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

Centers for Disease Control
and Prevention, Bacterial
& Mycotic Diseases,
Mailstop G-051, and Viral &
Rickettsial Diseases,
Mailstop G-132, 1600
Clifton Road, Atlanta,
GA 30333, USA

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-Rios *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 clarridgeiae* (Kordick *et al.*, 1997), *Bartonella vinsonii* subsp. *berkoffii* (Breitschwerdt *et al.*, 1995), *Bartonella talpae* (Birtles & Raoult, 1996), *Bartonella peromysci* (Birtles & Raoult, 1996), *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 *Grahamella* (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 & Wisseman, 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

The GenBank accession numbers for the *groEL* sequences determined in this paper are AF014830–AF014836.

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' CGTGAAGTTGCCTCAAAACC 3') and BbHS1630.n (5' AATCCATTCCGCCCATTC 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 × 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 Preps (Promega) according to manufacturer's specifications.

Cloning. All cloning was performed by using SureShot (Invitrogen) *Escherichia coli* (INVαF': F' *endA1 recA1 hsdR17* (*r_k⁻*, *m_k⁺*) *supE44 thi-1 gyrA96 relA1 φ80lacZΔM15 Δ(lacZYA-argF)U169* l-).

PCR-amplified fragments were ligated into pCR2.1 vector, transformed into INV1αF'-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 Qiaprep-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

Table 1. Bacterial strains studied

ATCC, American Type Culture Collection, Manassas, VA, USA; NCTC, National Collection of Type Cultures, Central Public Health Laboratory, London, UK. ^T, Type strain.

Organism	Strain	Source		<i>groEL</i>		16S rRNA	
		Collection no.	Reference	GenBank no.	Reference	GenBank no.	Reference
<i>Bartonella bacilliformis</i>	KC584T (LA6.3)*	ATCC 35686	Brenner <i>et al.</i> (1993)	M98257	Unpublished†	Z70003	Birtles & Raoult (1996)
<i>Bartonella clarridgeiae</i>	Houston-2 ^T	ATCC 51734	Clarridge <i>et al.</i> (1995)	AF014831	This study	X89208	Lawson <i>et al.</i> (1996)
<i>Bartonella doshae</i>	R18T	NCTC 12862	Birtles <i>et al.</i> (1995)	AF014832	This study	Z31351	Birtles <i>et al.</i> (1995)
<i>Bartonella elizabethae</i>	F9251 ^T	ATCC 49927	Daly <i>et al.</i> (1993)	AF014834	This study	L01260	Daly <i>et al.</i> (1993)
<i>Bartonella grahamii</i>	V2 ^T	NCTC 12860	Birtles <i>et al.</i> (1995)	AF014833	This study	Z31349	Birtles <i>et al.</i> (1995)
<i>Bartonella henselae</i>	Houston-1 ^T	ATCC 49882	Regnery <i>et al.</i> (1992b)	AF014829	This study	M73229	Regnery <i>et al.</i> (1992b)
<i>Bartonella quintana</i>	OK-90-268 ^T	ATCC VR-358	Brenner <i>et al.</i> (1993)	AF014830	This study	M73228	Regnery <i>et al.</i> (1992b)
<i>Bartonella vinsonii</i>	Baker ^T	ATCC VR-152	Baker (1946)	AF014835	This study	Z31352	Birtles <i>et al.</i> (1995)
<i>Bartonella vinsonii</i> subsp. <i>berkoffii</i>	93-CO1 ^T	ATCC 51672	Breitschwerdt <i>et al.</i> (1995)	AF014836	This study	U26258	Brenner <i>et al.</i> (1993)
<i>Sinorhizobium meliloti</i>	Rm1021 (104A14)*	–	–	U19726	Rusanganwa & Gupta (1993)	M28846	Shatters <i>et al.</i> (1989)
<i>Coxiella burnetii</i>	CB9MIC7 (Q177)*	–	–	M20482	Ferreira <i>et al.</i> (1993)	M21291	Weisburg <i>et al.</i> (1989)
<i>Brucella abortus</i>	s19 (11-19)*	–	–	M82975	Gor & Mayfield (1992)	X13695	Dorsch <i>et al.</i> (1989)
<i>Agrobacterium tumefaciens</i>	C58 (NCPB2437)*	–	–	X68263	Segal & Ron (1996)	D14500	Sawada <i>et al.</i> (1993)
<i>Ehrlichia chaffeensis</i>	Arkansas (Sapulpa)*	–	–	L10917	Sumner <i>et al.</i> (1993)	U60476	Unpublished†
<i>Bradyrhizobium japonicum</i>	110spc4 (USDA6)*	–	–	Z22603	Fischer <i>et al.</i> (1993)	U69638	Unpublished†
<i>Cowdria ruminantium</i>	Welgevonden (Crystal Springs)*	–	–	U13638	Lally <i>et al.</i> (1995)	X61659	Dame <i>et al.</i> (1992)

