Phylogeny of the *Mycoplasma mycoides* cluster based on analysis of five conserved protein-coding sequences and possible implications for the taxonomy of the group

Lucía Manso-Silván, Xavier Perrier and François Thiaucourt

INTRODUCTION

The *Mycoplasma mycoides* cluster, as described by Cottew et al. (1987), represents a group of six closely related mycoplasmas that are pathogenic for ruminants and which are currently referred to as *Mycoplasma mycoides* subsp. *mycoides* Small Colony biotype (MmmSC) and Large Colony biotype (MmmLC), *M. mycoides* subsp. *capri* (Mmc), *Mycoplasma capricolum* subsp. *capricolum* (Mcc), *M. capricolum* subsp. *capripneumoniae* (Mccp) and *Mycoplasma* sp. bovine group 7 of Leach (MBG7). In addition, on the basis of 16S rRNA gene sequence similarity, other species such as the pathogen *Mycoplasma putrefaciens* (Weisburg et al., 1989) and the related saprophytic species *Mycoplasma cottewii* and *Mycoplasma yeatsii* clustered in a group that was distinct from *Mycoplasma putrefaciens* and they were all clearly separated from the *M. mycoides* cluster. In conclusion, this approach has allowed us to assign phylogenetic positions to all members of the *M. mycoides* cluster and related species and has proved the need to adjust the existing taxonomy. Furthermore, this method may be used as a reference technique to assign an unequivocal position to any particular strain related to this cluster and may lead to the development of new techniques for rapid species identification.


The GenBank/EMBL/DDBJ accession numbers of the partial *fusA*, *glyQ*, *gyrB*, *lepA* and *rpoB* gene sequences of 26 strains analysed here are listed in Supplementary Table S1.

Accession numbers of the sequences used in this study are available with the online version of this paper.
group of strains that are serologically distinct from other bovine mycoplasmas (Leach, 1967) and which cause mastitis and polyarthritis in cattle.

Because the different members of the \textit{M. mycoides} cluster share many genotypic and phenotypic traits, their classification and evolutionary relationships have been difficult to establish. This may explain why the group of strains known as MBG7 remains unassigned, while Mccp was named only relatively recently (Leach et al., 1993). Initial classification within this cluster was attempted based on DNA–DNA hybridization and biochemical and serological methods as well as 2D PAGE patterns (Cottew et al., 1987), which resulted in differentiation of the \textit{M. mycoides} subspecies on one side and the \textit{M. capricolum} subspecies on the other. The position of MBG7 was controversial due to conflicting data obtained from the various methods applied and remains unresolved, though recent studies concur in favour of classification of this group of strains within a subspecies of \textit{M. capricolum} (Harasawa et al., 2000; Pettersson et al., 1996; Thiaucourt et al., 2000).

Another controversial issue is the distinction of MmmLC and Mmc as separate subspecies: numerous investigations, based on both genetic and antigenic analyses, suggest that they may be classed together as a single subspecies within the \textit{M. mycoides} subcluster (Costas et al., 1987; Leach et al., 1989; Monnerat et al., 1999; Olsson et al., 1990; Rodwell, 1982; Salih et al., 1983; Taylor et al., 1992; Vilei et al., 2006).

The phylogeny of the \textit{M. mycoides} cluster inferred from 16S rRNA gene sequences (Pettersson et al., 1996) relied on too few nucleotide variations to construct a reliable phylogenetic tree. It was then proposed that further studies be based on sequence analysis of alternative genes. To this aim, five protein-coding genes from 24 strains, as well as from MmmSC strain 8740-Rita, were deposited in GenBank under the accession numbers shown in Supplementary Table S1 (available in IJSEM Online). Sequences of these five genes were also retrieved from the published genomes of the type strains MmmSC PG1\textsuperscript{T} (GenBank accession no. BX293980) and Mcc California kid\textsuperscript{T} (CP000123). A total of 28 sequences were therefore available for each of the five genes studied.

\textbf{METHODS}

\textbf{Determination of partial gene sequences.} A sample of strains from various geographical origins was chosen carefully, including representatives of all members of the \textit{M. mycoides} cluster and related species and adapting the number of strains to the known variability within each group. The 28 strains analysed are listed in Table 1, with reference to their origin. The strains were characterized according to results obtained using two different techniques: growth inhibition with hyperimmune sera (Clyde, 1983) and dot immunobinding (Poumarat et al., 1992).

Mycoplasma strains were cultured in modified Hayflick’s broth [PPLO broth without crystal violet (21 g l\textsuperscript{-1}), 20 % horse serum de-complemented for 1 h at 56 °C, 5 % fresh yeast extract, 0.2 % glucose, 0.4 % sodium pyruvate] at 37 °C in 5 % CO\textsubscript{2} and were harvested in the late exponential phase of growth. Preparation of PCR templates in lysis buffer was based on the method of Miserez et al. (1997).

