Evaluation of intraspecies genetic variation within the 60 kDa heat-shock protein gene (groEL) of Bartonella species

Eric L. Marston,¹ John W. Sumner² and Russell L. Regnery²

Author for correspondence: Eric L. Marston. Tel: +1 404 639 3929. Fax: +1 404 639 3970. e-mail: efm4@cdc.gov

A phylogenetic investigation was done on the members of the genus Bartonella, based on the DNA sequence analysis of the groEL gene, which encodes the 60 kDa heat-shock protein GroEL. Nucleotide sequence data were determined for a near full-length fragment (1368 bp) of the groEL gene of the established Bartonella species and used to infer intraspecies phylogenetic relationships. Phylogenetic trees were inferred from multiple sequence alignments by using both distance and parsimony methods, which demonstrated an architecture composed of six well-supported lineages. The results are consistent with relationships deduced from recent sequence analysis studies based upon citrate synthase (gltA) and previously observed genotypic and phenotypic characteristics; however, they showed greater statistical support at the intragenus level. This suggests that groEL may be a more robust tool for phylogenetic analysis of Bartonella lineages.

Keywords: Bartonella, groEL, 16S rRNA, maximum-likelihood, parsimony analysis

INTRODUCTION

Members of the family Bartonellaceae are characterized as Gram-negative, oxidase-negative, fastidious, aerobic, pleomorphic bacteria belonging to the α-2 subclass of the Proteobacteria. The genus Bartonella currently consists of 11 recognized species (Birtles et al., 1995; Brenner et al., 1993; Daly et al., 1993). Four of the 11 species have been reported to cause human disease: Bartonella bacilliformis [aetiologic agent of bartonellosis or Carrion’s disease (Cáceres-Ríos et al., 1995)], Bartonella elizabethae (agent isolated from the blood of an immunocompromised patient suffering from endocarditis; Brenner et al., 1993), Bartonella henselae (aetiologic agent of cat scratch disease; Dolan et al., 1993) and Bartonella quintana (aetiologic agent of trench fever; Vinson, 1966). The remaining six species were isolated from the blood of vertebrate animals and are not known to cause recognizable human disease: Bartonella claridgeiae (Kordick et al., 1997), Bartonella vinsonii subsp. berkoffii (Breitschwerdt et al., 1995), Bartonella talpae (Birtles & Raoult, 1996), Bartonella peromysci (Birtles & Raoult, 1995), Bartonella vinsonii (Baker, 1946), Bartonella grahamii (Birtles & Raoult, 1996), Bartonella taylorii (Birtles & Raoult, 1996) and Bartonella doshiae (Birtles & Raoult, 1996). The latter three organisms were formerly part of the genus Grahameella (Birtles & Raoult, 1996). Cultures of Bartonella talpae and Bartonella peromysci are no longer available (Birtles & Raoult, 1996).

The family Bartonellaceae was recently reorganized using both genotypic and phenotypic characteristics. The genetic characterization has included 16S rRNA (16S rRNA) and citrate synthase (gltA) phylogeny (Birtles & Raoult, 1996), DNA–DNA hybridization and G+C content (Birtles et al., 1995; Brenner et al., 1993; Daly et al., 1993; Lawson et al., 1996; Myers & Wiseman, 1980; Myers et al., 1979; Weiss et al., 1978). The phenotypic characterization involved fatty acid analysis (Westfall et al., 1984), nutrient and growth requirements (Weiss et al., 1982), host–vector relationships, morbidity or mortality of host, host immune response, growth characteristics (colony morphology), presence or absence of flagella, antigenic cross-reactivity and substrate utilization based upon commercial enzymic profiling systems (Maurin et al., 1997).

Comparison of 16S rRNA sequences from a large number of genera has been used in establishing three
primary lines of bacterial evolutionary descent (Woese et al., 1990). Because of the limited variation in the 16S rRNA gene between members of the same genus, the usefulness of 16S rRNA for inferring intragenus and interspecies phylogeny is sometimes limited (Fox et al., 1980; Olsen & Woese, 1993). To circumvent this limitation, the phylogenetic relationships among the members of Bartonella (Birtles & Raoult, 1996) and Rickettsia (Roux et al., 1997) have been analysed by comparing 16S rRNA and gltA gene sequences. Unlike the 16S rRNA generated trees, the topologies for the respective gltA-inferred trees were supported by phenotypic and genotypic characteristics of the respective species. They, however, lacked the statistical support necessary to clearly establish all of the phylogenetic relationships.