* *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 distance 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 trees.

Coxiella 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

Table 2. Percentage similarity among *groEL* sequences and among 16S rRNA sequences

The values on the upper right are the uncorrected percentages of similarity between *groEL* sequences and the values on the lower left are the percentages of similarity between 16S rRNA gene sequences.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 <i>Coxiella burnetii</i>	—	55.0	55.7	60.7	61.4	61.9	60.0	61.6	61.8	57.8	62.5	61.9	61.3	60.9	55.8	62.4
2 <i>Cowdria ruminantium</i>	78.0	—	87.9	51.7	51.6	52.3	49.7	58.5	58.1	53.6	56.8	57.8	57.6	56.7	50.7	57.7
3 <i>Ehrlichia chaffeensis</i>	78.5	97.5	—	51.7	51.6	52.5	49.2	59.3	57.7	54.2	57.6	58.6	58.3	57.7	51.2	58.1
4 <i>Agrobacterium tumefaciens</i>	81.0	82.7	82.1	—	87.7	83.3	74.7	68.6	67.2	65.0	68.1	68.0	68.9	67.6	60.4	68.6
5 <i>Sinorhizobium meliloti</i>	81.3	82.4	82.3	94.5	—	83.8	77.5	67.2	66.2	63.6	68.0	67.1	68.1	66.4	59.5	68.2
6 <i>Brucella abortus</i>	80.4	82.0	82.1	93.7	94.2	—	76.2	71.0	70.6	66.7	72.2	71.6	72.8	71.1	62.9	72.6
7 <i>Bradyrhizobium japonicum</i>	80.0	80.8	80.2	87.5	88.1	88.0	—	63.0	62.1	58.7	63.9	63.2	64.2	62.6	55.4	64.3
8 <i>Bartonella bacilliformis</i>	80.2	82.1	83.0	92.5	92.1	94.2	86.3	—	84.0	78.4	83.0	83.3	84.5	83.3	73.1	84.7
9 <i>Bartonella clarridgeiae</i>	80.3	81.5	82.5	92.9	92.5	94.7	86.8	97.3	—	78.9	84.1	85.5	86.1	84.3	73.8	85.5
10 <i>Bartonella doshiae</i>	80.1	82.3	82.9	93.1	93.1	94.5	87.0	97.8	98.1	—	81.2	83.6	83.2	81.3	82.8	83.0
11 <i>Bartonella elizabethae</i>	80.5	81.7	82.5	92.8	92.5	94.9	86.7	97.9	97.3	98.6	—	89.6	87.2	85.5	77.0	87.3
12 <i>Bartonella grahamii</i>	80.2	82.3	82.9	93.7	93.4	94.5	87.5	98.2	98.2	99.2	98.9	—	90.0	86.8	79.4	89.1
13 <i>Bartonella henselae</i>	80.2	82.6	82.9	93.4	93.2	94.5	87.6	98.2	98.0	99.4	98.4	99.2	—	91.0	78.4	82.4
14 <i>Bartonella quintana</i>	80.1	82.5	82.8	93.0	92.8	94.5	87.1	97.9	97.8	98.9	98.2	98.8	98.9	—	90.8	77.5
15 <i>Bartonella vinsonii</i>	80.1	82.1	82.9	93.3	93.0	94.4	87.2	97.7	98.0	99.6	98.5	98.9	99.3	98.9	—	92.0
16 <i>Bartonella vinsonii</i> subsp. <i>berkoffii</i>	80.0	82.2	83.0	93.0	93.0	95.1	86.9	98.2	98.2	99.4	98.9	99.6	99.2	98.9	99.3	—

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 variable region found in *Bartonella* species corresponding to positions 183–187 (*Escherichia coli* structural model) identified by Birtles & Raoult (1996). Additionally, 12 bases (positions 152–163) representing a secondary structure unique to the 16S rRNA sequence of *Bartonella vinsonii* subsp. *berkoffii* were removed to ensure contiguous alignment.