The selection of primer pairs for amplification of \textit{fusA}, \textit{glpQ}, \textit{gyrB}, \textit{lepA} and \textit{rpoB} sequences (Table 2) was based on initial alignments of sequences obtained from published mycoplasma genomes (http://cbi.labri.fr/outils/molligen/home.php) as well as from ongoing genome sequencing projects (MmmLC strain 95010-C1 and MmmSC strain 8740-Rita). Amplification of each of the target sequences was performed in 50 μl reactions containing 1 × \textit{Taq} buffer (Qiagen) (giving a final MgCl\textsubscript{2} concentration of 1.5 mM), 150 μM dCTP and dGTP, 300 μM dATP and TTP (dNTPs from Roche), 0.4 μM each of the corresponding forward and reverse primer, 1 U \textit{Taq} polymerase (Qiagen) and 1 μl sample. The PCR thermal program consisted of an initial denaturation step of 2 min at 94 °C followed by 35 cycles of 30 s at 94 °C, 30 s at 52 °C and 45 s at 72 °C and a final extension step of 5 min at 72 °C. An exception was made for the PCR with primers GlpQ-L–F–R, for which the annealing temperature was 45 °C, with the reaction conditions otherwise remaining the same.

The primer pairs listed in Table 2 were also used for sequencing of the corresponding PCR products by Genome Express (Meylan, France). Sequences were aligned using the AlignX tool of Vector NTI 9.0 (Invitrogen) and they were trimmed to the same size, providing fragments of 561 bp (\textit{fusA}), 538 bp (\textit{glpQ}), 533 bp (\textit{gyrB}), 657 bp (\textit{lepA}) and 625 bp (\textit{rpoB}) for phylogenetic analyses. No deletions/insertions were observed except for the \textit{glpQ} sequences of the two \textit{M. putrefaciens} strains, which contained an extra codon. The partial nucleotide sequences determined in this study for each of the five protein-coding genes from 24 strains, as well as from MmmSC strain 8740-Rita and MmmLC strain 95010-C1, were deposited in GenBank under the accession numbers shown in Supplementary Table S1 (available in IJSEM Online). Sequences of these five genes were also retrieved from the published genomes of the type strains MmmSC PG1\textsuperscript{T} (GenBank accession no. BX293980) and Mcc California kid\textsuperscript{T} (CP000123). A total of 28 sequences were therefore available for each of the five genes studied.

\textbf{Phylogenetic analyses.} Sequence analyses were conducted using \textit{MEGA} 3.1 (Kumar et al., 2004; http://www.megasoftware.net), whereas Darwin 5.0 (2004) and \textit{MEGA} 3.1 were used for diversity analyses. Phylogenetic trees were constructed using the neighbour-joining algorithm (Saitou & Nei, 1987) based on pairwise dissimilarities between strains. Since other methods such as minimum evolution and maximum parsimony by close-neighbour interchange (\textit{MEGA} 3.1) generated nearly identical topologies, the neighbour-joining method was retained for its properties of convergence and consistency, as well as low computational complexity.

Pairwise distances were calculated as the proportion of non-matching sites between pairs of sequences. The three positions corresponding to an extra codon in the \textit{glpQ} sequences of the \textit{M. putrefaciens} strains were deleted to avoid gaps. Since the sequences were highly similar, the effect of multiple substitutions was considered negligible and no correction was applied to dissimilarities. The mean net distance between groups was also calculated as the arithmetic mean of the pairwise distances between two groups corrected by the corresponding mean distance within groups (\textit{MEGA} 3.1).

A bootstrap analysis with 500 replicates was performed to test the stability on randomly chosen sets of positions. To analyse the stability through the sample of strains, the influence of each sequence on the tree structure was estimated using a tool available in Darwin 5.0. In a jackknife-like procedure, each unit is successively removed and a partial tree is inferred from all units except the removed one. This tree is compared to the initial tree using a topological criterion, the quartet tree distance (Qd), which measures among all quartets
Table 1. *Mycoplasma* strains analysed in this study

