Use of rpoB sequences for phylogenetic study of Mycoplasma species

Kil-Soo Kim a,1,2, Kwan Soo Ko b,1,3, Myoung-Woong Chang c, Tae Wook Hahn d, Seong Karp Hong b, Yoon-Hoh Kook b,*

a Department of Laboratory Animal Sciences, Seoul National University College of Medicine, 28 Yongon-dong, Chongno-gu, Seoul 110-799, South Korea
b Department of Microbiology and Cancer Research Institute, Institute of Endemic diseases, SNUMRC, Seoul National University College of Medicine and Clinical Research Institute, Seoul National University Hospital, 28 Yongon-dong, Chongno-gu, Seoul 110-799, South Korea
c Department of Microbiology, Kosin Medical College, 24 Annam-dong, Seo-gu, Busan 602-702, South Korea
d Department of Veterinary Medicine, College of Animal Resources Science, Kangwon National University, Kangwon-do 200-701, South Korea

Received 1 September 2002; received in revised form 1 February 2003; accepted 17 July 2003
First published online 23 August 2003

Abstract

rpoB sequences encoding the β-subunit of RNA polymerase were determined in 26 Mycoplasma species for phylogenetic study. The portion of rpoB DNA used in this study showed a high degree of variation in terms of size and sequence among species. The rpoB phylogenies inferred from amino acid and nucleotide sequences were used to divide the mycoplasmas into two groups, a ‘pneumoniae group’ and a ‘hominis group’, which was consistent with the result from 16S rDNA sequence analysis. However, phylogenetic relationships within these groups differed in the two gene trees, which were supported by the incongruence length difference (ILD) test. This indicates that multiple gene sequences must be applied to infer accurate phylogenetic relationships among the mycoplasmas. The rpoB sequence, and especially the deduced amino acid sequence, offers a good alternative marker.

Keywords: Mycoplasma; Ureaplasma; rpoB; Phylogeny

1. Introduction

The members of Mycoplasma and Ureaplasma are free-living, wall-lacking bacteria with small genomes, a low G+C content of DNA, and unusual nutritional requirements [1]. They are widely distributed in nature as parasites of humans, mammals, birds, reptiles, arthropods, and plants [2]. So far, 13 species of Mycoplasma and two species of Ureaplasma have been isolated from humans [1,3]. It is known that five Mycoplasma species (Mycoplasma pneumoniae, Mycoplasma hominis, Mycoplasma genitalium, Mycoplasma fermentans, and Mycoplasma penetrans) and the two Ureaplasma species can cause human disease [1].

Moreover, many mycoplasmas are pathogenic to animals, and therefore are of concern in veterinary medicine [4]. In addition, mycoplasmas frequently cause serious problems by contaminating cell cultures. It has been reported that 10–87% of cell cultures are infected by mycoplasmas [2]. Mycoplasma hyorhinis, Mycoplasma orale, Mycoplasma arginini, M. fermentans, Mycoplasma pirum, and Acholeplasma laidlawii are the dominant contaminants [2].

Phylogenetic studies of the genera Mycoplasma and Ureaplasma have been based upon 16S rDNA sequences [3–6]. However, some discrepancies in clusters of closely related species exist, which seem to depend on the analytical methods and standards used. As reported previously, phylogenetic analysis based upon the 16S rDNA sequences alone may provide only a limited understanding of their relationships and evolutionary history. In addition, although 16S rDNA is by far the most widely used [7], it is critically limited as a phylogenetic marker in Mycoplasma, because there are two or more 16S rRNA operons in many Mycoplasma species [4] and these show high intraspecific variations [8,9]. Thus, protein-encoding genes have been suggested and used instead of 16S rDNA [10], because it is believed that a combination of multiple ge-
ngetic markers can increase the accuracy and reliability of a phylogenetic scheme [10]. However, no protein-coding gene has yet been applied to the phylogenetic study of Mycoplasma and Ureaplasma species.

In this study, we used a partial rpoB sequence for the phylogenetic study of mycoplasmas. rpoB encodes a β-subunit of RNA polymerase, and has previously been demonstrated to be an appropriate marker for the phylogenetic study of several bacteria, such as enteric bacteria, Mycobacterium, Borrelia, and Legionella [11–14]. In this study, based on the newly determined rpoB DNA sequences, phylogenetic relationships between the mycoplasmas were inferred and compared with the results obtained from 16S rDNA sequences.

2. Materials and methods

2.1. Strains

Twenty-three reference strains of Mycoplasma, two strains of Ureaplasma species, and one strain of A. laidlawii were used in this study (Table 1). M. genitalium, which was isolated from an AIDS patient, was kindly provided by Dr. Shyh-Ching Lo (Armed Forces Institute of Pathology, USA).