The trees inferred from the 16S rRNA data set using both distance (Fig. 1) and parsimony methods (Fig. 2) demonstrated similar topologies and bootstrap replicate values. The tree topology obtained from the distance methods of Fitch–Margoliash (Fig. 1) and neighbour-joining methods (data not shown) using the 16S rRNA data set was supported by previous genotypic and phenotypic observations for all species outside the *Bartonella* cluster (the close taxonomic relationship between *Cowdria ruminantium* and *Ehrlichia chaffeensis*; between the plant pathogen *Agrobacterium tumefaciens*, the nitrogen fixers *Sinorhizobium meliloti* and *Bradyrhizobium japonicum*, and the position of the cattle pathogen *Brucella abortus*). The bootstrap replicate values for this group range from 91 to 100 for distance methods (with the

exception of the node between the entire *Bartonella* cluster and *Brucella abortus* which has 54 bootstrap replicates). The bootstrap replicate values range from 76 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 distance methods and from 69 to 100 for parsimony). 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. *berkoffii*. 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. *berkoffii*, 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

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.

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. 2. Comparison of parsimony trees (PHYLIP analysis) 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. Bootstrap replicate values < 10 are represented by unlabelled nodes. Lengths of both the vertical and horizontal lines are not significant. 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 leftmost 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. Trees were generated using the deduced amino acid sequences of the *Bartonella* species and the other organisms used in

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 clarridgeiae*, 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 & Raoult, 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 & Raoult, 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 & Raoult, 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 & Raoult, 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 & Raoult, 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 & Raoult, 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