GroEL is one component of the highly conserved heat-shock chaperonin system of proteins in bacteria (Checa & Viale, 1997; Segal & Ron, 1995). It is ubiquitously distributed among the eubacteria, archaea and eukaryotic organelles, and its gene has been used for making inferences of eubacterial phylogenies (Masui et al., 1997; Viale et al., 1994).

The objectives of this study were (1) to obtain a major portion of the groEL gene of the currently recognized Bartonella species, (2) to determine the utility of groEL for inferring phylogenetic relationships between these species using multiple methods, and (3) to compare the results with previously published Bartonella phylogenies.

**METHODS**

**Strains and culture conditions.** Cultures were directly plated on commercially available rabbit blood/heart infusion agar (Becton-Dickinson Microbiology Systems), incubated at 32 °C in a humidified CO2-enriched environment (Regnery et al., 1992a) and kept up to 28 d. The sources of the Bartonella type strains used in this study can be found in Table 1.

**PCR.** Genomic DNA was extracted from bacterial cultures by using methods previously described (Norman et al., 1995) and stored at 4 °C until PCR analysis.

Two primers were selected from the groEL gene of Bartonella bacilliformis (GenBank accession no. Z15160): BbHS223.p (5'-CCTGAAGTGGCCCTCAAACCC 3') and BbHS1630.n (5'-GATCCATTCCGCCCATTC 3').

Primers were selected with the aid of the UNIX software package PRIME from the Genetics Computer Group (GCG, 1996). All primer positions are numbered relative to the groEL gene of Bartonella bacilliformis (GenBank accession no. Z15160). Primers were synthesized by the Biotechnology Core Facility, Centers for Disease Control and Prevention.

The PCR amplification was performed with 10 µl extracted DNA in a 100 µl reaction mixture containing 50 mM KCl, 10 mM Tris/ HCl (pH 8.3), 1.5 mM MgCl₂, 0.001% gelatin, 200 µM (each) dNTP (dATP, dCTP, dGTP and dTTP), 1-5 µM (each) primer, BbHS223.p and BbHS1630.n and 0.2 U thermostable AmpliTaq DNA polymerase (Perkin-Elmer Cetus). The thermocycler conditions used were as follows: 95 °C for 2 min and amplified for 40 cycles at 95 °C for 1 min, 56 °C for 1 min and 72 °C for 1 min and subsequently at 72 °C for 5 min. All PCR reactions were performed using a PTC200 DNA-Engine (MJ Research).

Fifteen microlitres of each amplification reaction was analysed by electrophoresis on a 1-5% agarose gel in 1 x Tris-borate-EDTA (TBE) buffer. Gels were stained with ethidium bromide and visualized by UV fluorescence (Maniatis et al., 1989). PCR products were purified from the post-PCR mix using Wizard PCR Prep (Promega) according to manufacturer's specifications.

**Cloning.** All cloning was performed by using SureShot (Invitrogen) Escherichia coli (INVaF' F' endA1 recA1 hsdR17 (r, mC) supE44 thi-1 gyrA96 relA1 phi80lacZAM15 Δ(lacZYA-argF)U169 1-). PCR-amplified fragments were ligated into pCR2.1 vector, transformed into INVaF'-competent cells with the TA Cloning System (Invitrogen) and selected on Luria-Bertani agar containing ampicillin (100 µg ml⁻¹), according to the manufacturer's protocol. Colonies were screened by insert size with a rapid plasmid miniscreen procedure (Hoekstra, 1988). Plasmid DNA was prepared from 5 ml overnight cultures by using a Qiamp-20 Plasmid Kit (Qiagen). Prior to dideoxy sequencing, insert size was confirmed by excising the insert from the plasmid using EcoRI (Promega) restriction endonuclease and resolving the fragments on a 1-5% TBE agarose gel (Maniatis et al., 1989). Size was established by using a 123 bp ladder molecular mass marker (Sigma).