2.2. DNA extraction, polymerase chain reaction (PCR) amplification, and sequencing

DNA extraction, PCR amplification, and sequencing reaction were performed as previously described by Ko et al. [14]. DNAs were extracted using the bead beater-phenol extraction method [12–14]. To amplify a portion of rpoB DNA in Mycoplasma and Ureaplasma species, two primer sets were used. A primer pair MPF1 (5′-GAT GAT CAC TTA GGT AA-3′) and MPRI (5′-CTT CNG GTG TTT CAA TNG G-3′) was used to amplify a portion of rpoB DNA in the pneumonieae group [5]. Another primer pair MPF2 (5′-AGA TGA YCC NGA TTC A-3′) and MPRI2 (5′-ACA AAT TCT TCC ATA GTG AGT-3′) was used in the hominis group [5]. The reaction mixture for PCR was then subjected to 30 cycles of: 30 s at 95°C for denaturation, 30 s at 42°C for annealing, 1 min at 72°C for extension, followed by a final extension at 72°C for 5 min (model 9700 Thermocycler; Perkin-Elmer Cetus).

2.3. Phylogenetic analyses

Raw sequences were analyzed and concatenated by DNASTAR (Madison, WI, USA). Multiple alignments were first accomplished for the inferred amino acid sequences. The amino acid sequences were deduced by the MegAlign program in DNASTAR (Madison, WI, USA). Based on the alignment of deduced amino acid sequences, aligned nucleotide sequences were obtained. Aligned sequences were checked manually. Phylogenetic trees of amino acid and nucleotide data sets were constructed using the neighbor-joining (NJ), maximum parsimony (MP), and maximum likelihood (ML) methods in the PAUP program [15]. The rpoB sequence of A. laidlawii determined in this study and that of Spiroplasma citri retrieved from the GenBank (accession number U25815) were used as outgroups (Table 1). Branch supporting values were evaluated with 1000 bootstrap replications. The 16S rDNA gene sequences of 26 strains were obtained from the GenBank and were used to infer phylogenetic relationships and these were compared with the phylogenetic relationships inferred from the results of the rpoB analysis. Incongruence length difference (ILD) test (partition homogeneity test in the PAUP program) [16,17] was conducted to determine whether the rpoB and the 16S rDNA gene data sets were coalesscent, using the PAUP program [15].

2.4. Nucleotide sequence accession numbers

The rpoB DNA sequences determined in this study were deposited in GenBank. Their accession numbers are listed in Table 1. The accession numbers of the 16S rDNA se-

---

Table 1

<table>
<thead>
<tr>
<th>Bacterial species and GenBank accession numbers of their rpoB sequences used in this study</th>
<th>Species</th>
<th>Strain number</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. anatis</td>
<td>ATCC 25524</td>
<td>AY191418</td>
<td></td>
</tr>
<tr>
<td>Mycoplasma arthritidis</td>
<td>ATCC 19611</td>
<td>AY191419</td>
<td></td>
</tr>
<tr>
<td>M. bovis</td>
<td>ATCC 25523</td>
<td>AY191420</td>
<td></td>
</tr>
<tr>
<td>Mycoplasma buccale</td>
<td>ATCC 23636</td>
<td>AY191421</td>
<td></td>
</tr>
<tr>
<td>M. fermentans</td>
<td>ATCC 19989</td>
<td>AY191422</td>
<td></td>
</tr>
<tr>
<td>M. gallinarum</td>
<td>ATCC 15319</td>
<td>AY191423</td>
<td></td>
</tr>
<tr>
<td>Mycoplasma gallisepticum</td>
<td>ATCC 19610</td>
<td>AY191424</td>
<td></td>
</tr>
<tr>
<td>M. genitalium</td>
<td>ATCC 33530</td>
<td>AY191425</td>
<td></td>
</tr>
<tr>
<td>M. hominis</td>
<td>ATCC 23114</td>
<td>AY191426</td>
<td></td>
</tr>
<tr>
<td>M. hyopneumoniae</td>
<td>ATCC 25617</td>
<td>AY191427</td>
<td></td>
</tr>
<tr>
<td>M. hyorhinis</td>
<td>ATCC 23839</td>
<td>AY191428</td>
<td></td>
</tr>
<tr>
<td>M. iowae</td>
<td>ATCC 33552</td>
<td>AY191429</td>
<td></td>
</tr>
<tr>
<td>M. lipophilum</td>
<td>ATCC 27104</td>
<td>AY191430</td>
<td></td>
</tr>
<tr>
<td>M. orale</td>
<td>ATCC 23714</td>
<td>AY191431</td>
<td></td>
</tr>
<tr>
<td>M. penetrans</td>
<td>ATCC 27160</td>
<td>AY191432</td>
<td></td>
</tr>
<tr>
<td>M. pirum</td>
<td>ATCC 19433</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. pneumoniae</td>
<td>ATCC 19531</td>
<td>AY191434</td>
<td></td>
</tr>
<tr>
<td>M. primatum</td>
<td>ATCC 25498</td>
<td>AY191435</td>
<td></td>
</tr>
<tr>
<td>M. pullorum</td>
<td>ATCC 33553</td>
<td>AY191436</td>
<td></td>
</tr>
<tr>
<td>M. pulmonis</td>
<td>ATCC 19612</td>
<td>AY191437</td>
<td></td>
</tr>
<tr>
<td>Mycoplasma salivarium</td>
<td>ATCC 23064</td>
<td>AY191438</td>
<td></td>
</tr>
<tr>
<td>Mycoplasma spermatothelium</td>
<td>ATCC 49695</td>
<td>AY191439</td>
<td></td>
</tr>
<tr>
<td>M. synoviae</td>
<td>ATCC 27399</td>
<td>AY191440</td>
<td></td>
</tr>
<tr>
<td>U. parvum</td>
<td>ATCC 27813b</td>
<td>AY191441</td>
<td></td>
</tr>
<tr>
<td>U. urealyticum</td>
<td>ATCC 27814b</td>
<td>AY191442</td>
<td></td>
</tr>
<tr>
<td>A. laidlawii</td>
<td>ATCC 23206</td>
<td>AY191443</td>
<td></td>
</tr>
<tr>
<td>S. citri</td>
<td>RHA2HP</td>
<td>U25815</td>
<td></td>
</tr>
</tbody>
</table>