- Baker, J. A. (1946).** A rickettsial infection in Canadian voles. *J Exp Med* **84**, 37–50.
- Benachenhou-Lahfa, N., Labedan, B. & Forterre, P. (1994).** PCR-mediated cloning and sequencing of the gene encoding glutamate dehydrogenase from the archaeon *Sulfolobus shibatae*: identification of putative amino-acid signatures for extremophilic adaption. *Gene* **140**, 17–24.
- Berry, V. & Gascuel, O. (1996).** On the interpretation of bootstrap trees: appropriate threshold of clade selection and induced gain. *Mol Biol Evol* **13**, 999–1011.
- Birtles, R. J. & Raoult, D. (1996).** Comparison of partial citrate synthase gene (*gltA*) sequences for phylogenetic analysis of *Bartonella* species. *Int J Syst Bacteriol* **46**, 891–897.
- Birtles, R. J., Harrison, T. G., Saunders, N. A. & Molyneux, D. H. (1995).** Proposals to unify the genera *Grahamella* and *Bartonella*, with descriptions of *Bartonella talpae* comb. nov., *Bartonella peromysci* comb. nov., and three new species, *Bartonella grahamii* sp. nov., *Bartonella taylorii* sp. nov., and *Bartonella doshiae* sp. nov. *Int J Syst Bacteriol* **45**, 1–8.
- Breitschwerdt, E. B., Kordick, D. L., Malarkey, D. E., Keene, B., Hadfield, T. L. & Wilson, K. (1995).** Endocarditis in a dog due to infection with a novel *Bartonella* subspecies. *J Clin Microbiol* **33**, 154–160.
- Brenner, D. J., O'Connor, S. P., Winkler, H. H. & Steigerwalt, A. G. (1993).** Proposals to unify the genera *Bartonella* and *Rochalimaea*, with descriptions of *Bartonella quintana* comb. nov., *Bartonella vinsonii* comb. nov., *Bartonella henselae* comb. nov., and *Bartonella elizabethae* comb. nov., and to remove the family Bartonellaceae from the order Rickettsiales. *Int J Syst Bacteriol* **43**, 777–786.
- Brown, J. R., Masuchi, Y., Robb, F. T. & Doolittle, W. F. (1994).** Evolutionary relationships of bacterial and archaeal glutamine synthetase genes. *J Mol Evol* **38**, 566–576.
- Cáceres-Rios, H., Rodríguez-Tafur, J., Bravo-Puccio, F., Maguina-Vargas, C., Díaz, C. S., Ramos, D. C. & Patarca, R. (1995).** Verruga peruana: an infectious endemic angiomatosis. *Crit Rev Oncog* **6**, 47–56.
- Checa, S. K. & Viale, A. M. (1997).** The 70-kDa heat-shock protein/DnaK chaperone system is required for the productive folding of ribulose-bisphosphate carboxylase subunits in *Escherichia coli*. *Eur J Biochem* **248**, 848–855.
- Claridge, J. E., Raich, T. J., Pirwani, D. & 8 other authors (1995).** Strategy to detect and identify *Bartonella* species in routine clinical laboratory yields *Bartonella henselae* from human immunodeficiency virus-positive patient and unique *Bartonella* strain from his cat. *J Clin Microbiol* **33**, 2107–2113.
- Creti, R., Citarella, F., Tibone, O., Sanangelantoni, A., Palm, P. & Cammarano, P. (1991).** Nucleotide sequence of a DNA region comprising the gene for elongation factor 1a (EF-1a) from the ultrathermophilic archaeote *Pyrococcus woesei*: phylogenetic implications. *J Mol Evol* **33**, 332–342.
- Daly, J. S., Worthington, M. G., Brenner, D. J. & 7 other authors (1993).** *Rochalimaea elizabethae* sp. nov. isolated from a patient with endocarditis. *J Clin Microbiol* **31**, 872–881.