Suppliers of strains are abbreviated as follows: Arist. U., Professor K. Sarris, Aristoteles University, Thessaloniki, Greece; Bern U., Professor J. Nicolet, Institute for Veterinary Bacteriology, University of Bern, Switzerland; BgVV, Dr K. Sachse, Federal Institute for Health Protection of Consumers and Veterinary Medicine, Division 4, Jena, Germany; Cal. U., Dr A. J. DaMassa, Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, USA; CIRAD, Centre de Coopération Internationale en Recherche Agronomique pour le Développement, UPR-15, Montpellier, France; CIRG, Dr Rajneesh Rana, Central Institute for Research on Goats, Makhdoom, Mathura, India; FDVR, Dr E. P. Lindley, Federal Department of Veterinary Research, Vom, Nigeria; IVRI, Dr V. P. Singh, Indian Veterinary Research Institute, Iztanagar, India; LANAVET, Dr A. Martrenchar, Laboratoire National Vétérinaire, Garoua, Cameroon; LPA, Dr J. Domenech, Laboratoire Central de Pathologie Animale, Bingerville, Ivory Coast; NVI-E, Dr Y. Fikrê, National Veterinary Institute, Debre Zeit, Ethiopia; NVI-S, Dr G. Böliske, National Veterinary Institute, Uppsala, Sweden; TiHo, Dr R. Schmidt, Tierärztliche Hochschule, Hannover, Germany.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Abbreviation</th>
<th>Host</th>
<th>Lesion/organ</th>
<th>Year</th>
<th>Country</th>
<th>Supplier/reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. mycoides</em> subsp. <em>mycoides</em> SC</td>
<td>PG1T* (=NCTC 10114T)</td>
<td>MmmSC-1</td>
<td>Bovine</td>
<td>Pleuropneumonia</td>
<td>NK</td>
<td>NK</td>
</tr>
<tr>
<td>8740-Rita*</td>
<td>MmmSC-2</td>
<td>Bovine</td>
<td>Pleuropneumonia</td>
<td>1987</td>
<td>Cameroon</td>
<td>LANAVET</td>
</tr>
<tr>
<td>95010-C1*</td>
<td>MmmLC-2</td>
<td>Caprine</td>
<td>Arthritis</td>
<td>1995</td>
<td>France</td>
<td>CIRAD</td>
</tr>
<tr>
<td>55507-1</td>
<td>MmmLC-3</td>
<td>Caprine</td>
<td>Arthritis</td>
<td>1998</td>
<td>Germany</td>
<td>TiHo</td>
</tr>
<tr>
<td>Kombolcha</td>
<td>MmmLC-4</td>
<td>Caprine</td>
<td>Pleuropneumonia</td>
<td>1975</td>
<td>Ethiopia</td>
<td>NVI-E</td>
</tr>
<tr>
<td>WK354</td>
<td>MmmLC-5</td>
<td>Caprine</td>
<td>Arthritis</td>
<td>1980</td>
<td>Switzerland</td>
<td>Bern U.</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>Mmc-2</td>
<td>Caprine</td>
<td>Mastitis</td>
<td>1975</td>
<td>France</td>
</tr>
<tr>
<td></td>
<td>N108</td>
<td>Mmc-3</td>
<td>NK</td>
<td>NK</td>
<td>1977</td>
<td>Nigeria</td>
</tr>
<tr>
<td>2003-045-C2</td>
<td>Mmc-4</td>
<td>Caprine</td>
<td>Pneumonia</td>
<td>&lt;2003</td>
<td>India</td>
<td>CIRG</td>
</tr>
<tr>
<td>2002-054 (VP9L)</td>
<td>Mmc-5</td>
<td>Caprine</td>
<td>NK</td>
<td>&lt;2002</td>
<td>India</td>
<td>IVRI</td>
</tr>
<tr>
<td><em>M. capricolum</em> subsp. <em>capricolum</em></td>
<td>California kidT* (=NCTC 10154T =ATCC 27343T =CIP 104620T)</td>
<td>Mcc-1</td>
<td>Caprine</td>
<td>Arthritis</td>
<td>1955</td>
<td>USA</td>
</tr>
<tr>
<td>96038</td>
<td>Mcc-2</td>
<td>Caprine</td>
<td>NK</td>
<td>&lt;2002</td>
<td>India</td>
<td>IVRI</td>
</tr>
<tr>
<td>8086-1</td>
<td>Mcc-3</td>
<td>Caprine</td>
<td>Arthritis</td>
<td>&lt;1996</td>
<td>Greece</td>
<td>Arist. U.</td>
</tr>
<tr>
<td>90122 (C1547)</td>
<td>Mcc-4</td>
<td>Caprine</td>
<td>Mastitis</td>
<td>1980</td>
<td>France</td>
<td>CIRAD</td>
</tr>
<tr>
<td><em>M. capricolum</em> subsp. <em>capripneumoniae</em></td>
<td>9231-Abomsa</td>
<td>Mccp-1</td>
<td>Caprine</td>
<td>Pleuropneumonia</td>
<td>1982</td>
<td>Ethiopia</td>
</tr>
<tr>
<td>95043</td>
<td>Mccp-2</td>
<td>Caprine</td>
<td>Pleuropneumonia</td>
<td>1995</td>
<td>Niger</td>
<td>CIRAD</td>
</tr>
<tr>
<td><em>Mycoplasma</em> sp. bovine group 7 of Leach</td>
<td>PG50K* (=NCTC 10133)</td>
<td>MBG7-1</td>
<td>Bovine</td>
<td>Arthritis</td>
<td>1963</td>
<td>Australia</td>
</tr>
<tr>
<td>9733</td>
<td>MBG7-2</td>
<td>Caprine</td>
<td>Milk (no symptoms)</td>
<td>1993</td>
<td>India</td>
<td>BgVV</td>
</tr>
<tr>
<td>D424</td>
<td>MBG7-3</td>
<td>Bovine</td>
<td>Preputium (no symptoms)</td>
<td>&lt;1990</td>
<td>Germany</td>
<td>NVI-S (Christiansen &amp; Ernø, 1990)</td>
</tr>
<tr>
<td>Calf 1</td>
<td>MBG7-4</td>
<td>Caprine</td>
<td>Septicaemia</td>
<td>&lt;1990</td>
<td>Nigeria</td>
<td>NVI-S (Christiansen &amp; Ernø, 1990)</td>
</tr>
<tr>
<td><em>Mycoplasma</em> sp.</td>
<td>8756-C13</td>
<td>M sp. (wild)</td>
<td>Caprine</td>
<td>NK</td>
<td>&lt;1987</td>
<td>USA</td>
</tr>
<tr>
<td>Tours 2</td>
<td>Mput-2</td>
<td>Caprine</td>
<td>Mastitis</td>
<td>1972</td>
<td>France</td>
<td>CIRAD</td>
</tr>
<tr>
<td><em>M. cottewii</em></td>
<td>VIS* (=NCTC 11732T =ATCC 51347T =CIP 105678T)</td>
<td>Mcot</td>
<td>Caprine</td>
<td>Ear (no symptoms)</td>
<td>1981</td>
<td>Australia</td>
</tr>
<tr>
<td><em>M. yeatsii</em></td>
<td>GH* (=NCTC 11730T =ATCC 51346T =CIP 105675T)</td>
<td>Myea</td>
<td>Caprine</td>
<td>Ear (no symptoms)</td>
<td>1981</td>
<td>Australia</td>
</tr>
</tbody>
</table>

*Strains for which sequences were obtained from published genomes or from ongoing genome-sequencing projects.
†Proposed reference strain for this taxon.
‡The taxonomic position of strain 8756-C13 is uncertain.
(subsets of four sequences) the proportion of those that do not share the same topology. This distance is regarded as a measure of the influence that the removed sequence has on the tree structure. A high Qd value for a given unit means that this sequence conveys a strong and atypical signal, largely incompatible with the common signal of the other units, and is highly influential on the tree structure. On the other hand, a low Qd value means that the removed unit conveys a signal similar to that of other units and is not particularly influential on the structure. Similar and low Qd values are expected for a coherent set of units. The detection of particularly high Qd values warns of the presence of atypical units. Since tree construction methods are known to be very sensitive to atypical units, sequences with unusual strong influence were excluded in order to infer more reliable trees.