*A strain isolated from an AIDS patient by Dr. Shyh-Ching Lo (Armed Forces Institute of Pathology, USA).

*bATCC 27813, U. urealyticum serovar 1; ATCC 27814, U. urealyticum serovar 2.

*pcoB sequence retrieved from GenBank.
quences retrieved for phylogenetic analysis are indicated within parentheses on the right of the species name in Fig. 3.

3. Results

3.1. \( \text{rpoB} \) sequences

Amplified \( \text{rpoB} \) DNAs showed marked size variations between species. In the 'pneumoniae group', the sizes of the PCR products, obtained using the primer set MPF1 and MPR1, ranged from 405 (\( \text{Mycoplasma iowae} \)) to 470 bp (\( \text{Ureaplasma urealyticum} \) and \( \text{Ureaplasma parvum} \)). In the 'hominis group', the PCR product obtained using the primersetMPF2 and MPR2, which are within MPF1 and MPR1, respectively, was 355 bp though \( \text{Mycoplasma hyopneumoniae} \) and \( \text{M. hyorhinis} \) produced larger products (358 bp).

Due to the extensive heterogeneity of the nucleotide sequences, rational alignment, using only nucleotide sequences, was not possible. Thus, amino acid sequences were deduced from the determined nucleotide sequences and used for phylogenetic analysis. To obtain sensible amino acid sequence alignments, three insertion–deletion (ID-1, -2, and -3) regions were needed. Seven amino acids of eight species in the 'pneumoniae group' are deleted in ID-1. In ID-2, gap of a single amino acid is present in 15 species of the 'hominis group' except \( \text{M. hyopneumoniae} \), \( \text{M. hyorhinis} \), and \( \text{A. laidlawii} \), and \( \text{S. citri} \).

Fig. 1. Multiple alignment of the deduced amino acid sequences of \( \text{rpoB} \). The identical sequences to consensus are indicated as dots. Three insertion–deletion regions (ID-1, -2, and -3) are contained by the gray box, and the region which is related to rifampin resistance in an open box. Asterisks indicate amino acids involved in the rifampin resistance of \( \text{E. coli} \).

In spite of the extensive insertions or deletions in multiple alignment of amino acid, no gap is required in the rifampin-resistant region (Fig. 1). It is well known that one of the peculiar traits of \( \text{Mollicutes} \) is its resistance to rifampin [2]. The N526 (numbering as in \( \text{Escherichia coli} \) [18]), which was reported as being related to the rifampin resistance in \( \text{Mollicutes} \) species [19], is taken as the consensus amino acid found in almost all mycoplasmas. However, the relationships of T526 in \( \text{M. hyopneumoniae} \), S526 in \( \text{M. hyorhinis} \), and Q526 in \( \text{A. laidlawii} \), with rifampin resistance are not known.