- Dame, J. B., Mahan, S. M. & Yowell, C. A. (1992).** Phylogenetic relationship of *Cowdria ruminantium*, agent of heartwater, to *Anaplasma marginale* and other members of the order Rickettsiales determined on the basis of 16S rRNA sequence. *Int J Syst Bacteriol* **42**, 270–274.
- Dolan, M. J., Wong, M. T., Regnery, R. L., Jorgensen, J. H., Garcia, M., Peters, J. & Drehner, D. (1993).** Syndrome of *Rochalimaea henselae* adenitis suggesting cat scratch disease. *Ann Intern Med* **118**, 331–336.
- Dorsch, M., Moreno, E. & Stackebrandt, E. (1989).** Nucleotide sequence of the 16S rRNA from *Brucella abortus*. *Nucleic Acids Res* **17**, 1765.
- Eisen, J. A. (1995).** The RecA protein as a model molecule for molecular systematics studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. *J Mol Evol* **41**, 1105–1123.
- Felsenstein, J. (1985).** Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.
- Felsenstein, J. (1989).** PHYLIP: phylogeny inference package. *Cladistics* **5**, 164–166.
- Ferreira, R. G., Soncini, F. C. & Viale, A. M. (1993).** Cloning, characterization, and functional expression in *Escherichia coli* of chaperonin (*groESL*) genes from the phototrophic sulfur bacterium *Chromatium vinosum*. *J Bacteriol* **175**, 1514–1523.
- Fischer, H. M., Babst, M., Kaspar, T., Acuna, G., Arigoni, F. & Hennecke, H. (1993).** One member of a *groESL*-like chaperonin multigene family in *Bradyrhizobium japonicum* is co-regulated with symbiotic nitrogen fixation genes. *EMBO J* **12**, 2901–2912.
- Fitch, W. M. & Margoliash, E. (1967).** Construction of phylogenetic trees. *Science* **155**, 279–284.
- Fox, G. E., Stackebrandt, E., Hepsell, R. B. & 16 other authors (1980).** The phylogeny of prokaryotes. *Science* **209**, 457–463.
- Genetics Computer Group (1996).** Wisconsin Sequence Analysis Package, version 9.0. Genetics Computer Group, Madison, USA.
- Gor, D. & Mayfield, J. E. (1992).** Cloning and nucleotide sequence of the *Brucella abortus* *groE* operon. *Biochim Biophys Acta* **1130**, 120–122.
- Gupta, R. S. (1995).** Evolution of the chaperonin families (Hsp60, Hsp10, and Tcp-1) of proteins and the origin of eukaryotic cells. *Mol Microbiol* **15**, 1–11.
- Hasegawa, M. & Hashimoto, T. (1993).** Ribosomal RNA trees misleading? *Nature* **361**, 23.
- Hoekstra, M. F. (1988).** Lid lysates: an economical and rapid method for plasmid analysis. *Biotechniques* **6**, 930–932.
- Jukes, T. H. & Bhushan, V. (1986).** Silent nucleotide substitutions and G + C content of some mitochondrial and bacterial genes. *J Mol Evol* **24**, 39–44.
- Karlin, S., Weinstock, G. M. & Brendel, V. (1995).** Bacterial classifications derived from RecA protein sequence comparisons. *J Bacteriol* **177**, 6881–6893.
- Kimura, M. (1980).** A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120.
- Kordick, D. L., Hilyard, E. J., Hadfield, T. L., Wilson, K. H., Steigerwalt, A. G., Brenner, D. J. & Breitschwerdt, E. B. (1997).** *Bartonella claridgeiae*, a newly recognized zoonotic pathogen causing inoculation papules, fever, and lymphadenopathy (cat scratch disease). *J Clin Microbiol* **35**, 1813–1818.

