Gene-sequence-based criteria for species definition in bacteriology: the *Bartonella* paradigm

Bernard La Scola, Zaher Zeaiter, Atieh Khamis and Didier Raoult

Unité des Rickettsies, CNRS UPRESA 6020, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 5, France

The definition of new species is currently based on polyphasic classification that includes both determination of phenotypic characteristics and DNA–DNA homology. However, none of these techniques is convenient for the rapid characterization of fastidious or non-culturable bacteria. Using sequences available in the GenBank database, we compared the similarities of gene fragments among the currently recognized *Bartonella* species. This comparison led to both the definition of similarity values that discriminated *Bartonella* at the species level and assessment of the relative discriminatory power of each gene examined. In this perspective, *rpo*B and *glt*A were found to be the most potent.

Definition of new species is currently based on results of DNA–DNA hybridization and description of phenotypic traits, so-called polyphasic characterization [1]. This method has two major drawbacks: the practical difficulty associated with performing DNA–DNA hybridization, and the frequent scarcity of distinguishing phenotypic characteristics. Moreover, for fastidious bacteria, detection from clinical material is frequently based on PCR detection methods without a culture step. Although development amplification and sequencing of the 16S rDNA has simplified the identification and detection of fastidious bacteria, isolation followed by DNA–DNA hybridization remains central to the delineation of species [2,3]. Recently however, the sequencing of several housekeeping genes has been proposed as an alternative to DNA–DNA hybridization [4]. Members of the genus *Bartonella* are fastidious bacteria that possess few phenotypic markers useful for species delineation. Comparison of DNA sequence data has served as the most commonly used approach to *Bartonella* species identification. Given that numerous loci have now been characterized in different *Bartonella* species, analysis of the relative value of each for the delineation of species is timely.

The Bartonella genus

The genus *Bartonella* contains 16 species and three subspecies. *Bartonella* possess a natural cycle that typically involves persistent intra-erythrocytic infection in a reservoir host, and an arthropod vector that transmits the bacteria between the reservoir hosts or, accidentally, to disease-susceptible hosts, including humans [5]. The ability of *Bartonella* to cause zoonoses has stimulated investigation of an increasing number of potential animal hosts, which has led to increasing encounters with new species. As for other bacteria, delineation of *Bartonella* species has been based on polyphasic characterization. However, *Bartonellae* remain inert in most biochemical tests. Indeed, phenotypic differentiation has relied solely on the presence or absence of flagellae and determination of pre-formed enzymes using systems elaborated for anaerobes. Although these assays served well when the genus was first described, its expansion quickly revealed limitations [6]. For example, *B. schoenbuchensis* is undistinguishable from several *Bartonella* species, including *B. henselae* and *B. quintana* [6]. Moreover, in clinical laboratories, identification is mostly based on results of gene amplification, and phenotypic characters are rarely if ever used. The actual abundance of genes proposed for taxonomic purposes in the genus *Bartonella* reflect the need for an alternative procedure to polyphasic classification [7]. Accession numbers for sequences where the complete or partial sequence is known are available for nearly all *Bartonella* species (Table 1). Specifically, these sequences comprise those of 16S rDNA, *glt*A, *gro* EL, *rpo* B, *fts* Z, *rib*C and internal transcribed spacer (ITS).

Comparison of *Bartonella* sequences

Multi-sequence alignment was performed for each sequence using CLUSTAL W software (Version 1.81) (http://spiral.genes.nig.ac.jp/homology/clustalw-e.shtml) [8], and DNA sequence similarities were calculated using the MEGA 2.1 software (http://www.megasoftware.net) [9]. The extent of the differences between sequences was calculated with pairwise deletion parameters. Results of the similarities were used to construct a matrix in which, for each locus examined, the percentage of similarity with the closest relative was included. This matrix (Fig. 1) indicates the discriminating power and minimal similarity percentage of each gene that defines a species. Discriminatory power (DP) is well-defined by the median value of inter-species similarity, whereas minimal similarity (MS) is defined as the highest inter-species similarity score (Table 2). 16S rDNA, with a DP of 99.7% and an MS of 99.7%, is the least discriminatory gene. Nevertheless, it ensures good differentiation of *Brucella abortus* (with 94%

Corresponding author: Didier Raoult (didier.raoult@medecine.univ-mrs.fr).