3.2. Phylogenetic relationships of \( \text{Mycoplasma} \) and \( \text{Ureaplasma} \)

Fig. 2 shows the phylogenetic relationships of 25 species in genera \( \text{Mycoplasma} \) and \( \text{Ureaplasma} \) based on the deduced amino acid and nucleotide sequence of \( \text{rpoB} \). These trees were constructed by the method of NJ, which were
wholly consistent with those by MP and ML methods (data not shown). Both trees clearly separate the *Mycoplasma* species into two groups, a ‘pneumoniae group’ and a ‘hominis group’, which is concordant with the results of previous studies [4,6]. Cluster B (ten species) and cluster C (five species), which were previously classified in a partial 16S rDNA analysis [3], were also well preserved in the ‘hominis group’ of rpoB trees (Fig. 2), which were quite similar but showed some differences in detail (Fig. 2). For examples, clade of *M. penetrans* and *M. iowae* in the ‘pneumoniae group’ formed a subgroup with two species of *Ureaplasma* in the amino acid tree (Fig. 2A) but did not in the nucleotide tree (Fig. 2B). In the hominis group, *Mycoplasma bovis* clustered with *Mycoplasma primatum* by amino acid phylogeny, with high bootstrap support (91%), but not by nucleotide phylogeny. Placement of *Mycoplasma anatis* was also different in two trees.

The phylogeny inferred from the 16S rDNA sequences from GenBank is also shown in Fig. 3. Five clusters defined by 16S rDNA sequences by Pettersson et al. [6] (Fig. 3) did not coincide completely with those of the rpoB phylogenies. Five species of ‘*M. hominis* cluster’, which corresponds to ‘cluster C’ as described by Yoshida et al. [3], also formed a group in the rpoB trees. Their relationships in the 16S rDNA tree (Fig. 3) were identical to those of the rpoB tree constructed by nucleotide sequences (Fig. 2B). *M. hyopneumoniae* and *M. hyorhinis*, which form the ‘*Mycoplasma neurolyticum* cluster’, were also grouped in the rpoB trees. However, members of other clusters in the 16S rDNA phylogeny were not congruent with those in the rpoB phylogeny. For example, *Mycoplasma pulmonis* grouped with *M. hyopneumoniae* and *M. hyorhinis* of the ‘*M. neurolyticum* cluster’ in the 16S rDNA tree (Fig. 3), but not in the rpoB trees (Fig. 2). *Mycoplasma gallinarum* was included in the ‘*Mycoplasma lipophilum* cluster’ by 16S rDNA analysis, but it clustered with *Mycoplasma synoviae*, *M. anatis*, and *Mycoplasma pullorum* in the rpoB analysis, which are members of the ‘*M. synoviae* cluster’ by the 16S rDNA phylogeny. In addition, though six species of the ‘*M. lipophilum* cluster’ formed a distinct group in the 16S rDNA phylogeny (Fig. 3), *M. lipophilum* was not included in any cluster of the rpoB amino acid tree. *M. lipophilum*, *M. primatum*, and *M. bovis* were also not included in any cluster of the rpoB nucleotide tree (Fig. 2).

### 3.3. ILD test

To evaluate the significance of incongruence between the 16S rDNA and rpoB phylogenies, we performed the ILD test [16] using 1000 replicates. The result of ILD test

**A) Amino acid**

![Amino acid tree](FEMSLE_11180_23-9-03)

**B) Nucleotide**

![Nucleotide tree](FEMSLE_11180_23-9-03)
showed non-combinability ($P < 0.001$), which means significant incongruence between the two data sets [17], and that it is not an artifact of tree reconstruction. To measure the extent of incongruence between two data sets, inconsistent species were progressively removed. The analysis was repeated until the incongruence was no longer significant ($P > 0.05$) [17,20]. Six species, $M$. gallinarum, $M$. synoviae, $M$. primatum, $M$. lipophilum, M. bovis, and $M$. pulmonis, had to be excluded from the data sets to achieve a significant threshold ($P > 0.05$).

4. Discussion

So far, the inferred phylogenetic relationships of mycoplasmas have been exclusively based on the 16S rDNA sequences [4,5,7,21]. Phylogenetic analysis based on 16S rDNA sequences classified mycoplasmas into several groups or clusters [3–6], and this was found to be generally concordant with biochemical properties [6]. However, it has been frequently indicated that phylogenetic conclusions based on single gene sequences may be hasty and insufficient. Thus, the importance of comparing results obtained from multiple loci has been highlighted for organismal phylogeny [10,22]. In particular, multiple copies of rrn operons [4] and intraspecific sequence polymorphisms in Mycoplasma species [8,9] require some care in the interpretation of the results of a single 16S rDNA sequence.