- Kosoy, M., Regnery, R., Tzianabos, T., Marston, E., Jones, D., Green, D., Maupin, G., Olson, J. & Childs, J. (1997). Distribution, diversity and host specificity of *Bartonella* in rodents from the southeastern United States. *Am J Trop Med Hyg* **57**, 578–588.
- Kruse, M., Gamulin, V., Cetkovic, H., Pancer, Z., Müller, I. M. & Müller, E. G. (1996). Molecular evolution of the metazoan protein kinase C multigene family. *J Mol Evol* **43**, 374–383.
- Kumada, Y., Benson, D. R., Hillemann, D., Hosted, T. J., Rochefor, D. A., Thompson, C. J., Wohlleben, W. & Tateno, Y. (1993). Evolution of the glutamine synthetase gene, one of the oldest existing and functioning genes. *Proc Natl Acad Sci USA* **90**, 3009–3013.
- Lally, N. C., Nicoll, S., Paxton, E. A., Cary, C. M. & Sumption, K. J. (1995). The *Cowdria ruminantium groE* operon. *Microbiology* **141**, 2091–2100.
- Lawson, P. A., Clarridge, J. E. & Collins, M. D. (1996). Description of *Bartonella-clarridgeiae* sp. nov. isolated from the cat of a patient with *Bartonella henselae*-septicemia. *Med Microbiol Lett* **5**, 64–73.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1989). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Masui, S., Sasaki, T. & Ishikawa, H. (1997). *groE*-homologous operon of *Wolbachia*, an intracellular symbiont of arthropods: a new approach for their phylogeny. *Zool Sci* **14**, 701–706.
- Maurin, M., Birtles, R. & Raoult, D. (1997). Current knowledge of *Bartonella* species. *Eur J Clin Microbiol Infect Dis* **16**, 487–506.
- Myers, W. F. & Wisseman, C. L. (1980). Genetic relatedness among the Typhus group of *Rickettsiae*. *Int J Syst Bacteriol* **30**, 143–159.
- Myers, W. F., Wisseman, C. L., Fiset, P., Oaks, E. V. & Smith, J. F. (1979). Taxonomic relationship of vole agent to *Rochalimaea quintana*. *Infect Immun* **26**, 976–983.
- Norman, A. F., Regnery, R., Jameson, P., Greene, C. & Krause, D. C. (1995). Differentiation of *Bartonella*-like isolates at the species level by PCR-restriction fragment length polymorphism in the citrate synthase gene. *J Clin Microbiol* **33**, 1797–1803.
- Olsen, G. J. & Woese, C. R. (1993). Ribosomal RNA: a key to phylogeny. *FASEB J* **7**, 113–123.
- Regnery, R. L., Anderson, B. E., Clarridge, J. E., Rodríguez-Barradas, M. C., Jones, D. C. & Carr, J. H. (1992a). Characterization of a novel *Rochalimaea* species, *R. henselae* sp. nov., isolated from blood of a febrile, human immunodeficiency virus-positive patient. *J Clin Microbiol* **30**, 265–274.
- Regnery, R., Martin, M. & Olson, J. (1992b). Naturally occurring “*Rochalimaea henselae*” infection in domestic cat. *Lancet* **340**, 557–558.
- Roux, V., Rydkina, E., Ereemeeva, M. & Raoult, D. (1997). Citrate synthase comparison, a new tool for phylogenetic analysis, and its application for the *Rickettsiae*. *Int J Syst Bacteriol* **47**, 252–261.
- Rusanganwa, E. & Gupta, R. S. (1993). Cloning and characterization of multiple *groEL* chaperonin-encoding genes in *Rhizobium meliloti*. *Gene* **126**, 67–75.
- Saitou, N. & Nei, M. (1987). The neighbor joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.
- Sawada, H., Ieki, H., Oyaizu, H. & Matsumoto, S. (1993). Proposal for rejection of *Agrobacterium tumefaciens* and revised descriptions for the genus *Agrobacterium* and for *Agrobacterium radiobacter* and *Agrobacterium rhizogenes*. *Int J Syst Bacteriol* **43**, 694–702.
- Segal, G. & Ron, E. Z. (1995). The *groESL* operon of *Agrobacterium tumefaciens*: evidence for heat shock-dependent mRNA cleavage. *J Bacteriol* **177**, 750–757.
- Segal, R. & Ron, E. Z. (1996). Regulation and organization of the *groE* and *dnaK* operons in Eubacteria. *FEMS Microbiol Lett* **138**, 1–10.
- Shatters, R. G., Somerville, J. E. & Kahn, M. L. (1989). Regulation of glutamine synthetase II activity in *Rhizobium meliloti* 104a14. *J Bacteriol* **171**, 5087–5094.
- Staden, R. (1982). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Res* **10**, 2951–2961.
- Steel, M. A., Lockhart, P. J. & Penny, D. (1993). Confidence in evolutionary trees from biological sequence data. *Nature* **364**, 440–442.
- Sumner, J. W., Sims, K. G., Jones, D. C. & Anderson, B. E. (1993). *Ehrlichia chaffeensis* expresses an immunoreactive protein homologous to the *Escherichia coli* GroEL protein. *Infect Immun* **61**, 3536–3569.
- Swofford, D. (1991). Phylogenetic analysis using parsimony (PAUP), version 3.1.1. Champaign, IL: Illinois Natural History Survey.
- Tiboni, O., Cammarano, P. & Sanangelantoni, A. M. (1993). Cloning and sequencing of the gene encoding glutamine synthetase I from the archeum *Pyrococcus woesei*: anomalous phylogenetics inferred from analysis of archeal and bacterial glutamine synthetase I sequences. *J Bacteriol* **175**, 2271–2277.
- Viale, A. M., Arakaki, A. K., Soncini, F. C. & Ferreyra, R. G. (1994). Evolutionary relationships among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisons. *Int J Syst Bacteriol* **44**, 527–533.
- Vinson, J. W. (1966). In vitro cultivation of the rickettsial agent of trench fever. *Bull WHO* **35**, 155–164.
- Wang, W. & Shakes, D. C. (1996). Molecular evolution of the 14-3-3 protein family. *J Mol Evol* **43**, 384–398.
- Weisburg, W. G., Dobson, M. E., Samuel, J. E., Dasch, G. A., Mallavia, L. P., Mandelco, L., Sechrest, J. E., Weiss, E. & Woese, C. R. (1989). Phylogenetic diversity of the *Rickettsiae*. *J Bacteriol* **171**, 4202–4206.
- Weiss, E., Dasch, G. A., Woodman, D. R. & Williams, J. C. (1978). Vole agent identified as a strain of the trench fever rickettsia, *Rochalimaea quintana*. *Infect Immun* **19**, 1013–1020.
- Weiss, E., Mamay, H. K. & Dasch, G. A. (1982). Ornithine metabolism in the genus *Rochalimaea*. *J Bacteriol* **150**, 245–250.
- Westfall, H. N., Edman, D. C. & Weiss, E. (1984). Analysis of fatty acids of the genus *Rochalimaea* by electron capture gas chromatography: detection of nonanoic acid. *J Clin Microbiol* **19**, 305–310.
- Woese, C. R., Kandler, O. & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* **87**, 4576–4579.