Investigation of the phylogenetic relationships within the genus *Bartonella* based on comparative sequence analysis of the *rnpB* gene, 16S rDNA and 23S rDNA

Christian Pitulle, Christoph Strehse, James W. Brown and Edward B. Breitschwerdt

Author for correspondence: Christian Pitulle. Tel: +1 919 513 6433. Fax: +1 919 513 6336. e-mail: Christian.Pitulle@NCSU.edu

A variety of genes and analytical methods have been applied to the study of phylogenetic relationships within the genus *Bartonella*, but so far the results have been inconsistent. While previous studies analysed single protein-encoding genes, we have analysed an alignment containing the sequences of three important phylogenetic markers, RNase P RNA, 16S rRNA and 23S rRNA, merged by catenation, to determine the phylogenetic relationships within the genus *Bartonella*. The dataset described here comprises 13 different *Bartonella* strains, including the seven strains that are known to be human pathogens. A variety of algorithms have been used to construct phylogenetic trees based on the combined alignment and, for comparison purposes, each individual gene. Trees generated from the catenated alignment are more consistent (independent of algorithm) and robust (higher bootstrap support). It is suggested that a phylogenetic analysis incorporating the RNase P RNA, 16S rRNA and 23S rRNA be used to study the phylogenetic relationships within the genus *Bartonella*.

**Keywords:** 16S rDNA, 23S rDNA, RNase P RNA, phylogeny, genus *Bartonella*

**INTRODUCTION**

Members of the genus *Bartonella* have been associated with a broadening spectrum of clinical syndromes in both human and veterinary medicine (Anderson & Neuman, 1997; Breitschwerdt & Kordick, 2000), making the determination of the phylogenetic relationships between these strains and the identification of species-specific markers for clinical diagnostic testing imperative. The most contemporary tools used to identify micro-organisms and to study their natural relationships are the molecular approaches that utilize the sequence variability of certain genes, especially those that encode highly structured RNA molecules. The use of well-defined secondary structure in the identification of homologous residues (i.e. the alignment process) is one important advantage of RNA sequences over genes that encode proteins. In particular, comparative analysis of the 16S rDNA (Woese, 1987; Ludwig & Schleifer, 1999) is considered to be the ‘gold standard’ for phylogenetic analysis. However, among very closely related groups of organisms, such as the species of the genus *Bartonella*, phylogenetic differentiation at the species level and below (Fox et al., 1992; Suksawat et al., 2001; Houpikian & Raoult, 2001) is not always possible because of the high degree of 16S rDNA sequence conservation. This has led to difficulties in the identification of medically important *Bartonella* strains when using 16S rDNA-based molecular diagnostics. Alternative genes have been used successfully to increase the sensitivity and specificity of diagnostic methods for the detection of these *Bartonella* species. Useful alternative genes include a 17 kDa antigen gene (Sweger et al., 2000), flaA (Sander et al., 2000), gltA (Birtles & Raoult, 1996; Roux et al., 2000; Houpikian et al., 2001), htrA (Anderson & Neuman, 1997), ftsZ (Ehrenborg et al., 2000) and groEL (Marston et al., 1999), as well as the 16S–23S rDNA internally transcribed spacer sequences (Jensen et al., 2000; Birtles et al., 2000). However, when used to infer the phylogenetic relationships within the genus

The GenBank accession numbers for the 16S rDNA, 23S rDNA and *rnpB* sequences described in this work are listed in Table 1.
Bartonella, these sequences suggested conflicting relationships. Variations in the datasets, in terms of the strains and number of species used in the studies and analytical methods, have made it difficult to compare and evaluate the proposed phylogenetic relationships. Additionally, some earlier studies did not provide detailed information regarding the data analysis, including sequence accession numbers and alignments. Furthermore, all previous studies were based on the analysis of a single gene. The limitations of single gene analysis of Bartonella species have been recognized, especially for the use of 16S rRNA, groEL and gltA (Houpikian & Raoult, 2001).