RESULTS AND DISCUSSION

Phylogenetic approach

Five conserved protein-coding genes, presenting greater variability than 16S rRNA gene sequences, were selected for phylogenetic analysis of the M. mycoides cluster. These genes were located at distant positions, distributed along the genome of sequenced mycoplasmas belonging to this cluster. Four of them are informational genes: fusA encodes the elongation factor G, gyrB encodes the DNA gyrase subunit B, lepA encodes the GTP-binding membrane protein or elongation factor LepA and rpoB encodes the DNA-dependent RNA polymerase β-subunit. These universally conserved genes have been selected as complementary to 16S rRNA gene analysis for bacterial identification and phylogenetic studies (Santos & Ochman, 2004) and rpoB has already been used for phylogenetic analysis of Mycoplasma species (Kim et al., 2003; Vilei et al., 2006). An operational gene, glpQ, was also included. It encodes glycerophosphoryl diester phosphodiesterase and has also been used to define prokaryote taxa (Schwan et al., 2005).

Assuming that the evolutionary history of organismal lineages in the M. mycoides cluster can be properly estimated in the form of a phylogenetic tree, it was proposed that this ‘species tree’ may be inferred from the phylogenies based on the five selected genes, the ‘gene trees’. However, it is now well established that different genes sampled from the same set of taxa can produce markedly different phylogenies because of limited variability, inadequate sampling or random convergence, but also because of biological phenomena related to evolutionary events such as horizontal transfer or selection pressure. The problem is then to reconcile the trees derived from different genes to approximate the species tree. Two approaches are generally proposed: consensus methods and total evidence. The former (strict consensus, relaxed consensus or, more recently, consensus networks) considers discordance as uncertainty and produces trees that are often largely unresolved. In contrast, the total evidence approach infers a tree from a set of concatenated gene sequences assuming that each gene tree contains, at least in part, a common phylogenetic signal that will emerge with sufficient sampling. It has been shown that total evidence is often efficient in overcoming the incongruences present in single-gene analyses (Rokas et al., 2003). Therefore, the approach chosen in this study was to infer the phylogeny of the M. mycoides cluster from a sequence generated by concatenation head-to-tail of the five selected gene sequences. However, since the expected statistical smoothing is not guaranteed with only five genes, the confidence attached to the inferred phylogeny would be strengthened if the common structure could be retrieved, at least at the level of the main subclusters, in each of the individual gene phylogenies or if the biological causes for strong incongruence on a particular gene could be recognized and explained. For this reason, each of the five gene trees was analysed separately and compared to the total evidence tree.

The M. mycoides cluster tree as inferred from a set of five concatenated protein-coding sequences

Primer pairs for amplification of fusA, glpQ, gyrB, lepA and rpoB sequences were selected on the basis of homologous regions detected by alignment of the available sequences of the type strains MmmSC PG1T and Mcc California kidT and sequences obtained through ongoing genome

<table>
<thead>
<tr>
<th>Table 2. PCR primers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR product sizes and 5’ positions of forward primers for each target were taken from the genome sequences of MmmSC PG1T (GenBank accession no. BX293980) and Mcc California kidT (CP000123).</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target</th>
<th>Name</th>
<th>Sequence (5’–3’)</th>
<th>Name</th>
<th>Sequence (5’–3’)</th>
<th>Product (bp)</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>fusA</td>
<td>FusA-F</td>
<td>TGAATTTTTTATGTTGGGAGAA</td>
<td>FusA-R</td>
<td>GGTAATTTTAATGTTTCAGCATATGAA</td>
<td>781</td>
<td>680</td>
</tr>
<tr>
<td>glpQ</td>
<td>GpQ-L-F</td>
<td>TGGATATGACGTATCTTACG</td>
<td>GpQ-L-R</td>
<td>TCAGGATAACCTGAATATACCC</td>
<td>695</td>
<td>1</td>
</tr>
<tr>
<td>gyrB</td>
<td>GyrB-F</td>
<td>TTTGGATGATCTGTTACCTGAATATACCC</td>
<td>GyrB-R</td>
<td>TTGTTGATGATCTGTTACCTGAATATACCC</td>
<td>635</td>
<td>359</td>
</tr>
<tr>
<td>lepA</td>
<td>LepA-L-F</td>
<td>AGAGAAAATTTTATGTTGGGAGAA</td>
<td>LepA-L-R</td>
<td>CTTGGGACGTTGGAATTTGAT</td>
<td>1097</td>
<td>19</td>
</tr>
<tr>
<td>rpoB</td>
<td>RpoB-D2</td>
<td>GCTCAATCTATGAATTACGATG</td>
<td>RpoB-D2</td>
<td>TCTTCTGGGAGAAATGTTGTC</td>
<td>824</td>
<td>1963</td>
</tr>
</tbody>
</table>
sequencing projects (MmmLC strain 95010-C1 and MmmSC strain 8740-Rita). The primers listed in Table 2 yielded the expected PCR products from all of the 24 additional strains belonging to the *M. mycoides* cluster and related species analysed in this study (Table 1), whereas unrelated mycoplasmas such as *Mycoplasma auris* UIAT failed to yield a PCR product. It must be noted that, as seen by analysis of available, unrelated mycoplasma gene sequences, it is unlikely that these primer pairs will be adequate for PCR amplification from other taxa. In all, 28 sequences were obtained for each of the five genes, which were aligned and trimmed to the same size, providing gene fragments ranging from 533 to 657 bp (Table 3). The five gene fragments corresponding to each of the 28 strains were concatenated head-to-tail, providing a set of sequences of 2914 bp for phylogenetic analysis.