Therefore, in this study, we applied the rpoB sequence to the phylogenetic analysis of mycoplasmas and compared the results obtained with the results of 16S rDNA sequences. Several interesting findings were observed as a result of the rpoB sequence analysis. First, the rpoB DNA regions of mycoplasmas were found to show considerable sequence divergence. To obtain an aligned data set, the introduction of several gapped regions (ID-1, -2, and -3) was required (Fig. 1). The aligned gap patterns were found to differ for closely related species, such as $M$. penetrans and $M$. iowae, which showed 91.9 and 87.0% sequence similarity of 16S rDNA and rpoB, respectively, and these were consistently grouped together in 16S rDNA and rpoB trees. Although the rpoB DNA region has been utilized in previous studies of the phylogenetic analysis and identification of bacteria such as Mycobacterium [12], Borrelia [13], Bartonella [23], Staphylococcus [24], and Legionella [14], no gap within the same genus was necessary to align their rpoB sequences. Thus, the sequence divergence of rpoB in mycoplasmas is quite unusual. Possibly, it was caused by an accelerated evolutionary rate due to the absence of genes related to mutation repair, such as the mut homologs [25]. Alignment of gaps due to the insertion–deletion and considerable sequence divergence can be advantageous for the design of species-specific primers for
the detection of mycoplasmas in humans, animals, or contaminated cell lines. In spite of extensive divergences between species, high similarities among the wild-type strains within the same species such as *M. pneumoniae* and *M. hyorhinis* also make them possible. For example, *rpoB* sequences of Korean isolates, 10 strains of *M. pneumoniae* and 19 strains *M. hyorhinis* showed more than 99.7 and 99.1% similarities, respectively. Considering the remarkable sequence divergences and discrete groupings in both the 16S rDNA and *rpoB* analyses and their different biochemical properties [6], it would be more plausible to classify members of *Mycoplasma* as two separate genera. To suffice this taxonomic consideration, however, more evidence, such as other gene sequences and phenotypic characteristics are necessary.

The second interesting finding concerns the existence of an incongruence between the *rpoB* and 16S rDNA trees. The ILD test supported the notion that the apparent inconsistent grouping of several species is not due to an artifact in tree reconstruction, but rather to the incongruent nature of the two data sets. The difference between the two data sets was caused by six species, namely *M. synoviae*, *M. bovis*, *M. primatum*, *M. gallinarum*, *M. lipophilum*, and *M. pulmonis*. This result emphasizes that molecular markers other than the 16S rDNA sequence must be applied to obtain a precise phylogenetic and evolutionary scheme. However, the 'pneumoniae group' and the 'hominis group', as defined by Weisburg et al. [5], are also well conserved by *rpoB* analysis.

Third, differences were found between trees based on the deduced amino acid and nucleotide sequences of *rpoB* DNA. This phenomenon seems to have originated from several synonymous mutations at the third nucleotide of the corresponding codons, which do not result in amino acid changes. In this case, multiple substitutions at the same site may obscure the actual phylogenetic relationships. Thus, phylogeny from amino acid standpoint rather than from nucleotides may be considered to better reflect the actual evolutionary route of the organisms. In previous studies using *rpoB* sequences of other genera, the amino acid sequence could not be utilized because of its very low divergence within the same genus. However, this is not the case for mycoplasmas. The amino acid sequence of RpoB may be a good molecular marker for its evolutionary study. For example, the grouping of *M. primatum* and *M. bovis* in the 16S rDNA phylogeny (Fig. 3) is conserved in the *rpoB* amino acid tree (Fig. 2A) but not in the *rpoB* nucleotide tree (Fig. 2B). In this case, it may be suggested that the grouping of *M. primatum* and *M. bovis* is more parsimonious and thus is more plausible.

In conclusion, this study shows that *rpoB* is a good alternative molecular marker to 16S rDNA for the phylogenetic analysis of mycoplasmas. The extreme sequence divergence shown by *rpoB* DNA in the *Mycoplasma* species is also widespread enough to differentiate or identify wild types. In particular, considerable heterogeneities of amino acid sequences within the same genus exist in this species.

**Acknowledgements**

This work was supported by a grant of the Korea Health 21 R&D Project, Ministry of Health and Welfare, Seoul, Republic of Korea (01-PJ10-PG6-01GM03-0002), and in part by the BK21 project for Medicine, Dentistry, and Pharmacy.

**References**