**Direct dideoxy sequencing.** PCR products were sequenced in both directions with the dideoxy-sequencing Prism Ready Reaction Dye-terminator Kit (Applied Biosystems) using a Cetus 9600 thermocycler (Perkin-Elmer Cetus). The sequencing reactions were resolved with a 4-25% polyacrylamide gel at 51 °C with constant voltage (1500 V) using an ABI Prism model 377 autosequencer (Applied Biosystems). Primary sequencing was carried out with T7 and M13 primers. Sequencing beyond the plasmid-based primers was carried out with primers designed from the novel sequences.

**Sequencing data analysis.** Data extracted from the sequencing gels were analysed first by the Macintosh software ABI-Autoanalyser Sequence Analysis package (Applied Biosystems). The sequence data were then analysed and aligned by using the UNIX software package (Staden, 1982). Contiguous sequences from completed sequencing projects generated in Staden were compared with current nucleic acid entries in GenBank (release 101.0) by using BLAST and FASTA (GCG, 1996).

**Construction of phylogenetic trees.** Levels of similarity between groEL and 16S rRNA sequences were calculated with GCG (1996) in the program OLDDISTANCES by using the uncorrected implementation.

The primary 16S rRNA gene sequences were obtained from GenBank (release 101.0). The 16S rRNA gene sequences of recognized Bartonella species were aligned with each other and with 16S rRNA gene sequences of selected members of Eubacteria by using the multisequence alignment algorithm utilized by PILEUP (GCG, 1996). 16S rRNA sequences were aligned on the basis of conserved regions and secondary elements, as previously described (Birtles & Raoult, 1996) (16 species and 1327 characters). The groEL sequences
groEL amino acid sequences were generated by using methods [the least-squares method of Fitch-Margoliash]. Phylogenetic trees were generated from sequence alignments by using computer algorithms implemented in the PHYLIP and PAUP computer software packages. Trees inferred from the groEL sequences were calculated using PROTDIST in PHYLIP. Pairwise distances between GroEL proteins were calculated using PROTDIST in PHYLIP. Pairwise distances between respective sequences for both 16s rRNA and groEL were calculated by using the two-parameter method of Kimura (1980) as provided by DNADIST in PHYLIP and used for the Fitch–Margoliash and neighbour-joining methods.

*groEL.* GenBank strain designation (16s rRNA GenBank strain designation).

† Unpublished GenBank data.

obtained in this study were aligned with each other, with the sequence of Bartonella bacilliformis, and with the sequences available for selected members of Eubacteria by using the same method (16 species and 1368 characters). Phylogenetic trees were generated from sequence alignments by using computer algorithms implemented in the PHYLIP 3.572 (Felsenstein, 1989) and PAUP 3.1.1 (Swofford, 1991) computer software packages. Trees inferred from the groEL and 16s rRNA sequences were obtained by using two parsimony methods (the DNAPARS program in PHYLIP and the heuristic search algorithm of PAUP) and two search methods [the least-squares method of Fitch–Margoliash (Fitch & Margoliash, 1967) and neighbour-joining methods (Saitou & Nei, 1987) as executed in PHYLIP]. Trees based on groEL amino acid sequences were generated by using parsimony methods (the PROTPARS algorithm of PHYLIP). For the trees generated by the PROTPARS, DNAPARS, PAUP, Fitch–Margoliash and neighbour-joining methods, 100 bootstrap replicates were conducted by the method described by Felsenstein (1985) as implemented by SEQBOOT within PHYLIP and PAUP. For the distance-based phylogenetic methods listed above, estimated evolutionary distances between each pair of sequences were calculated for input into the tree-reconstruction algorithms. Pairwise distances between GroEL proteins were calculated using PROTDIST in PHYLIP. Pairwise distances between respective sequences for both 16s rRNA and groEL were calculated by using the two-parameter method of Kimura (1980) as provided by DNADIST in PHYLIP and used for the Fitch–Margoliash and neighbour-joining methods.

*coxella burnetii,* part of the γ-subclass of Proteobacteria, was selected as an outgroup in the unrooted analysis. This organism was maintained as the outgroup for all analyses conducted for this study.

Nucleotide sequence accession numbers. The groEL gene and 16s rRNA GenBank accession numbers for the organisms used for comparison can be found in Table 1.