As shown by pairwise distance analysis (Fig. 1), the strains of the *M. mycoides* cluster were closely related, with maximal divergence limited to 4.6 %, whereas strains of *M. putrefaciens*, *M. cottewii* and *M. yeatsii* strains were easily differentiated from them (14.6–16.3 % distance). The mean intragroup distance was 0 % for MBG7 and also very limited for MmmSC and Mccp (0.1 % dissimilarity), whereas MmmLC and Mmc respectively showed 1.5 and 0.8 % intraspecific divergence. Analysis of the mean net distance between groups revealed a very close identity of MmmLC and Mmc strains (0.3 % net distance) and placed MBG7 closer to Mcc (1.9 % net divergence, compared with 3.9 % net distance from MmmSC).

Phylogenetic analysis using the neighbour-joining method resulted in the tree topology shown in Fig. 2. The mycoplasmas from the *M. mycoides* cluster were arranged in two subclusters, one grouping the *M. mycoides* subspecies and the other comprising the *M. capricolum* subspecies and MBG7 strains. MmmSC strains constituted a separate branch within the *M. mycoides* subcluster, whereas MmmLC and Mmc strains could not be separated clearly from each other. The *M. capricolum* subcluster also comprised two branches, one containing Mcc and Mccp strains and the other represented by MBG7 strains. The uncharacterized strain 8756-C13 (labelled ‘M sp.’) was

<table>
<thead>
<tr>
<th>Gene Sequence</th>
<th>M. mycoides cluster and related species</th>
<th>M. mycoides cluster only</th>
</tr>
</thead>
<tbody>
<tr>
<td>fuaA</td>
<td>561 106 90</td>
<td>36 29</td>
</tr>
<tr>
<td>glpQ</td>
<td>538 185 165</td>
<td>55 47</td>
</tr>
<tr>
<td>gyrB</td>
<td>533 172 150</td>
<td>52 41</td>
</tr>
<tr>
<td>lepA</td>
<td>657 182 158</td>
<td>68 51</td>
</tr>
<tr>
<td>rpoB</td>
<td>625 149 126</td>
<td>50 35</td>
</tr>
<tr>
<td>Concatenated set</td>
<td>2914 794 686</td>
<td>261 203</td>
</tr>
</tbody>
</table>

**Table 3.** Comparison of the variability of different gene sequences analysed among members of the *M. mycoides* cluster and related species

The numbers of variable and phylogenetically informative sites (variables excluding singletons) were calculated from sequence alignments of 28 strains of the *M. mycoides* cluster and related species listed in Table 1 and from the alignment of the 23 strains belonging to the *M. mycoides* cluster. *M. putrefaciens*, *M. cottewii* and *M. yeatsii* strains were excluded from these latter comparisons, as well as the uncharacterized strain 8756-C13. For strain details see Table 1.
placed in an intermediate position between the cluster and the other related species. *M. cottewii* and *M. yeatsii* clustered together, separate from *M. putrefaciens*, and they were all clearly separated from the *M. mycoides* cluster. All the main branches indicated here were supported by maximum bootstrap values (100%).

**Assessment of the concordance of individual gene trees with the tree based on concatenated sequences**

To assess the adequacy of the phylogeny inferred from concatenated sequences, separate analyses of each of the five genes were performed on 23 strains belonging to the *M. mycoides* cluster. To facilitate comparisons, the trees were rooted using the uncharacterized strain 8756-C13. However, since atypical sequences such as the outgroup can deeply disturb tree structures, this was not included in the set of analysed strains but was grafted onto the inferred trees afterwards. For the same reason, influential units were investigated systematically by calculating Qd values, as described in Methods. In order to verify the stability of the phylogenetic signal along the gene sequences, separate analyses were performed on sequential subsequences. Finally, independent analyses were also performed on each of the three codon positions, which do not undergo the same evolutionary processes. The first two codon positions are strongly constrained, minimizing the risk of homoplasy, while the third codon position is often free. Modifications in the first two positions tend to result in a change in the amino acid residue. Substitutions in the second codon position were infrequent and were often associated with amino acid changes, accompanied by a change in the first position. The first codon position was therefore retained to avoid this possible redundancy.

The structure of each of the trees based on individual gene sequences was compared to the topology of the tree based on concatenated sequences, focusing comparisons on the distribution of the different subspecies according to the main branches supported by high bootstrap values.

