Complete *rpoB* gene sequencing as a suitable supplement to DNA–DNA hybridization for bacterial species and genus delineation

Toïdi Adékambi,^{1,2} Thomas M. Shinnick,² Didier Raoult¹
and Michel Drancourt¹

Correspondence
Michel Drancourt
Michel.Drancourt
@medecine.univ-mrs.fr

¹Unité des Rickettsies, CNRS UMR 6020 IFR 48, Faculté de Médecine, Université de la Méditerranée, Marseille, France

²Mycobacteriology Laboratory Branch, Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA

DNA–DNA hybridization (DDH), the gold standard for bacterial species delineation, is a laborious method and the alternative, average nucleotide identity (ANI), a genomic sequence-derived parameter, is not applicable to non-sequenced species. A universal cut-off value to delineate bacterial species does not exist, yet a DDH value <70 % and ANI <95 ± 0.5 % have proved useful in selected examples. We herein compare published values for DDH and ANI with sequence similarity of *rpoB* gene sequences retrieved from GenBank for strains of 230 bacterial species representative of 45 genera. Intraspecific *rpoB* sequence similarity was 98.2–100 %. We observed that an *rpoB* gene sequence similarity ≤97.7 % significantly correlated with a DDH value <70 % and an ANI value <94.3 %. An *rpoB* gene sequence similarity <85.5 % correlated with membership of different genera. When applied to fastidious and as-yet-uncultivated organisms lacking experimental DDH values, these cut-off values suggested that '*Candidatus* Blochmannia pennsylvanicus' and '*Candidatus* Blochmannia floridarius' may belong to different genera, that the different endosymbiotic *Buchnera aphidicola* organisms may belong to different genera and that, while the tsetse fly enteric symbiont *Sodalis glossinidius* may belong to the *Enterobacteriaceae*, the endocellular obligate symbiont *Wigglesworthia glossinidia* from the same host may belong to the group of as-yet-uncultivated gammaproteobacteria. *rpoB* gene sequence similarity provides an efficient supplement to DDH and ANI measurements to delineate bacterial species and genera, including delineation of as-yet-uncultivated, non-sequenced organisms.

INTRODUCTION

DNA–DNA hybridization (DDH) has been adopted as the primary molecular tool to delineate bacterial species (Stackebrandt & Goebel, 1994; Wayne *et al.*, 1987), and a 70 % cut-off is most often used to place organisms into different species, although some species have higher threshold values. Despite being introduced several decades ago, DDH data are consistent with recent results from complete genome sequences. Nevertheless, DDH is a time-consuming and labour-intensive method, requiring the manipulation of large amounts of biomass of possibly hazardous organisms, and is ill-suited to rapid delineation of prokaryotes. Also, it is not applicable to the study of

prokaryotes that are currently non-culturable, now recognized to represent the majority of prokaryotes in the biosphere (Amann *et al.*, 1995; Rappé & Giovannoni, 2003; Riesenfeld *et al.*, 2004). It also requires pairwise comparisons of DNA isolated from the test organism and members of characterized species sharing >97 % 16S rRNA gene sequence similarity (Stackebrandt & Goebel, 1994); individual strains cannot be analysed and compared with a database using a codified set of criteria to assign them to a known taxon or to propose a new taxon (Gevers *et al.*, 2005).

At the beginning of the 1990s, it was found that bacterial isolates exhibiting <97 % 16S rRNA gene sequence similarity usually shared <70 % DDH and belonged to different species. On the other hand, isolates exhibiting ≥97 % similarity might or might not meet the 70 % DDH criterion for inclusion in the same species (Fox *et al.*, 1992; Stackebrandt & Goebel, 1994). A recent study increased the recommended 16S rRNA gene sequence similarity from 97

Abbreviations: ANI, average nucleotide identity; DDH, DNA–DNA hybridization.

Details of strains included in this study and a maximum-likelihood tree based on *rpoB* gene sequences of gammaproteobacteria are available as supplementary material with the online version of this paper.

to 98.7–99 % in order to facilitate taxonomic studies without sacrificing the quality and precision of a 'species' description (Stackebrandt & Ebers, 2006). In this study (Stackebrandt & Ebers, 2006), about 20 % of 16S rRNA gene sequences exhibited a similarity value >99 % with corresponding DDH values <70 %.