To circumvent the limitations of single gene analysis, we performed a phylogenetic study using three phylogenetic markers, 16S rRNA, 23S rRNA and the RNA subunit of the endoribonuclease RNase P, to determine the phylogenetic relationship within the genus Bartonella. rRNAs are used universally for phylogenetic studies. Historically, 16S rRNA was chosen over 23S rRNA for phylogenetic studies because its smaller size made it more tractable for sequence analysis. As a result, several thousand 16S rRNA sequences are available, whereas the number of available 23S rDNA sequences remains rather small. However, the 23S rRNA contains more than twice the phylogenetic information contained in 16S rRNA (Woese, 1987; Ludwig & Schleifer, 1999). Because of vast improvements in molecular techniques, especially amplification methods and automated DNA sequencing, 23S rRNA sequences have become more easily accessible for phylogenetic analysis. RNase P RNA demonstrates a greater extent of sequence divergence per nucleotide position than the larger rRNAs. Therefore, it can be phylogenetically more informative in cases of very closely related organisms (Haas & Brown, 1998). These sequences can be used to study the phylogenetic relationships of organisms from all three kingdoms of life (Pitulle et al., 1998; Woese et al., 1990) or to differentiate very closely related medically important bacteria (Herrmann et al., 2000).

Of the 13 strains investigated in this study, Bartonella bacilliformis, Bartonella claridgeiae, Bartonella quintana, Bartonella henselae, Bartonella elizabethae, Bartonella grahamii, Bartonella vinsonii subsp. arupensis and Bartonella vinsonii subsp. berkhoffii are known to be human pathogens (Houpikian & Raoul, 2001).

METHODS

DNA isolation and PCR amplification. Genomic DNA was isolated from bacterial cells (approx. 100 µg) by following the protocol of the Ultraclean Soil DNA isolation kit (MoBio Laboratories). All subsequent PCRs were performed on a Progene thermocycler (Techne). The 16S rDNA amplification for B. bacilliformis KC584 was performed as described previously (Pitulle et al., 1999).

The 23S rDNA was amplified by PCR using 20 pmol each of primers 125F (5'-GATTTCGAAATTGGGMAACC-3'; M = A or C) and 2650R (5'-CCATCCCGGCTCCTCCTCG-TACT-3'). Each 100 µl PCR contained dGTP, dATP, TTP and dCTP at 0.2 mM each, 2 mM MgCl₂ (Promega), 5 U Taq polymerase (Promega), 1 × buffer (Promega), 5 µl of the DNA extraction (approx. 200 ng DNA) and double-distilled water. All reactions started with an initial denaturation step of 2 min at 94 °C, followed by 30 amplification cycles (1 min at 92 °C, 1 min at 50 °C and 1 min at 72 °C) and a final extension step of 10 min at 72 °C.

The genes for the RNase P RNA were amplified by PCR using 20 pmol each of primers 59F BamHI (5'-CGGGATCCGAGGAAAGTGCCTG-3'; I = inosine) and 347R EcoRI (5'-CCGAAATCTAAGCGGRTTCTGTG-3'; R = A or G), 5 U Taq polymerase (Promega), 200 ng genomic DNA, 1.5 mM MgCl₂ and 0.2 mM each dNTP in 1 × reaction buffer (Promega). All PCRs started with an initial denaturation step for 2 min at 94 °C, followed by 35 amplification cycles (1 min at 92 °C, 30 s at 53 °C and 30 s at 72 °C) and a final extension step for 5 min at 72 °C.

PCR products were separated on a 1% (w/v) agarose gel containing 0.5 µg ethidium bromide ml⁻¹ and were visualized by UV illumination and compared with the DNA size standards hyperladders I and IV (Denville Scientific). PCR fragments were purified according to the protocol of the QIAquick PCR purification kit (Qiagen).