**Tree based on fusA sequences.** The phylogenetic tree constructed from *fusA* sequences (Fig. 3a) was in general agreement with the tree based on concatenated sequences.
The two subclusters where clearly distinguished: *M. mycoides* subspecies on one side, *M. capricolum* subspecies and MBG7 on the other. Within the *M. mycoides* subscluster, the two MmmSC strains clustered together, though they were not neatly separated from the rest of the subcluster, where MmmLC and Mmc remained indistinguishable. As for the *M. capricolum* subgroup, Mccp strains could not be clearly separated from Mcc strains, whereas MBG7 constituted a distinct branch. All these main branches were supported by high bootstrap values (91–100 %).

**Tree based on glpQ sequences.** The general topology of the tree based on glpQ sequences (Fig. 3b) was similar to that based on fusA, with MmmSC strains more deeply embedded within the *M. mycoides* subcluster, being indistinguishable from the other strains.

**Tree based on gyrB sequences.** The tree inferred from gyrB sequences was incongruent with the phylogeny based on concatenated sequences (Fig. 4a). It differed from the fusA and glpQ trees in the position of the MmmSC strains, which left the *M. mycoides* subcluster to join the *M. capricolum* subcluster (98 % bootstrap). The MBG7 strains were attached to the *M. capricolum* subcluster, where Mccp constituted a separate branch. Independent analyses on each codon position revealed the same incongruent topology, and no influential units were identified among the 23 sequences. However, closer examination of variable nucleotide positions showed that the MmmSC strains shared several variant nucleotides with MBG7 strains, which attracted them towards the *M. capricolum* subcluster. These nucleotides were concentrated (7 of 10) in a short subsequence between positions 140 and 240. Therefore, an analysis restricted to the second half of the sequence (Fig. 4b) agreed globally with the concatenated sequence phylogeny, with MmmSC placed in the *M. mycoides* subcluster.

**Tree based on lepA sequences.** The tree derived from lepA sequences (Fig. 5a) was largely incongruent with the common structure and did not allow a clear discrimination of the species *M. mycoides* and *M. capricolum*. As in the
case of gyrB, the attraction between MmmSC and MBG7 was apparent, in this case inducing a separate group (74 % bootstrap) linked to the M. mycoides subcluster. The strains referred to as MmmLC-5 and MmmLC-3 were linked to the M. capricolum subcluster. These, as well as strain MmmLC-2, were detected as highly influential units, with Qd values over 0.22, while the remaining values were below 0.12. MmmLC-2 and MmmLC-3 showed specific nucleotide variations, while MmmLC-5 shared variant nucleotides with the M. capricolum subcluster. In fact, the nine variable sites present in the first half of the lepA sequence of MmmLC-5 were characteristic of M. capricolum, whereas the second half contained seven variant nucleotides corresponding to a typical M. mycoides sequence. Analysis of the second half of this sequence (Fig. 5b) restored the three erratic strains to their subcluster but did not suppress the attraction between MmmSC and MBG7. When the analysis was restricted to the initial codon position (Fig. 5c), the common structure was recovered (except for strain MmmLC-5), with MmmSC linked to the M. mycoides subcluster and MBG7 to the M. capricolum subcluster, indicating that their shared nucleotide variations, as well as the specific variant nucleotides of the MmmLC-2 and MmmLC-3 sequences, mainly concerned the least-constrained codon positions.

**Tree based on rpoB sequences.** The tree derived from rpoB sequences (Fig. 6a) was highly incongruent with the common structure and neither of the two subclusters could be clearly discriminated. Analyses based on partial sequences and on each codon position did not allow this structure to be clarified. However, strain MmmLC-1, corresponding to strain Y-goatR, was identified as a strongly influential unit (Qd value of 0.2, whereas the others did not exceed 0.02). In six variable positions, the sequence of strain Y-goatR shared variant nucleotides characteristic of M. capricolum, which were not shared by MBG7. This resulted in attraction of the MmmLC/Mmc group towards the M. capricolum subspecies, accompanied by repulsion of MBG7. The tree constructed by exclusion of strain Y-goatR (Fig. 6b) restored an M. capricolum subcluster including MBG7 and an M. mycoides subcluster, though variant nucleotides specific to MmmSC, partly shared by strain MmmLC-4, kept them in a separate group.

**Phylogeny of the M. mycoides cluster and taxonomic implications**

The phylogeny of the M. mycoides cluster has been inferred from a set of five concatenated protein-coding gene sequences and strengthened by separate analyses performed on each individual gene. Although incongruences were observed for certain genes, which explain the ambiguities that have been found in previous studies, the ‘mean structure’ represented by the concatenated sequence tree could be recognized behind gene specificities when reinforced in this manner. This phylogeny does not contradict that based on 16S rRNA gene sequences (Pettersson et al., 1996). However, the reduced variability of the 16S rRNA gene sequence, added to the specific evolutionary processes of this genetic marker, has not allowed the close relationships existing between these organisms to be completely resolved: only 36 variable positions were found in a comparison of 1467 nucleotides corresponding to the sequences of the two rRNA operons of 10 strains of the M. mycoides cluster, including all type strains, increasing to 66 (30 unique positions) when M. putrefaciens KS1 was included. The use of five housekeeping gene sequences has not only reduced the bias generated by using a single gene sequence but has also

---

**Fig. 5.** Evolutionary distance trees based on lepA sequences of 657 bp (a) and based on the second half of these lepA sequences, from nucleotide position 328 (b), and a tree inferred from the first codon positions of the 657 bp lepA sequences (c). For further details refer to Fig. 3. See Table 1 for strain details and Supplementary Table S1 for sequence accession numbers.
Phylogeny and taxonomy of the *M. mycoides* cluster

provided greater resolution by considerably increasing variability. The partial gene sequences used for phylogeny inference provided 106–185 variable positions in a comparison of the 28 strains analysed and 36–68 when limiting the comparison to the 23 strains of the *M. mycoides* cluster (Table 3). Concatenation of the five sequences increased the number of variable positions to 261, of which 203 were informative for phylogenetic analysis of the *M. mycoides* cluster.