RESULTS AND DISCUSSION

Levels of similarity between 16s rRNA and groEL nucleotide sequences

The 16s rRNA levels of similarity (Table 2) for all species used in this analysis ranged from 78-0 to 99-6%. The levels of similarity among Bartonella species ranged from 97-3 to 99-6%. The non-Bartonella species had levels of similarity amongst each other ranging from 78-0 to 97-5%.

The groEL sequences obtained from the current study and GenBank revealed lower percentage similarity values when compared to 16s rRNA. A pairwise comparison of groEL sequences (Table 2) revealed similarity values ranging from 73-1 to 92-0% for Bartonella species. A range of similarity values from 49-2 to 87-9% was obtained from all non-Bartonella species studied.

16s rRNA phylogeny

Multiple phylogenetic methods were used to analyse the data obtained from this study to determine the consistency of the trees between methods. One method...
of analysis would not be adequate because of the different assumptions, algorithms and tree-building techniques called upon by each phylogenetic method.

The 16S rRNA sequences from organisms used in the analysis were chosen based upon the relative relatedness to the genus *Bartonella* and the availability of the organism’s respective groEL DNA sequence from GenBank (release 101). Proper alignment of the 16S rRNA sequences required the removal of a sequence of groEL (~~~~) representing a secondary structure unique to the 16S rRNA sequence of *Escherichia coli* (Escherichia coli). Additionally, 12 bases (positions 152-163) representing an arbitrary cutoff point of <70 bootstrap replicates to be considered insignificant. Species within the *Bartonella* cluster demonstrated a range of bootstrap replicates from very low (unlabelled nodes represent <10 bootstrap replicates) to high (ranging from 52 to 100 for parsimony methods). Both methods have bootstrap replicate values that indicate significant statistical support for species outside the *Bartonella* cluster. Authors have argued that bootstrap proportion is an approximation of a statistical confidence interval (Berry & Gascuel, 1996; Felsenstein, 1985). By extension, this would allow an arbitrary cutoff point of <70 bootstrap replicates to be considered insignificant. Species within the *Bartonella* cluster demonstrated a range of bootstrap replicates from very low (unlabelled nodes represent <10 bootstrap replicates) to high (ranging from 52 to 100 for parsimony methods). Distance methods were unable to resolve the relationships within the *Bartonella* cluster among *Bartonella doshiae*, *Bartonella grahamii*, *Bartonella henselae*, *Bartonella quintana*, *Bartonella vinsonii* and *Bartonella vinsonii* subsp. *berkofii*. Though the close relationships shared between the human pathogens *Bartonella quintana* and *Bartonella henselae* and between the vole isolate *Bartonella vinsonii* and the subspecies *Bartonella vinsonii* subsp. *berkofii*, as observed with 16S rRNA and *gltA* phylogeny and DNA-DNA hybridization studies, are clearly established, they were not supported in the distance analysis. Parsimony methods were also unable to resolve the relationships within the *Bartonella* cluster. The high bootstrap replicate values were not supported by the observed relationships between species. The *Bartonella quintana–Bartonella henselae* cluster and distant
relatedness of the flagellated *Bartonella clarridgeiae* to the rest of the *Bartonella* species are supported by bootstrap replicates and prior observations; however, the clustering of the non-flagellated *Bartonella elizabethae* with the flagellated *Bartonella bacilliformis* is supported by bootstrap replicates but not by prior 16S rRNA, DNA–DNA hybridization and *gltA* studies.

**groEL phylogeny**

The *groEL* sequences generated from the *Bartonella* species were 1368 bp and represented sequence between reference positions 223 and 1630 of the *Bartonella bacilliformis* *groEL* sequence. The primers and conditions selected for this amplification allowed us to amplify all of the *Bartonella* species, with the exception of *Bartonella vinsonii*, which required the annealing temperature of the PCR reaction to be dropped to 48 °C.