Cloning of PCR products and sequencing of recombinant DNA. The purified PCR products were ligated into the pgEM-T vector system (Promega) followed by transformation of Escherichia coli JM109 high-efficiency competent cells according to the protocol outlined by Promega. Recombinants were selected by the blue/white colour of colonies. Plasmid DNA from several clones was isolated by using the QIAprep plasmid kit (Qiagen). Both strands of recombinant plasmid DNA (200–400 fmol) were sequenced by using the SequiTherm EXCELII Long Read DNA sequencing kit LC (Epicentre Technologies) using 1 pmol each of fluorescently labelled primers T7-800 (5'-GTAATA-CCGACTCATAAGGG-3') and SP6-700 (5'-ATTAGTGACACTATAG-3') and primers 515F-800 (5'-TGACACTATAG-3') for sequencing of 16S rDNA, the primers 515F-800 (5'-TGTGCCAGCMGCCGCGTAA-3') and 1391R-700 (5'-GACGGGCGGTGCGWTRCA-3'; W = A or T) were used in addition to primers T7-800 and SP6-700. The 23S rDNAs were sequenced with primers 125F-700, 2650R-800, 1608F-700 (5'-CYACCTGTGWCAGGTGT-3'; Y = C or T) and 2235R-800 (5'-GAGAGCGCCGGC-CCCCCGTCAAAT-3') in addition to T7-800 and SP6-700. After an initial denaturation step of 2 min at 92 °C, 30 sequencing cycles (30 s at 92 °C, 15 s at 50 °C and 15 s at 70 °C) were executed. The sequencing reactions were analysed by PAGE [5-5% (v/v) for mpR; 3.75% (v/v) for rDNA] on a Li-Cor 4200 automated DNA sequencer. Sequences have been deposited in GenBank (Benson et al., 2000); the accession numbers are listed in Table 1.

Sequence alignment and phylogenetic tree analysis. The previously available Bartonella and Agrobacterium tumefaciens ribosomal and RNase P RNA sequences (see Table 1) were respectively downloaded as aligned sets from the Ribosomal Database Project (Maidak et al., 2001) and the RNase P database (Brown, 1999). The additional sequences were added to the alignment, inserting alignment gaps on the basis of differences in secondary structure. Sequences at either end that were not available for all sequences were discarded, leaving 313 positions in the RNase P RNA alignment. 1388 positions in the 16S rRNA alignment and 2424 positions in the 23S RNA alignment. Considering only the Bartonella data, there were 36, 88 and 216 phylogenetically informative positions (i.e. positions that varied in sequence) in the RNase P, 16S rRNA and 23S
Table 1. Strains used for the phylogenetic analysis within the genus Bartonella

The GenBank accession numbers for sequences of the 16S rDNA, 23S rDNA and rnpB are listed. Sequence data that were determined in this study were deposited in GenBank under the accession numbers indicated by asterisks.

<table>
<thead>
<tr>
<th>Strain</th>
<th>16S rDNA</th>
<th>23S rDNA</th>
<th>rnpB</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. vinsonii subsp. arupensis ATCC 700727^</td>
<td>AF214558</td>
<td>AF410937*</td>
<td>AF441295*</td>
</tr>
<tr>
<td>B. claridgeiae NCSU 94-F40^ (= ATCC 700095^)</td>
<td>U64691</td>
<td>AF410938*</td>
<td>AY033649*</td>
</tr>
<tr>
<td>B. doshiae R18^ (= ATCC 700133^)</td>
<td>Z31351</td>
<td>AF410939*</td>
<td>AF441294*</td>
</tr>
<tr>
<td>B. elizabethae ATCC 49927^</td>
<td>L01260</td>
<td>AF410940*</td>
<td>AY033770*</td>
</tr>
<tr>
<td>B. vinsonii subsp. berkhoorfi 93CO-1^ (= ATCC 51672^)</td>
<td>L35052</td>
<td>AF410941*</td>
<td>AY033783*</td>
</tr>
<tr>
<td>B. grahamii V2^ (= NCTC 12860^ = ATCC 700132^)</td>
<td>Z31349</td>
<td>AF410942*</td>
<td>AY0441293*</td>
</tr>
<tr>
<td>B. henselae Houston-1^ (= ATCC 49882^)</td>
<td>M73229</td>
<td>AF410943*</td>
<td>AY033897*</td>
</tr>
<tr>
<td>Bartonella strain N40</td>
<td>AF204274</td>
<td>AF410944*</td>
<td>AY0441292*</td>
</tr>
<tr>
<td>Bartonella strain deer 159/660/1</td>
<td>AF373845*</td>
<td>AF410945*</td>
<td>AF376051*</td>
</tr>
<tr>
<td>B. quintana Fuller^ (= ATCC VR-358^)</td>
<td>M11927</td>
<td>AF410946*</td>
<td>AY033948*</td>
</tr>
<tr>
<td>‘B. weissii’ 99-BO1</td>
<td>AF291746</td>
<td>AF410947*</td>
<td>AF376050*</td>
</tr>
<tr>
<td>B. vinsonii subsp. vinsonii Baker^ (= ATCC VR-152^)</td>
<td>Z31352</td>
<td>AF411589*</td>
<td>AY033502*</td>
</tr>
<tr>
<td>B. bacilliformis KC584 (Minnick et al., 1995)</td>
<td>AF442955*</td>
<td>L39095</td>
<td>AF440224*</td>
</tr>
</tbody>
</table>

