 5.2 were added. After mixing by inversion, samples were placed on ice for 10 min, before centrifugation in a microcentrifuge for 10 min. The supernatant fluid was transferred to another tube, and the DNA precipitated by adding 0.1 vol. 3 M-sodium acetate and 2.5 vols of ethanol (Maniatis et al., 1982). The precipitated nucleic acids were washed, dried, redissolved in TE and stored at −20 °C.

The 16S rRNA sequences from the species given in Table 1 were obtained as described previously (Edwards et al., 1989) and the relevant gene fragment coding for 16S rRNA was synthesized using either primer combination 1 (PA, AGA GTT TGA TCC TGG CTC AG; and pE', CCG TCA ATT CCT TGG AGT TT) or primer combination 2 (PA, AGA GTT TGA TCC TGG CTC AG; and pI', TGC ACA CAG GCC ACA AGG GA). The thermal profile involved 36 cycles of denaturation at 93 °C for 1 min, primer annealing for 2 min, and extension at 72 °C for 6 min. The annealing temperature chosen was dependent on the primers used: 37 °C for primer combination 1 and 60 °C for primer combination 2. Following amplification, the DNA was extracted with phenol and precipitated with ethanol. Alternatively, the volume of the amplification reaction (100 μl) was simply reduced under vacuum. The relevant fragment was isolated by 0.8% (w/v) agarose gel electrophoresis and electroeluted (Maniatis et al., 1982), then purified on Elutip-D columns (Schleicher & Schuell) and ethanol-precipitated. For sequencing, template DNA (0.05–0.2 pmol; the PCR routinely yielded 2–4 pmol of the desired fragment) was annealed to 2.0 pmol of the respective sequencing primer in 10 μl (40 mM-Tris/HCl pH 7.5, 20 mM-MgCl2, 50 mM-NaCl; 5 min at 94 °C, 2 min at 65 °C, 15 min at 37 °C). Sequencing primers used were pB, TAA CAC ATG CAA GTC GAA CG; pC', CCC ACT GCC TCC CGT AG; pC, CTA CGG GAG GCA GCA GTG GG; and pD', GTA TTA CCG CGG CTG CTG. The labelling step was carried out for 5 min at 25 °C by adding 1.0 μl 0.1-M-dithiothreitol, 5-0 μCi (185 kBq) [α-32P]dCTP, 2.0 μl labelling nucleotide mix (dGTP, dATP, TTP at 200 nm) and 3 units sequenase (United States Biochemical Corporation) (total volume 150 μl). From this reaction mix, 3.5 μl was

<table>
<thead>
<tr>
<th>Species</th>
<th>Source*</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. tuberculosis</td>
<td>Inst. Med. Microbiology, Hannover</td>
</tr>
<tr>
<td>M. bovis</td>
<td>DSM 43216</td>
</tr>
<tr>
<td>M. bovis BCG</td>
<td>ATCC 14472</td>
</tr>
<tr>
<td>M. tuberculosis H37</td>
<td>ATCC 14474T</td>
</tr>
<tr>
<td>M. avium</td>
<td>ATCC 14470</td>
</tr>
<tr>
<td>M. chelonei</td>
<td>ATCC 15754T</td>
</tr>
<tr>
<td>M. avium complex</td>
<td>ATCC 15755</td>
</tr>
<tr>
<td>M. fortuitum</td>
<td>ATCC 15985</td>
</tr>
<tr>
<td>M. gordonae</td>
<td>ATCC 33016S</td>
</tr>
<tr>
<td>M. intracellulare</td>
<td>ATCC 33013; Dr Kazda</td>
</tr>
<tr>
<td>M. kansasi</td>
<td>ATCC 33014</td>
</tr>
<tr>
<td>M. kansasii</td>
<td>Dr Kazda, Forschungsinst. Borstel</td>
</tr>
<tr>
<td>M. marinum</td>
<td>Dr Schröder, Forschungsinst. Borstel</td>
</tr>
<tr>
<td>M. malmoense</td>
<td>ATCC 29571T</td>
</tr>
<tr>
<td>M. nonchromogenicum</td>
<td>ATCC 19530T</td>
</tr>
<tr>
<td>M. paratuberculosis</td>
<td>ATCC 19698; Dr Jorgensen, National Veterinary Lab., Copenhagen</td>
</tr>
<tr>
<td>M. phlei</td>
<td>ATCC 35784</td>
</tr>
<tr>
<td>M. scrofulaceum</td>
<td>ATCC 19981T</td>
</tr>
<tr>
<td>M. simiae</td>
<td>ATCC 25275T</td>
</tr>
<tr>
<td>M. spagni</td>
<td>ATCC 33027T; Dr Kazda</td>
</tr>
<tr>
<td>M. smegmatis</td>
<td>ATCC 14468</td>
</tr>
<tr>
<td>M. szulgai</td>
<td>ATCC 25799T</td>
</tr>
<tr>
<td>M. terrae</td>
<td>ATCC 15755T</td>
</tr>
<tr>
<td>M. xenopi</td>
<td>ATCC 19520T</td>
</tr>
</tbody>
</table>

