RNA polymerase $\beta$-subunit-based phylogeny of Ehrlichia spp., Anaplasma spp., Neorickettsia spp. and Wolbachia pipientis

Anne-Véronique Taillardat-Bisch, Didier Raoult and Michel Drancourt

Sequence analysis of $rpoB$, the gene encoding the $\beta$-subunit of RNA polymerase, was used in a phylogenetic investigation of nine species from the genera Ehrlichia, Neorickettsia, Wolbachia and Anaplasma. The complete nucleotide sequences obtained for Anaplasma phagocytophilum (HGE agent), Ehrlichia chaffeensis, Neorickettsia sennetsu, Neorickettsia risticii, Anaplasma marginale and Wolbachia pipientis were amongst the longest $rpoB$ sequences in GenBank and ranged from 4074 bp for N. sennetsu to 4311 bp for W. pipientis. Additional partial $rpoB$ sequences were obtained for Ehrlichia canis, Ehrlichia ruminantium and Ehrlichia muris. Identical phylogenetic trees were inferred from multiple sequence alignments of the nucleotide sequences and the derived amino acid sequences using either distance, maximum-likelihood or parsimony methods. This study confirms the phylogeny previously inferred from sequence analyses of the 16S rRNA gene, groESL and gltA and allows the confirmation of four monophyletic clades. The $rpoB$ nucleotide sequences were more variable than the 16S rRNA gene and groESL sequences at the species level.

Bacteria previously classified in the tribe Ehrlichiae in the order Rickettsiales were first described as small, Gram-negative, obligate intracellular micro-organisms that occurred in peripheral blood monocytes or granulocytes, where they multiplied in a modified parasitophorous endosome (Weiss & Moulder, 1984). Taxonomic and phylogenetic studies of these bacteria were initially based on their morphological, antigenic and metabolic characters (Weiss & Moulder, 1984). Using such a previous classification scheme, the family Rickettsiaceae was found to be composed of, amongst others, the tribe Ehrlichieae, containing the genera Ehrlichia, Cowdria and Neorickettsia, and the tribe Wolbachieae, containing the genus Wolbachia. The genus Anaplasma was classified in a different family, the Anaplasmataceae. With the introduction of molecular techniques, new approaches became available to study the phylogenetic and taxonomic relationships amongst bacteria (Roux, 1999). Phylogenetic studies based on comparisons of 16S rRNA gene sequences of ehrlichiae showed that the bacteria belonged to the $\alpha$-subclass of the Proteobacteria and were closely related to the members of the genus Rickettsia (Weisburg et al., 1989). Ehrlichia, Wolbachia pipientis, Neorickettsia, Cowdria and Anaplasma grouped together in the same Ehrlichia cluster (Drancourt & Raoult, 1994), whereas Wolbachia persica grouped with the facultatively intracellular bacterium Francisella tularensis (Forsman et al., 1994). Additional phylogenetic trees made by comparing groESL (Sumner et al., 1997, 2000) and gltA (Inokuma et al., 1997) sequences confirmed the phylogenetic relationships derived from comparisons of 16S rRNA gene sequences. Recently, based on these data, Dumler et al. (2001) proposed to reorganize this clade into four genera, Anaplasmataceae, Ehrlichia, Neorickettsia and Wolbachia, in a single family, the Anaplasmataceae.

It has been emphasized that reliable bacterial phylogeny is best obtained by comparing results obtained from multiple molecular tools, and it has been suggested that any alternative genes used should encode central cellular functions (Olsen & Woese, 1993). We chose to study $rpoB$ as a suitable gene for additional phylogenetic studies of the ehrlichiae. The gene encodes the $\beta$-subunit of RNA polymerase and belongs to an operon comprising $rplL$, $rpoB$ and $rpoC$ (Yura & Ishihama, 1979). These genes encode an enzyme complex that is highly conserved amongst the Bacteria (Ovchinnikov et al., 1981) and the Archaea (Klenk & Zillig, 1994) and their nucleotide sequences have been widely and successfully used to infer phylogenetic relationships amongst other bacteria (Klenk & Zillig, 1994; Mollet et al., 1997; Pühler et al., 1989; Rowland et al., 1993; Renesto et al., 2000a, b).
The objectives of our study were (i) to determine the complete sequences of the rpoB genes of six species currently belonging to the genera *Ehrlichia*, *Neorickettsia*, *Wolbachia* and *Anaplasma* and partial sequences from related species, (ii) to determine the rpoB-based phylogenetic relationships between these species using multiple methods of analysis and (iii) to compare the results with previously established phylogenies for these bacteria.

The strains used in our study are described in Table 1. *Anaplasma phagocytophilum* (HGE agent) was cultured in HL-60 cells while *Ehrlichia chaffeensis*, *Neorickettsia sennetsu*, *Neorickettsia risticii*, *Ehrlichia canis* and *Ehrlichia ruminantium* were co-cultured with DH82 cells. *W. pipiens* was co-cultured with Aa23 cells (mosquito cell line) (O’Neill et al., 1997). Organisms were purified on renografin gradients and DNA was extracted using the QIAamp Tissue kit (Qiagen) according to the manufacturer’s recommendations. DNA extracted from *Anaplasma marginale* and *Ehrlichia muris* was kindly provided by Dr G. Palmer (Washington State University, Pullman, USA) and Dr M. Kawahara (Nagoya City Public Health Research Institute, Japan), respectively.