This approach has allowed us to construct a reliable tree to assign phylogenetic positions to all members of the *M. mycoides* cluster, showing once more that the existing taxonomy needs to be reviewed.

**Phylogenetic position of Mycoplasma sp. bovine group 7 of Leach.** The greatest incongruences, found in two of the five gene sequences analysed in this study, resulted from the attraction between MmmSC and MBG7. For *gyrB*, this proximity was limited to a subsequence and, in the case of *lepA*, it essentially concerned the least-constrained codon positions. The common structure corresponding to *fusA* and *glpQ* (and *rpoB* when excluding the influential unit MmmLC-1), which places MmmSC in the *M. mycoides* subcluster and MBG7 in the *M. capricolum* subcluster, is not strongly contradicted by analyses of the other two genes and can therefore be retained. Furthermore, the partial attraction observed between MBG7 and MmmSC explains the ambiguity that persists in the present classification and why, unlike the other members of the *M. mycoides* cluster, the group of strains referred to as MBG7 has not yet been assigned a species name. Investigations based on DNA–DNA hybridization (Christiansen & Ernø, 1982), identification with a DNA probe (Taylor et al., 1992) and electrophoretic analysis of isoenzymes and serology (Salih et al., 1983) have suggested the classification of MBG7 as a distinct species. DNA–DNA hybridization studies by Askaa et al. (1978), lipoprotein sequence similarity and immunological cross-reactions (Frey et al., 1998), as well as sequence similarity of the glycerol transport locus *gtsABC* (Djordjevic et al., 2003), have revealed a close relatedness to MmmSC. In contrast, serological cross-reactions instead suggested that MBG7 was closer to the *M. capricolum* subspecies (Cottew et al., 1987; Ernø et al., 1983), which was confirmed by 1D and 2D SDS-PAGE profile comparisons (Costas et al., 1987; Olsson et al., 1990). In addition, MSC4, a small stable RNA, was found to be specific to the species *M. capricolum* and MBG7 (Ushida et al., 2003), and recent phylogenetic studies based on 16S rRNA gene sequences (Pettersson et al., 1996), 16S–23S rRNA intergenic spacer sequences (Harasawa et al., 2000) and a putative membrane protein gene (Thiaucourt et al., 2000) all grouped MBG7 in the *M. capricolum* subcluster. The results presented here provide additional evidence regarding the genetic identification of this group of strains and support its classification as a subspecies of *M. capricolum*.

Interestingly, all strains of MBG7 provided identical sequences for each of the five genes analysed. This seems to indicate a very limited intraspecies variability for this group of strains and, presumably, a recent differentiation within the species *M. capricolum*.

All members of the *M. mycoides* cluster have been isolated from small ruminants (Atalaia et al., 1987; Kusiluka et al., 2001), where multiple infections are frequent (Cottew & Yeats, 1982). These conditions may well promote horizontal gene transfer between the different species. Genomic exchanges were suspected for several strains analysed in this study. This is the case for strain MmmLC-5, for which the first half of the *lepA* sequence was characteristic of *M. capricolum* while the second half corresponded to a typical *M. mycoides* sequence. A lateral transfer of genes or gene fragments may explain the local proximity that exists between MBG7 and MmmSC that has been observed previously and is again shown in two of the five genes analysed in this work.

**Relatedness of *M. mycoides* subsp. *mycoides* LC and *M. mycoides* subsp. *capri*.** The distinction of MmmLC and

**Fig. 6.** Evolutionary distance trees based on *rpoB* sequences of 625 bp constructed from the set of 23 strains belonging to the *M. mycoides* cluster (a) and constructed as above but excluding strain MmmLC-1 (b). For further details refer to Fig. 3. See Table 1 for strain details and Supplementary Table S1 for sequence accession numbers.
Mmc as separate subspecies has been questioned for some time. Numerous investigations have provided evidence to suggest that they may be grouped as a single subspecies within the M. mycoides subcluster. Although early DNA–DNA hybridization studies by Askaa et al. (1978) suggested that Mmc was closer to MmmSC than to MmmLC, analyses of 1D and 2D PAGE protein patterns proved the opposite (Costas et al., 1987; Leach et al., 1989; Olsson et al., 1990; Rodwell, 1982; Salih et al., 1983). Identification using a DNA probe (Taylor et al., 1992) and analysis of LppA sequence similarity and serological cross-reactions (Monnerat et al., 1999) also suggested a very close relatedness between MmmLC and Mmc, and two phylogenetic studies of the M. mycoides cluster concluded that they should be grouped into a single entity (Pettersson et al., 1996; Thiaucourt et al., 2000). However, the International Committee on Systematics of Prokaryotes Subcommittee on the taxonomy of Mollicutes has indicated that additional comparative analysis is required in order to clarify this matter (Johansson & Bradbury, 2007).