 0966-842X/03/$ - see front matter doi:10.1016/S0966-842X(03)00143-4

Table 1. Bacterial strains and sequences used in the study of the *Bartonella* genus[a]

Species (strain)	Collection number	Accession number						
		16S	*Glt*A	*Rpo*B	ITS	*Gro*EL	*Fts*Z	*Rib*C
Bartonella alsatica (IBS382^T)	CIP 105477	AJ002139	AF204273	AF165987	AF312506	AF299357	AF467763	AY116630
Bartonella bacilliformis (KC584^T)	ATCC 35686	Z11683	U28076	AF165988	L26364	Z15160	AF007266	AJ236918
Bartonella birtlesii (IBS 325^T)	CIP 106294	AF204274	AF204272	AF165989	AY116640	AF355773	AF467762	AY116632
Bartonella bovis (91-4^T)	CIP 106692	AF199502	AF293394	AF166581[b]	AY116638	AF071194	AF467761	AY116637
Bartonella clarridgeiae (Houston-2^T)	ATCC 51734	U64691	U84386	AF165990	AF167989	AF014831	AF141018	AJ23616
Bartonella doshiae (R18^T)	NCTC 12862	Z31351	AF207827	AF165991	AJ269786	AF014832	AF467754	AY116627
Bartonella elizabethae (F9251^T)	ATCC 49927	L01260	U28072	AF165992	L35103	AF014834	AF467760	AY116633
Bartonella grahamii (V2^T)	NCTC 12860	Z31349	Z70016	AF165993	AJ269785	AFD14833	AF467753	AY1666583[b]
Bartonella henselae (Houston-1^T)	ATCC 49882	M73229	L38987	AF171070	L35101	AF014829	AF061746	AJ132928
Bartonella koehlerae (C-29^T)	ATCC 700693	AF076237	AF176091	AY166580[b]	AF312490	AY116641	AF467755	AY116634
Bartonella quintana (Fuller^T)	ATCC VR-358	M11927	Z70014	AF165994	L35100	AF014830	AF061747	AJ236917
Bartonella schoenbuchensis (R1^T)	NCTC 13165T	AJ278187	AJ278183	AY167409[b]	AY116639	AY116642	AF467765	AY116628
Bartonella taylorii (M6^T)	NCTC 12861	Z31350	AF191502	AF165995	AJ269784	AF304017	AF467756	AY116635
Bartonella tribocorum (IBS 506^T)	CIP 104576	AJ003070	AJ005494	AF165996	AF312505	AF304018	AF467759	Not done
Bartonella vinsonii subsp. *arupensis* (OK 94-513^T)	ATCC 700727	AF214558	AF214557	AY166582[b]	AF312504	AF304016	AF467758	AY116631
Bartonella vinsonii subsp. *berkhoffii* (93-CO1^T)	ATCC 51672	U26258	U28075	AF165989	AF167988	AF014836	AF467764	AY116629
Bartonella vinsonii subsp.*vinsonii* (Baker^T)	ATCC VR-152	M73230	Z70015	AF165997	L35102	AF014835	AF467757	AY116636

[a]Abbreviations: ATCC, American Type Culture Collection; CIP, Collection de l'Institut Pasteur; ITS, internal transcribed spacer; NCTC, National Collection of Type Culture.
[b]These were sequenced specifically for this manuscript.

similarity), which is one of the closest genera to *Bartonella* (Fig. 1). The *glt*A, *gro* EL, *rpo* B and *fts* Z genes, and ITS all have good discriminating power, with DPs ranging from 92.6% to 94.4% (Table 2), but the gene with the highest discriminating power appears to be *rib*C, which has a DP of 86.5%. Nevertheless, a low DP is not sufficient to ensure that two species are clearly differentiated, because for *gro* EL, *fts* Z, *rib*C and ITS, *B. schoenbuchensis*, *B. bovis*, *B. grahamii* and *B. taylorii* have similarity values ranging from 97.9% to 99.8%. Only two genes – *rpo* B and *glt*A – have an MS of ≤96%. The relationship between the *B. vinsonii* subspecies was investigated using these two genes. *Bartonella vinsonii vinsonii* and *B. vinsonii berkhoffii* were the closest taxa, with 96% similarity for both *glt*A and *rpo* B sequences. The closest taxon to *B. vinsonii arupensis*, assessed using *glt*A, was *B. vinsonii vinsonii* (95.7%). Using an *rpo* B comparison, *B. vinsonii vinsonii* and *B. vinsonii berkhoffii* both shared 95.5% similarity with *B. vinsonii arupensis*.