* ATCC, American Type Culture Collection, Rockville, USA; DSM, Deutsche Stamm-sammlung für Mikroorganismen, Braunschweig, FRG; T, type strain.
Identification of mycobacteria

added to 2.5 µl of the appropriate termination mix (200µM-dNTPs, 5 µM-ddNTP, 50 µM-NaCl) and incubated for 5 min at 37 °C. Formamide/dye stop mix (4 µl) was added, and samples were heated for 5 min at 94 °C before loading 2.5 µl onto a 6% sequencing gel (Maniatis et al., 1982). After electrophoresis, gels were fixed in 10% (v/v) acetic acid/12% (v/v) methanol, dried, and exposed to X-ray film for 12 h.

Oligonucleotides were synthesized on a Gene Assembler Plus (Pharmacia) and purified by shadow-casting polyacrylamide gel electrophoresis (Maniatis et al., 1982).

The sequences were aligned using a multisequence alignment algorithm (Krüger & Oosterburg, 1983) provided in the program package SAGE (Technoma) designed for the IBM XT/AT/PS2 and compatible systems. The alignment was adjusted manually to account for common secondary structure.

Results and Discussion

The PCR primers directed the synthesis of an approximately 1 kb gene fragment containing the 5' part of the gene coding for 16S rRNA. Fig. 1 shows the physical location of the oligonucleotides used in this study. As can be seen in Fig. 2, as little as 1 ng of purified DNA, is sufficient to yield a fragment by the PCR which can be sequenced.

An alignment of the sequences from the organisms investigated is given in Fig. 3. The depicted part of the small subunit rRNA gene (positions 123-273) contains a region known to be highly variable in eubacteria (Woese et al., 1983), i.e. positions 200-240 according to the numbering system for Escherichia coli (Brosius et al., 1978). By comparing the aligned sequences obtained, regions specific for each species can be defined.

In detail, the M. tuberculosis complex can be identified at positions 123-157, 175-213 and 215-273. M. tuberculosis is highly related to M. bovis and M. africanum (Baess, 1979), such that a proposal has been put forward to reduce M. bovis and M. africanum to synonyms of M. tuberculosis, with a possible subdivision at the subspecific or infrasubspecific level (Wayne & Kubica, 1986). Accordingly, M. tuberculosis cannot be differentiated from M. kansasii at the small subunit ribosomal RNA level, because the 16S rRNA sequence from both species is identical, indicating a previously unreported close phylogenetic relationship between these two species (unpublished observation).