rRNA alignments, respectively, or 41, 106 and 229 for those algorithms that include comparison of bases with gaps. Although the number of informative positions in the RNase P sequence alignment is relatively low, there is more variation within these positions than in the informative rRNA sequences, i.e. the information content per position is higher. The separate sequence alignments were used independently or as a merged set in which the sequence data from each strain were concatenated. These alignments were used to build phylogenetic trees, using PHYLIP version 3.573c (Felsenstein, 1989) programs Dnaspars (parsimony analysis), DNAML (maximum likelihood), DNAPENNY (exhaustive parsimony searches), NEIGHBOR (neighbour-joining) and FITCH (least-squares distance matrix). In all cases, except DNAPENNY (because of the prohibitive computational demand), a consensus tree from 100 bootstrapped datasets (generated using BOOT) was generated along with the single best tree for each alignment. Global rearrangements were allowed in the construction of all trees in which this is a possibility. All trees were explicitly treated as rooted, with the A. tumefaciens sequences as the outgroup. In bootstrap analyses, sequences were added in random order. These alignments and the resulting output files can be accessed online at http://www.mbio.ncsu.edu/Bartonella/.

RESULTS AND DISCUSSION

The phylogenetic analysis of certain DNA sequences is a powerful tool for the determination of the natural relationships between micro-organisms. Although the quality of phylogenetic trees can be estimated by bootstrap analysis, these values alone do not entirely reflect the correctness of the trees. The analysis also depends on the type of gene used, its molecular clock tempo and mode and, most importantly, a meaningful sequence alignment. In the case of RNA sequences, alignments are based on homology identified by comparative analysis (Gutell, 1993; James et al., 1989) rather than simply maximizing sequence similarity, as is customary in protein alignments.

We have isolated and determined the nucleotide sequences of the genes for 13 RNase P RNAs, 12 23S rRNAs and two 16S rRNAs from 13 Bartonella strains (Table 1). The PCR products for the RNase P RNAs represent about 90% of the complete RNA molecules, all of which resemble E. coli-like RNase P RNAs (A-type) with minimal differences in size, ranging from 345 to 349 nt. The RNase P RNAs, at $\geq 94\%$ sequence identity, are the least conserved of the sequences, and the differences are dispersed in the RNA sequence. Therefore, the use of Bartonella RNase P RNA for diagnostic testing based on size-selective or species-specific PCR is unlikely to be readily useful. We did not find a single restriction endonuclease or any combination of any two of these enzymes that can digest all 13 partial rnpB genes to generate unique restriction-fragment patterns. However, because of its small size, sequencing of the entire amplified rnpB gene on both strands is relatively easy and is a very accurate method for clear discrimination of the 13 Bartonella strains investigated in this study.

Of the two rRNAs, the 23S rRNAs show sequence identities of $\geq 96\%$, whereas the 16S rRNAs are $\geq 98\%$ identical. We determined 1406 and 1407 nt, respectively, of the 16S rRNAs derived from Bartonella sp. deer isolate 159/160/1 and B. bacilliformis KC584. The sequence and size differences (2404–2451 nt) in the 23S rDNA are more significant for diagnostic PCR testing than those seen in the 16S rDNAs or the RNase P RNAs. The use of 23S rDNA for diagnostic purposes will be presented in a forthcoming publication.