The *A. phagocytophilum rpoB* gene was amplified by PCR using consensus primer pairs determined after alignment of complete *rpoB* and *rplL* sequences available in GenBank. Primers D1760AGB (5'-GGITTIGAITGCGAGC-3'; positions 1631–1648 relative to the *Escherichia coli rpoB* nucleotide sequence), R1820AGB (5'-GACCTTCGGGTTC(A/G)AIIGGAC-3'; positions 1699–1676) and R4060AGB (5'-GAIITTAACGITAICATTTTG-3'; positions 3884–3862) were designed by reference to an alignment of *rpoB* sequences of *Escherichia coli*, *Coxiella burnetii*, *Salmonella enterica* Typhimurium and *Rickettsia prowazekii*. Primer D420GL (5'-CTIGGAAIGCGGGGC-3'; positions 331–347) was designed by reference to an alignment of *rplL* sequences of *Escherichia coli* and *S. enterica* Typhimurium. DNA extracted from uninfected HL60 cells into sterile distilled water was used for negative controls. Sequencing was carried out using the dRhodamine Terminator cycle-sequencing ready reaction kit with AmpliTaq FS (Perkin Elmer Applied Biosystems) as described by the manufacturer and reaction products were sequenced with an automated DNA sequencer (ABI 310 Genetic Analyzer; Perkin Elmer Applied Biosystems). The sequences were edited and assembled with the auto-assembler program of the ABI 310 Genetic Analyzer. The Universal Genome Walker kit (Clontech) was used to determine the 3’ end of the *A. phagocytophilum rpoB* gene. PCR was performed using an adaptor primer (AP1) supplied by the manufacturer and an *rpoB*-specific primer, WALK2 (5’-CGATGTGATCCGTCGGCCAGGGTGGA-3’), designed specifically for the 3’ end of the ongoing *A. phagocytophilum rpoB* sequence. As we were unable to amplify *rpoB* from other species of the *Ehrlichiae* with the above primers, we designed the additional PCR primers D4z (5’-TATGGGIGACIAATGCA-3'; positions 2108–2127 relative to the *A. phagocytophilum rpoB* sequence) and R7u (5’-GCCCAACATTCCATITCCIC-3'; positions 3944–3924 relative to the *A. phagocytophilum rpoB* sequence) by reference to an alignment of *rpoB* sequences of *R. prowazekii*, *Bartonella henselae*, *Bartonella quintana* and *A. phagocytophilum* (determined in this study). The Genome Walker method was used to determine the 5’ and 3’ ends of the genes of *Ehrlichia chaffeensis*, *N. risticii*, *N. sennetsu*, *A. marginale* and *W. pipiens*. The specific primers used for genome walking and their positions on the *rpoB* gene are available as supplementary material in IJSEM Online (http://ij.sgmjournals.org). With primers D4z and R7u, partial sequences were obtained for three other species, *Ehrlichia canis*, *Ehrlichia ruminantium* and *Ehrlichia muris*. Amplification and sequencing were performed under the conditions described above.

Phylogenetic trees were constructed based on the *rpoB* gene sequences and deduced amino acid sequences after alignment using the CLUSTAL W program (Thompson et al., 1994) supported within the Biscane workstation (Dessen et al., 1990). Phylogenetic trees were inferred using the PHYLIP software package (Felsenstein, 1989). Evolutionary distance matrices, generated by DNADIST and PROTDIST, were constructed by the method of Kimura (1980). The matrices were used to infer dendrograms using the neighbour-joining method (Saitou & Nei, 1987). The data were also examined using parsimony (DNAPARS and PROTPARS in PHYLIP) and maximum-likelihood analysis (DNAML in PHYLIP). Bootstrap values were obtained for a consensus based on 100 randomly generated trees using SEQBOOT and CONSENSE. Tree figures