The tree topology obtained from the distance methods of Fitch–Margoliash (Fig. 1) and neighbour-joining methods (data not shown) using the *groEL* dataset was supported by previous genotypic and phenotypic observations for all species. Outside the genus *Bartonella*, the bootstrap replicate values range from 86 to 100, indicating statistical support. With the exception of the *Bartonella bacilliformis–Bartonella clarridgeiae* cluster and the branch dividing the *Bartonella henselae–Bartonella quintana* and *Bartonella vinsonii–Bartonella vinsonii* subsp. *berkoffii* clusters, the bootstrap replicate values for all of the *Bartonella* species were significant (range 82–100). The relationships that the human pathogens *Bartonella quintana* and *Bartonella henselae*, the flagellated *Bartonella bacilliformis* and *Bartonella clarridgeiae*, *Bartonella vinsonii* and its subspecies *Bartonella vinsonii* subsp. *berkoffii* and *Bartonella elizabethae* and *Bartonella grahamii* (the previously observed *gltA* phylogenetic relationships; Birtles & Raoult, 1996) share were maintained by the respective branches observed in the current analysis. The tree topology and bootstrap replicate values obtained using neighbour-joining distance methods were very similar (data not shown).

Using parsimony methods (Fig. 2), the tree topology

---

**Fig. 1.** Comparison of Fitch distance matrix trees for members of the genus *Bartonella* and other phylogenetically proximal Eubacteria derived from 16S rRNA (a) and *groEL* (b) data. The support at each branch is indicated by the value at each node based upon 100 bootstrap replicates. Lengths of the vertical lines are not significant. Scale bar represents evolutionary distance as calculated using Kimura two-parameter distance calculation. Heavily dashed branches represent branches that were placed over one another for illustrative purposes and are not to be interpreted as branches originating from the rightmost node. Trees are unrooted.
approximates the results obtained using distance methods. The topologies observed in the trees obtained from both methods of analysis were identical; however, the bootstrap replicates showed greater support using parsimony analysis. The topologies and statistical support of all trees regardless of the method used, are consistent with previously observed genotypic and phenotypic characteristics.

**GroEL protein phylogeny**

Base compositions and codon usage differences in different phylogenetic and taxonomic lineages have been reported to constitute potential sources of inconsistencies when nucleotide sequences (including those of rRNA) are compared in phylogenetic studies (Hasagawa & Hashimoto, 1993; Olsen & Woese, 1993; Steel et al., 1993). Substitutional bias is thought to be minimized in highly conserved proteins (Hasegawa & Hashimoto, 1993; Jukes & Bhushan, 1986). Inferences based on comparisons of the amino acid sequences of highly conserved proteins have been proposed to be more reliable than those based on the corresponding nucleotide sequences (Hasegawa & Hashimoto, 1993). Sequencing the GroEL proteins of each *Bartonella* species was beyond the scope of this study and were compared to trees generated by nucleic acid sequences (data not shown). The trees inferred from the amino acid sequences presented topologies having little statistical support within the *Bartonella* genus. However, like 16S rRNA, the extragenus branching was supported (greater than 71% of the bootstrap replicates). Within the *Bartonella* genus, the cluster containing *Bartonella vinsonii* and *Bartonella vinsonii* subsp. *berkoffii* was supported (greater than 95% of the bootstrap replicates). The observed relationship between the flagellated bartonellae, *Bartonella bacilliformis* and *Bartonella claridgeae*, was well supported (greater than 75% of bootstrap replicates). Placement of the other *Bartonella* species in the inferred trees was not supported statistically (ranging from 10 to 54% of bootstrap replicates) or with genotypic and phenotypic observations (data not shown).

**Comparison of 16S rRNA and groEL phylogenies**

Our results indicate that outside the *Bartonella* cluster, the statistical significance of the proposed branching order and overall topology of groEL-inferred trees were better resolved than those of 16S rRNA-inferred trees. When parsimony analysis was performed using both PHYLIP (Fig. 2) and PAUP (data not shown), all
five extragenus branches in the groEL-generated trees were retained with 100% of bootstrap samples. The 16S rRNA-generated trees were equally robust at this level (100% of bootstrap samples at each node), though the node between Brucella abortus and the cluster of A. tumefaciens and S. meliloti was retained with 76% of bootstrap samples. When a distance matrix was performed with groEL data, all five extragenus branches had support of more than 86% of bootstrap samples. Similarly, the stability of the proposed branching order of the groEL-inferred trees showed more support (range of 53–100 bootstrap replicates for distance methods and 80–100 bootstrap replicates for parsimony methods) within the genus Bartonella than the 16S rRNA (range of 4–100 bootstrap replicates for distance methods and 12–100 bootstrap replicates for parsimony methods). Additionally, the relationships observed between the plant pathogen Agrobacterium tumefaciens, the nitrogen fixers Sinorhizobium meliloti and Bradyrhizobium japonicum, and the position of the cattle pathogen Brucella abortus vary slightly in topology between the two types of analysis and between the two genes, indicating that even closely related species as used in this analysis are not fully resolved, though each has strong support.