We combined our sequences with existing 16S rDNA and 23S rDNA data extracted from GenBank (Benson et al., 2000) to create a single alignment set based on the secondary structures of these molecules. Only those regions defined as homologous on the basis of
A. tumefaciens serves as the outgroup in this tree. The tree shown was generated using FITCH (see text), which best matches the consensus bootstrap trees. The horizontal axis is the estimated evolutionary distance. The four numbers shown at each node are the numbers of times that node appears among 100 bootstrapped trees: FITCH, NEIGHBOR, DNA Pars and DNAML from left to right. The final + or – indicates whether or not this node appeared in the single most parsimonious tree identified using DNA PENNY. Bar, estimated evolutionary distance of 0.01 substitutions per position.

Fig. 1. Phylogenetic tree of *Bartonella* species based on the combined RNase P RNA, 16S and 23S rRNA sequence alignment. A. tumefaciens serves as the outgroup in this tree.
Bartonella birtlesii (Bermond et al., 2000) is an indistinct one. The relationship between these organisms is not well supported by bootstrap analysis. If these organisms are specifically related, the relationship is not consistently supported by bootstrap values. Although most trees place both organisms into one subgroup, this relationship is not as clear; neither Bartonella strains N40, IBS 325\(^T\) and IBS 358 are three bacterial strains that were isolated from Apodemus spp. Their partial 16S rDNA sequences as well as their partial gltA gene sequences were found to be identical. On the basis of additional biochemical and molecular tests, all three strains were considered to represent the same Bartonella species, Bartonella birtlesii, although IBS 325\(^T\) became the type strain (Bermond et al., 2000).

B. quintana and B. henselae seem to represent independent branches within the tree, although they appear together in a subset of trees. The relative branching order of these sequences is not well supported by bootstrap values. Determining the relationship between these organisms and the remaining Bartonella species will require sequence data from additional isolates of this genus.

The use of 16S rDNA data analysis remains the generally accepted gold standard when studying phylogenetic relationships. The high degree of 16S rDNA sequence identity within the genus Bartonella, however, has been recognized as a problem because branching arrangements of the trees often elicit very little statistical support. It is doubtful whether phylogenetic relationships should be based solely on 16S rRNA in cases where sequence identities are \(> 99\%\) (Drancourt et al., 2000). Although several protein-encoding genes have been suggested as models for the study of phylogenetic relationships within the genus Bartonella, the citrate synthase gene (gltA) has been most widely used as an alternative to 16S rDNA analysis. However, the use of gltA consistently resulted in low bootstrap support for the resulting trees. Although gltA has been used successfully to differentiate Bartonella species for diagnostic purposes, the phylogenetic interpretation of these diagnostically important changes in the DNA sequences remains unclear.

Although single gene analysis has been used extensively to study the phylogenetic relationships within the genus Bartonella, the present study clearly demonstrates that predictions concerning the natural relationships based on only one gene are not always possible. Results are even more confusing when two or more genes are analysed independently and relationships are concluded from comparisons of the separate trees (Houpikian & Raoult, 2001).

In contrast, the combined use of the phylogenetic markers RNase P RNA, 16S rRNA and 23S rRNA can contribute significantly to clarifying the phylogenetic relationships among Bartonella species. The alternative phylogenetic algorithms used resulted in more consistent tree topologies, and the statistical support for the suggested relationships is generally better than for any single gene, as indicated by higher bootstrap values. One might expect the molecules with the lowest or highest sequence similarities to skew the results. However, there was no single gene in our study that consistently dominated the data analysis. Also, because of its size, the combined dataset is not as susceptible to sequencing errors and allelic heterogeneity as data derived from a single gene.

In conclusion, we suggest that a combined phylogenetic analysis incorporating the RNase P RNA, 16S rRNA and 23S rRNA be used to study the phylogenetic relationships within the genus Bartonella. Because of their phylogenetic value, all three genes could be used when describing a novel Bartonella species or when novel Bartonella isolates are reported.

**ACKNOWLEDGEMENTS**

The work presented was supported by a grant to C.P. from the State of North Carolina, and by NIH grant GM5(2894) to J.W.B. We would like to thank R. Birtles for the gift of Bartonella strain N40, B. grahamii NCTC 12860\(^T\), B. doshiae R18\(^T\) and B. vinsonii subsp. arupensis. We also thank M. Minnick and B. Chomel for B. bacilliformis strain KC584 and the Bartonella sp. deer isolate 159/660/1, respectively.

**REFERENCES**