Cut-off sequence similarity values for species definition in *Bartonella*

With the exception of 16S rDNA, all the sequences studied have good DPs, ranging from 94.46% to 87.4% for *fts* Z and *rib*C, respectively. Nevertheless, despite having low DPs, the usefulness of *gro* EL, *rib*C, *fts* Z and ITS is restricted by high similarity values between some species. This high similarity is reproducibly observed between pairs of phylogenetically close relatives, such as *B. schoenbuchensis* and *B. birtlesii*, or *B. taylorii* and *B. grahamii*. The only genes with lower inter-species similarity values are *rpo* B and *glt*A. The discriminatory power of *glt*A is restricted by the short length of most sequences deposited in Genbank (http://www.ncbi.nlm.nih.gov/) [typically only 327 base pairs (bp)]. The high discriminatory power of *rpo* B has been described in numerous bacterial genera, including spirochetes, *Legionella, Staphylococcus, Mycobacterium* and *Enterobacteriaceae* [10–16]. We propose that newly encountered *Bartonella* isolates should be considered a new species if a 327-bp *glt*A fragment shares <96.0% sequence similarity with those of the validated species, and if an 825-bp *rpo* B fragment shares <95.4% sequence similarity with those of the validated species. Subspecies definition is more difficult because no clear definition of a 'subspecies' is available to date. *Bartonella vinsonii vinsonii* and *B. vinsonii berkhoffii,* with 96% similarity in *glt*A and *rpo* B gene sequences, fulfil the criteria for being considered as the same species. For *B. vinsonii arupensis,* an *rpo* B sequence similarity of 95.5% and a *glt*A sequence similarity of 95.7% with the two other *B. vinsonii* subspecies are on either side of the limits defined earlier. Failure of the *glt*A comparison to comply with the defined limit could be interpreted as a shortfall. However, when the subspecies was characterized by DNA–DNA hybridization, it was found that, although it possessed >70% homology with the other *B. vinsonii* subspecies, the calculated Δ_{Tm} values were 5–7% – a range beyond that required for species validation [17]. Therefore, conclusions drawn from *glt*A–*rpo* B sequence analysis and DNA–DNA hybridization appear to be congruent.

Table 2. Percentage sequence similarity of *Bartonella* with its closest relative[a]

Gene	Percentage similarity		
	Median value	Highest value	Lowest value
16S rDNA	99.7	99.8	98.3
ITS	93.9	99.8	89.2
*Glt*A	93.6	96.0	87.1
Gro EL	92.6	99.4	86.2
Rpo B	92.8	95.4	90.9
Fts Z	94.4	97.9	88.8
*Rib*C	86.5	99.5	82.8

[a]Percentages for 16S rDNA, *glt*A, *gro* EL, *rpo* B, *fts* Z, *rib*C and internal transcribed spacer (ITS) were calculated on the basis of sites 1356, 1340, 327, 1185, 825, 788, 621, respectively.

Concluding remarks

To date, developments in gene amplification and sequencing – especially for the gene encoding 16S rDNA – have not particularly facilitated the procedure required for classification and identification of new bacterial species. For several species, including *Bacillus* or *Bosea* [3,18,19], comparison of 16S rDNA alone is not sensitive enough to