### Table 1. Strains from which DNA was extracted for PCR amplification of *rpoB* sequences

<table>
<thead>
<tr>
<th>Strain</th>
<th>Natural host</th>
<th>Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Anaplasma phagocytophilum</em> WebsterT</td>
<td>Sheep, horses, dogs, deer</td>
<td>Ixodes ticks</td>
</tr>
<tr>
<td><em>Anaplasma marginale</em> Florida</td>
<td>Cattle, sheep, goats</td>
<td>Dermacentor ticks</td>
</tr>
<tr>
<td><em>Ehrlichia chaffeensis</em> ArkansasT</td>
<td>Humans</td>
<td>Amblyomma americanum</td>
</tr>
<tr>
<td><em>Neorickettsia sennetsu</em> MiyayamaT</td>
<td>Humans</td>
<td>Fish helminths</td>
</tr>
<tr>
<td><em>Neorickettsia risticii</em> IllinoisT</td>
<td>Horse</td>
<td>Helminths</td>
</tr>
<tr>
<td><em>Wolbachia pipiens</em> O’Neill</td>
<td>Insects</td>
<td>Culex pipiens</td>
</tr>
<tr>
<td><em>Ehrlichia canis</em> OklahomaT</td>
<td>Dogs</td>
<td><em>Rhipicephalus sanguineus</em></td>
</tr>
<tr>
<td><em>Ehrlichia muris</em> AsukeT</td>
<td>Mice</td>
<td>Not determined</td>
</tr>
<tr>
<td><em>Ehrlichia ruminantium</em> Kum</td>
<td>Goats, sheep, cattle</td>
<td>Amblyomma</td>
</tr>
</tbody>
</table>
were generated using TreeView version 1.61 (Page, 1996) for parsimony and maximum-likelihood analysis and using MEGA (http://evolgen.iob.metro-u.ac.jp/mega5/) for the neighbour-joining method (Sudhir et al., 1993). Sequence similarity comparisons were made with the GAP program of the GCG package (Genetics Computer Group), again supported within the Bisance workstation.

In this study, we developed rpoB consensus primers by comparing the sequences of the rpoB genes of various bacteria in GenBank. We used these primers in PCRs to amplify the rpoB gene from ehrlichieae. The primer pairs D420GL/R1820AGB and D1760AGB/R4060AGB enabled the amplification of 3890 bp of the rpoB of A. phagocytophilum in two fragments, F1 (1230 bp) and F2 (2650 bp). These extended from rplL (upstream of rpoB) beyond a part of rpoB. The primer pair D4x and R7u enabled the amplification of fragments of rpoB of 2000 bp from A. marginale, Ehrlichia chaffeensis, N. sennetsu, N. risticii and W. pipientis. The 5' and 3' ends were determined using the Universal Genome Walker kit, resulting in rpoB gene sequences of 4074 bp in N. sennetsu, 4092 bp in N. risticii, 4137 bp in Ehrlichia chaffeensis, 4146 bp in A. marginale, 4185 bp in A. phagocytophilum and 4311 bp in W. pipientis. The G+C content of the rpoB gene varied from 33-1 mol% for Ehrlichia chaffeensis to 48-4 mol% for A. marginale and deduced amino acid sequences varied from 1357 to 1436 aa. The percentage similarity varied from 57-1 % (A. marginale vs N. sennetsu) to 98-1 % (N. sennetsu vs N. risticii) for the nucleotide sequences and from 60-2 % (Ehrlichia chaffeensis vs N. sennetsu) to 95-2 % (N. sennetsu vs N. risticii) for the deduced amino acid sequences. The percentages of similarity of the rpoB sequences between the two species N. sennetsu and N. risticii and other ehrlichial species were lower (58-5-65-4%) than those between other Ehrlichia species and B. henselae (44-8-66-3%). Using the primer pair D4x and R7u, we also established partial rpoB nucleotide and deduced amino acid sequences for Ehrlichia canis, Ehrlichia ruminantium and Ehrlichia muris. Full descriptions of the rpoB genes determined in this study are available as supplementary material in IJSEM Online (http://ijsem.sgmjournals.org).

The topologies observed in the rpoB trees obtained using the three methods of analysis were identical, and showed the ehrlichieae to be closely related to the genus Rickettsia in the α-subclass of the Proteobacteria. The bootstrap values showed greatest support using the neighbour-joining method and ranged from 99 to 100 %, except at the node where Ehrlichia chaffeensis diverged, which had a bootstrap value of 70 %. The phylogetic tree for the ehrlichieae had four branches. The first branch comprised the genus Anaplasma, although there was relatively low similarity (74-5%) between A. marginale and A. phagocytophilum. The second branch contained Ehrlichia chaffeensis, the third W. pipiens and the fourth branch contained N. sennetsu and N. risticii, which had very high similarity (98-1%). When more ehrlichieae, namely Ehrlichia canis, Ehrlichia muris and Ehrlichia ruminantium, were included in a partial rpoB-based tree, the same topologies were found and had high bootstrap values (Fig. 1).

Our data confirm those of Dumler et al. (2001) and are in agreement with the findings of groEL- and gltA-based phylogenetic studies (Sumner et al., 1997, 2000; Inokuma et al., 2001) and those postulated previously (Drancourt & Raoult, 1994). The number of clades depends on the level of rpoB sequence divergence used to define them (Fig. 1): three clades would be delineated using a 5% divergence level, four clades would be delineated at a divergence level between 4-5 and 5% and five clades would be delineated at a divergence level of >2-3-4-5%. Biological and phenotypic properties may help to resolve the question of the appropriate level of divergence.

Phylogenetic relationships often cannot be accurately established using sequence information from only a single gene. Sequence analysis of the rpoB gene, which has more sequence variation than the 16S rRNA gene, is an additional tool on which to base the phylogenetic relationships of the ehrlichieae.

Acknowledgements
The authors acknowledge Dr Patrick Kelly for reviewing the manuscript.

References