Citrate synthase (gltA) has been demonstrated to be a more reliable molecule for inferring phylogenetic relationships between Rickettsia species (Roux et al., 1997) and Bartonella species (Birtles & Raoul, 1996), when compared to 16S rRNA. The groEL gene has been demonstrated to be a robust and information-rich evolutionary tool for inferring phylogenetic relationships among eubacteria (Viale et al., 1994). Citrate synthase and groEL sequences are not available at present for most bacterial species, though the number of sequences for each is increasing.

The current study was limited by the availability of groEL sequences in GenBank, many of which are unavailable for taxa used in previous studies. The taxa selected have closer relationships with each other phylogenetically, which explains much of the ambiguity observed with the 16S rRNA analysis. Previously published gltA sequence similarities among the Bartonella taxa ranged from 83.8 to 91% (Birtles & Raoul, 1996), indicating that the DNA sequence diversity of groEL falls unevenly between 16S rRNA and gltA. Comparing the results of these phylogenetic reports on Bartonella species (namely 16S rRNA and gltA) with this report is hampered because these two genes and groEL are not equivocal, since each group of sequences (Birtles & Raoul, 1996) is not of equal length or equally represented in GenBank, but represent either the entire gene (16S rRNA) or greater than 80% (both gltA and groEL, 946 and 1368 bp, respectively).

Multiple gene data sets for microbial phylogenetic analysis have been extremely useful in determining the precise molecular relationships between genera and species (Eisen, 1995; Olsen & Woese, 1993; Roux et al., 1997). Phylogenies substantially different from the Woese concept have been proposed, based on alignments of several housekeeping genes (Benachenhou-Lahfa et al., 1994; Birtles & Raoul, 1996; Brown et al., 1994; Creti et al., 1991; Gupta, 1995; Karlin et al., 1995; Kruse et al., 1996; Kumada et al., 1993; Roux et al., 1997; Tiboni et al., 1993; Viale et al., 1994; Wang & Shakes, 1996). These studies and others have clearly shown that phylogenies based on different genes show discrepancies.

Current molecular identification and classification of bacteria (many of which are at present uncultivable) to the genus level rely solely, if not mainly, on 16S rRNA gene sequences. 16S rRNA can be considered ineffectual as a tool for inferring phylogenetic relationships at the Bartonella species level compared to gltA (Birtles & Raoul, 1996). Classification based upon only one gene of an entire genome does not allow for accurate resolution of phylogenetic relationships, though it does help classify newly discovered organisms to the genus level. DNA–DNA hybridization may not be appropriate for routine identification and classification of Bartonella species because of the cost and time involved in preparing large-scale matrix analyses each time a new isolate is obtained.

The groEL gene offers an ideal tool to resolve the topological and statistical ambiguities created when using 16S rRNA. The groEL gene provides better statistical support than gltA within the Bartonella species, though both offer topologies that are supported by current knowledge of these organisms. The combination of these three genes allows for one to place an organism in the Bartonella genus and to infer its relationship to the other Bartonella species with greater confidence than if one only uses a single molecule for determining the intraspecies relationships.

Putative new isolates of Bartonella, based upon 16S rRNA and gltA sequence comparisons, are being identified in large numbers in rodent populations (Birtles & Raoul, 1996; Kosoy et al., 1997; Maurin et al., 1997). At present these new isolates, as well as many of the currently recognized species, require a more thorough phenotypic and genetic characterization.

As data on the members of the genus Bartonella accumulate and phylogenetic inference programming takes advantage of faster computers and larger data sets, it would be interesting to combine genetic, protein, structural, conformational and physiological interactions of 16S rRNA, GltA, GroEL, as well as other genes that are later characterized and sequenced, with phenotypically characteristic data sets to better infer the evolutionary relationships within the genus and better understand the natural history of Bartonella species.
ACKNOWLEDGEMENTS

We thank Barbara Ellis and Lorenza Beati for their helpful suggestions and review of this manuscript and Jim Olson for his support.

REFERENCES


Bartonella phylogeny using groEL