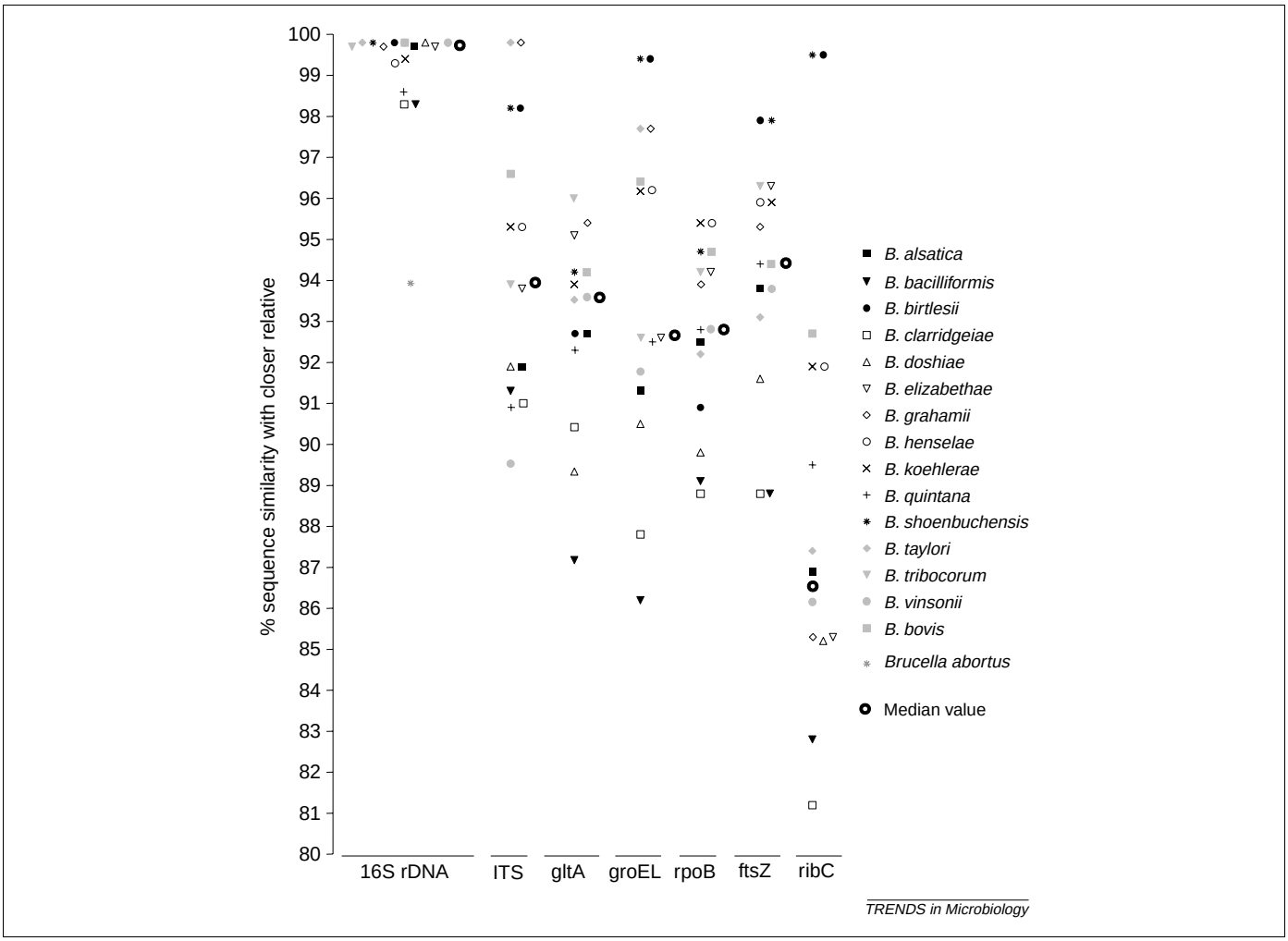

Fig. 1. Percentage sequence similarity for all species of *Bartonella* with its closest relative for 16S rDNA, *glt*A, *gro*EL, *rpo*B, *fts*Z, *rib*C and internal transcribed spacer (ITS). Percentages were calculated for 16S rDNA, *glt*A, *gro*EL, *rpo*B, *fts*Z, *rib*C and ITS on the basis of sites 1356, 1340, 327, 1185, 825, 788, 621, respectively. The median for each gene is represented by a bold circle. The *rib*C sequence was not available for *B. tribocorum*.

discriminate confidently between bacterial species in some genera and must therefore be accompanied by DNA–DNA hybridization if reliable delineations are to be made [2]. To date, two strains can be assigned to the same species if their DNA–DNA relatedness is of $<70\%$ homology and if the heteroduplex has a Δ_{Tm} of $<5\%$. However, the 16S rDNA sequence appears to be the best tool to order prokaryotic taxa hierarchically among the ranks of genera [4], as observed for *Bartonella* (Fig. 1). Comparison of sequences from more divergent parts of the genome has enabled better delineation of species with genera that are difficult to differentiate by phenotypic methods.