Additional evidence has recently been presented by Vilei et al. (2006) based on rpoB sequence comparisons. In the present work, comparisons have been extended to five genes using a similar sample of strains, though strain WK354, included as an Mmc strain in previous studies (Monnerat et al., 1999; Thiaucourt et al., 2000), was characterized in this work as MmmLC by both growth inhibition and dot immunobinding techniques. As in previous studies, the data obtained in this work show that MmmLC is closer to Mmc than to the biotype MmmSC and that MmmLC and Mmc are not sufficiently distant to constitute separate entities. Furthermore, neither of these two organisms represents the aetiological agent of a characteristic disease, which may have been a justification for their separation. For example, this had been the rationale for the establishment of Mccp as a separate subspecies of M. capricolum.

In conclusion, MmmLC and Mmc appear to constitute a single entity, which could justify their combination into the same taxonomic subspecies.

Position of M. putrefaciens and other related species. M. putrefaciens is not included in the classical M. mycoides cluster and may be easily differentiated from the members of this cluster by biochemical and serological methods (Tully et al., 1974) as well as by a specific PCR test (Peyraud et al., 2003). However, on the basis of 16S rRNA gene sequence similarity, it was included in the phylogenetic M. mycoides cluster, within the Spiroplasma group (Weisburg et al., 1989). Also based on 16S rRNA sequence data, the closely related species M. cottewii and M. yeatsii were included in this phylogenetic cluster, branching with M. putrefaciens (Heldtander et al., 1998). Although the latter authors suggested that these three species should actually constitute a separate cluster, their precise taxonomic position remains uncertain. In this study, M. putrefaciens, M. cottewii and M. yeatsii were distant from the members of the classical cluster, with an important number of strictly specific nucleotide variations, confirming that they should not be included in the M. mycoides cluster.

Application to identification and classification of strains

The great relevance of diseases such as CBPP and CCPP and the fact that several members of the M. mycoides cluster produce similar clinical signs in small ruminants with different virulence mean that precise identification is of paramount importance. However, some of these mycoplasmas are difficult to differentiate by conventional techniques (Cottew et al., 1987). Growth inhibition, which is often used for species identification, is frequently hampered both by interspecies cross-reactions and by intraspecies variability. Cross-reactions are frequent between biotypes MmmSC and MmmLC (Cottew & Yeats, 1978; Thigpen et al., 1983) and between MBG7 and Mccp (Cottew et al., 1987; Ernø et al., 1983; Kibe et al., 1985), hindering confirmation of CBPP and CCPP, respectively. A high degree of intraspecies heterogeneity has also been shown using several techniques, particularly within MmmLC/Mmc and also Mcc to a certain extent, whereas MmmSC and Mccp appear rather homogeneous (Abu-Groun et al., 1994; Christiansen & Ernø, 1990; Costas et al., 1987; Kusiluka et al., 2001). Our data correlate with these observations and emphasize once more the limitations encountered in the classification of certain strains and, furthermore, in the definition and delimitation of species and subspecies within the M. mycoides cluster.

Given the increased resolving power of the strategy presented in this work, it was hypothesized that it might allow strains that have proved difficult to affiliate to be classified. This is the case for strain 8756-C13, isolated from a Rocky Mountain goat, for which identification by classical methods is challenging. Cultural parameters indicated that it might be an MmmLC strain, though growth-inhibition tests were negative for both MmmLC and Mmc. It was finally classed as MmmLC on the basis of immunofluorescence results, although electrophoretic patterns were inconclusive. The data from this study show that this is not a typical strain of the M. mycoides cluster, but a nearby relative, located closer to this cluster than to other related species such as M. putrefaciens, M. cottewii and M. yeatsii. Presumably, study of new isolates originating from diverse geographical and host origins will result in the identification of further ‘atypical’ strains related to this cluster and, until whole-genome sequence comparisons become accessible to diagnostic laboratories, protein gene sequence data may prove to be a good addition to 16S rRNA genes for fine strain characterization. However, analyses based on multiple gene sequences are still difficult to use for routine diagnostic purposes. On the other hand, fusA may be an interesting alternative for rapid identification and phylogenetic positioning by PCR and sequencing, for which the set of sequences presented in this work may
serve as a reference: although variability was low among the fusA sequences from *M. mycoides* strains, the tree inferred from this gene was very similar to that based on the concatenated sequences.

**Concluding remarks**

Conserved protein-coding gene sequences may be used to complement and extend 16S rRNA gene sequence data to allow differentiation of closely related groups of organisms, such as the members of the *M. mycoides* cluster. However, a single gene tree can be largely incongruent with the species tree, which represents its true evolutionary history. Therefore, several genes must be analysed simultaneously. It is proposed here to consider the tree inferred from five concatenated gene sequences as a reliable estimation of the species phylogeny, since this ‘mean structure’ can be detected in each individual gene tree, even if it is partially masked by specific gene evolution.

This approach has resulted in the construction of a reliable phylogenetic tree, which will be a universal tool for the classification of strains of the *M. mycoides* cluster, including those difficult to identify by classical methods. Furthermore, it may lead to the development of new techniques for rapid species identification and molecular epidemiology.

**ACKNOWLEDGEMENTS**

We are grateful to Valerie Barbe from Genoscope (Evry, France) for the results of MmmLC strain 95010-C1 whole-genome sequencing and to François Pounararat and Florence Tardy from AFSSA (Lyon, France) for characterization of the strains by dot immunobinding. We also thank Alan Blanchard and Pascal Sirand-Pugnet (INRA-Bordeaux, France) for invaluable discussions and are indebted to Janet Bradbury for critical reading of the manuscript and for her precious suggestions.

**REFERENCES**