As expected from the location of this thoroughly investigated region in an rRNA segment characterized by its interspecies variability (Woese et al., 1983; Woese, 1987), an intensive search of published 16S rRNA sequences (Dams et al., 1988) gave no hint of convergent sequence homology in unrelated organisms within this part of the gene. Another less conserved region, which may be useful for confirmation of mycobacterial species already characterized by sequencing region 123–273, is shown in Fig. 4. The sequence determination in this part of the 16S rRNA is especially useful for differentiating M. gastri/M. kansasi from M. simiae and M. scrofulaceum. Sequencing with the appropriate primer, pC or pD', can be performed using the same amplified DNA fragment as that for determining the sequence of region 123–273.
Identification of mycobacteria 1919

Fig. 4. Alignment of 16S rRNA sequences. The first nucleotide shown corresponds to E. coli 16S rRNA position 429 (Brosius et al., 1978). M. tuberculosis was used as the reference sequence; nucleotides different from those of M. tuberculosis are indicated; bars indicate deletions. The nucleotide sequence was determined using oligonucleotides pC or pD'. Both primers result in identical sequence information, but obtain it from opposite strands.

It is well established that the variability of 16S rRNA sequences is conserved at the species level (Woese, 1987). In order to prove that different strains belonging to the same species exhibit identical variable regions within the 16S rRNA molecule, clinical isolates from the M. tuberculosis complex were obtained from K. Schröder, Forschungsinstitut Borstel, FRG. Thirteen isolates of M. tuberculosis, two isolates of M. bovis and two isolates of M. africanum were investigated, all of which were found to possess a 16S rRNA sequence identical to that of the M. tuberculosis strain shown in Fig. 3 (data not shown). These results demonstrate conservation of the variable nucleotides within the 16S rRNA molecule at the species level.

Nucleic acid technology has allowed bacterial classification to include elucidation of evolutionary relationships by comparative analysis of functionally equivalent macromolecules (Zuckerkandl & Pauling, 1965). Indeed, the determination of rRNA sequences has become a powerful tool for the systematic classification of bacteria (Fox et al., 1980; Woese et al., 1983; Gray et al., 1984; Woese, 1987). It has only recently emerged that parts of the 16S rRNA may be used to identify micro-organisms (Göbel & Stanbridge, 1987; Musial et al., 1988; Rossau et al., 1989; Stackebrandt et al., 1989), thus providing an alternative to the classical determinative approach, in which phenotypic criteria are used to identify species. Because of the overall conserved nature of rRNA, it is difficult to define stretches of nucleotides specific for very closely related taxa which may only be divided at a subspecific level. Thus, phenotypic distinctions can be made for M. tuberculosis, M. bovis and M. africanum, but 100% homology in the 16S rRNA structure rules out differentiation at this level. Similarly, M. gasitri differs from M. kansasii in phenotypic features, although both species share an identical small subunit rRNA sequence.

Until now, the identification of cultural isolates of mycobacteria based on phenotypic and chemotaxonomic features has traditionally been laborious and time-consuming. In addition, identification of mycobacterial isolates is complicated by disparities in the growth and metabolic rates of members of different species, and by the need for a variety of specialized and complex tests. In contrast, our approach is directed at the rapid determination of 16S rRNA sequences shown to be species-specific. Barely grown primary cultures suffice, circumventing the need for complex biochemical tests, and for subcultivating strains after their initial isolation. Currently 5000 mycobacterial cells are sufficient to yield a sequenceable amplified gene fragment. This level of sensitivity can be further improved (to 100 cells) by first reverse-transcribing the rRNA into cDNA before using it in the PCR (data not shown). The rapidity (from cell pellet to information on the nucleotide sequence within 2 d) and technical simplicity of the approach presented in this paper compare favourably with the traditional determination of phenotypic characters.

The authors thank K. Schröder, J. Kazda and J. Jorgensen for providing strains, M. Emde and A. Ohrdorf for technical assistance, E. C. Maibom for typing the manuscript and D. Bitter-Suermann for continuous encouragement. This work was supported by a grant from the Bundesministerium für Forschung und Technologie (01 K1 89117) to E. C. Böttger.
References