The *Ad Hoc* committee for the re-evaluation of the species definition in bacteriology has proposed that, in the future, new species description could be based on house-keeping-gene-sequence analysis with a comparison of <5 genes, provided there is a sufficient degree of congruence between the technique used and DNA–DNA reassociation [4]. We agree that sequencing of multiple genes is necessary for sequence-based delineation of species. Study of the sequences of these genes for all formally accepted species will then enable determination of the sequence similarity cut-off above which a new strain might be assigned to a new species, provided there is congruency with results of phylogenetic analysis based on gene sequence analyses giving high bootstrap values [20]. Nevertheless, after the analysis is completed, only the sequences of gene(s) that ensure good discrimination should be necessary for the identification of new species. In the event that a putative new species of *Bartonella* is isolated, DNA–DNA hybridization will need to be used in the unlikely case where *glt*A and *rpo*B analyses do not fulfil the required criteria. In other genera, such as *Legionella* or *Mycobacteria*, where phenotypic characteristics are not sufficient for comprehensive characterization of increasingly described species (as for *Bartonella*), recent publications of complete sets of gene sequences (i.e. 16S rDNA, *mip*, *rpo*B, *hsp*65, rpoB and ITS) provide the opportunity to use the procedure described here. However, the definition of similarity cut-off for these genera will be required first, as it is probable that these cut-offs will be genera-specific and will thus have to be re-validated.

Acknowledgements

We are indebted to Richard J. Birtles for reviewing this manuscript. Z.Z. was awarded by the 'Fondation pour la Recherche Médicale'.

References

1 Wayne, L.G. *et al.* (1987) Report of the *Ad hoc* committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* 37, 463–464

2 Stackebrandt, E. and Goebel, B.M. (1994) A place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* 44, 846–849

3 Fox, G.E. *et al.* (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int. J. Syst. Bacteriol.* 42, 166–170

4 Stackebrandt, E. *et al.* (2002) Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 52, 1043–1047

5 Breitschwerdt, E.B. *et al.* (2000) *Bartonella* infection in animals: carriership, reservoir potential, pathogenicity, and zoonotic potential for human infection. *Clin. Microbiol. Rev.* 13, 428–438

6 Dehio, C. *et al.* (2001) *Bartonella schoenbuchii* sp. nov., isolated from the blood of wild roe deer. *Int. J. Syst. Evol. Microbiol.* 51, 1557–1565

7 Houpikian, P. and Raoult, D. (2001) Molecular phylogeny of the genus *Bartonella*: what is the current knowledge? *FEMS Microbiol. Lett.* 200, 1–7

8 Thompson, J.D. *et al.* (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680

9 Kumar, S. *et al.* (2001) *MEGA2: Molecular Evolutionary Genetics Analysis Software*, Arizona State University

10 Renesto, P. *et al.* (2001) Use of *rpoB* gene analysis for detection and identification of *Bartonella* species. *J. Clin. Microbiol.* 39, 430–437

11 Renesto, P. *et al.* (2000) RpoB gene analysis as a novel strategy for identification of spirochetes from the genera *Borrelia*, *Treponema*, and *Leptospira*. *J. Clin. Microbiol.* 38, 2200–2203

12 Drancourt, M. and Raoult, D. (2002) rpoB gene sequence-based identification of *Staphylococcus* species. *J. Clin. Microbiol.* 40, 1333–1338

13 Mollet, C. *et al.* (1997) *rpoB* sequence analysis as a novel basis for bacterial identification. *Mol. Microbiol.* 26, 1005–1011

14 Ko, K.S. *et al.* (2002) Application of RNA polymerase β-subunit gene (*rpoB*) sequences for the molecular differentiation of *Legionella* species. *J. Clin. Microbiol.* 40, 2653–2658

15 Abbott, K.C. *et al.* (1991) Hemophagocytic syndrome – a cause of pancytopenia in human ehrlichiosis. *Am. J. Hematol.* 38, 230–234

16 Kim, B.J. *et al.* (1999) Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (*rpoB*). *J. Clin. Microbiol.* 37, 1714–1720

17 Welch, D.F. *et al.* (1999) Isolation of a new subspecies, *Bartonella vinsonii* subsp. *arupensis*, from a cattle rancher: identity with isolates found in conjunction with *Borrelia burgdorferi* and *Babesia microti* among naturally infected mice. *J. Clin. Microbiol.* 37, 2598–2601

18 Ash, C. *et al.* (1991) Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequençing of 16S rRNA. *Int. J. Syst. Bacteriol.* 41, 343–346

19 La Scola, B. *et al.* (2002) *Bosea eneae* sp. nov., *Bosea massiliensis* sp. nov. and *Bosea vestrisii* sp. nov., isolated from water supplies, and emendation of the genus *Bosea* (Das 1996). *Int. J. Syst. Evol. Microbiol.* 53, 15–20

20 Pitulle, C. *et al.* (2002) Investigation of the phylogenetic relationships within the genus *Bartonella* based on comparative sequence analysis of the *rnpB* gene, 16S rDNA and 23S rDNA. *Int. J. Syst. Evol. Microbiol.* 52, 2075–2080