In the present study, we first compared values of published *rpoB* gene sequence similarity derived from complete genome sequences and GenBank with DDH values and secondarily we examined the relationship of complete *rpoB* gene sequence similarity and the genomic sequence-derived parameter average nucleotide identity (ANI) (Konstantinidis & Tiedje, 2005a). The ANI of conserved genes in two sequenced strains represents a robust measure of the genetic and evolutionary distance, it correlates with 16S rRNA gene sequence similarity and the genome mutation rate, it is not affected by lateral transfer or variable recombination rates of single (or a few) genes and it offers resolution at the subspecies level (Konstantinidis & Tiedje, 2005a). *rpoB* is a single-copy gene, it belongs to the common set of genes (Koonin, 2003) and it is long enough to contain phylogenetically useful information (Adékambi & Drancourt, 2004; Drancourt & Raoult, 1999; Khamis *et al.*, 2003, 2004; La Scola *et al.*, 2006; Mollet *et al.*, 1998; Taillardat-Bisch *et al.*, 2003). It might be less prone than the 16S rRNA gene to lateral gene transfer (Case *et al.*, 2007; Korczak *et al.*, 2006) due to its housekeeping function and it has already been used for bacterial species delineation in selected cases (Adékambi *et al.*, 2004, 2006a, b, c) as well as to estimate the DNA G + C content of whole bacterial genomes (Fournier *et al.*, 2006). Among some alternative protein-coding genes, the *rpoB* gene showed the highest correlation to average amino-acid identity, which reflects whole genome-level relatedness (Konstantinidis & Tiedje, 2005b).

METHODS

DDH and ANI data mining. We screened DDH values contained in articles published during the last three decades (see Supplementary Table S1 in IJSEM Online). The DDH values were obtained using the microtitre plate technique (Ezaki *et al.*, 1989), the spectrophotometric technique (De Ley *et al.*, 1970), the membrane filter method (Tourova & Antonov, 1987) and dot hybridization (Yip *et al.*, 2007). We further analysed ANI measurements previously determined for 85 bacterial genomes (Goris *et al.*, 2007; Konstantinidis & Tiedje, 2005a) (Supplementary Table S2).

Sequence analysis. We recovered complete 16S rRNA gene and *rpoB* gene sequences from the same bacterial strains in GenBank in order to constitute a local database. A BLASTN search (Altschul *et al.*, 1990) identified the best hits for the 16S rRNA gene and *rpoB* gene in other species (<http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi>). However, genes appearing to be the most similar based on BLAST hits may often not represent the closest relative phylogenetically (Koski & Golding, 2001). Also, with direct BLASTN with *rpoB* sequences exhibiting <85.5 % similarity, manual analysis was necessary in most situations. Percentages of similarity between 16S rRNA gene sequences or between *rpoB* gene sequences were determined using the CLUSTAL W program supported by the PBIL

website. The *rpoB* sequences were obtained from strains of the same bacterial species used in DDH experiments and ANI measurements (Supplementary Tables S1 and S2), although the strains sequenced were not necessarily the same. The maximum-likelihood tree was reconstructed using PhyML software with GTR (gamma distribution and invariable sites) as the substitution model and 100 bootstrap replications (Guindon & Gascuel, 2003).

RESULTS AND DISCUSSION

Correlation between *rpoB* gene sequence similarities and DDH values

GenBank contains over 25 000 sequences annotated as being partial or complete *rpoB* gene sequences. Complete *rpoB* gene sequences (~3000 sequences) ranged from 3411 bp (*Staphylococcus aureus*) to 4185 bp (*Neisseria meningitidis*). The quality of the *rpoB* gene sequence subset is likely to be reliable since most of the *rpoB* gene sequences in GenBank are recent (deposited in the last 10 years). The plot of DDH values against *rpoB* gene sequence similarity for 318 pairwise comparisons fits a linear model (Fig. 1). The correlation between the two parameters is high ($r^2=0.82$) and the linear relationship between the two parameters was best described by the regression formula DDH value = $5.98(\textit{rpoB similarity}) - 516.1$. However, the r^2 value was probably <0.9 because reciprocal DDH values may differ by up to 15 % and DDH values varied according to the method used. None of these methods is straightforward to apply without thorough training. Also, a significant number of physico-chemical parameters, genome size, the presence of large plasmids, DNA purity and other factors, influence the hybridization (Stackebrandt & Ebers, 2006). DDH values may also not represent the actual sequence identity because detectable DNA heteroduplexes will only form between strands that show at least 80 % sequence identity (Coenye & Vandamme, 2004; Rosselló-Mora & Amann, 2001). Multilocus sequence analysis (MLSA) has been proposed as an alternative method that can replace DDH experiments (Gevers *et al.*, 2005; Kuhnert & Korczak, 2006; Yamamoto *et al.*, 1999). However, only partial gene sequences were involved in these analyses, limiting the accuracy of the conclusions. Also, these studies were limited to one genus or family, limiting their usefulness across a large range of bacteria.

Species delineation cut-off

Our analysis indicated that an *rpoB* gene sequence similarity above 97.7 % should be useful to delineate species (Fig. 2). Using this cut-off value, species exhibiting a 16S rRNA gene sequence similarity ≥ 99 % and a DDH value <70 % with closest related species were resolved by the complete *rpoB* gene sequence in 82/85 cases (96.5 %) (Fig. 3). According to the *rpoB*-based criterion, a 99.7 % *rpoB* sequence similarity confirmed the close phylogenetic relationship between *Borrelia recurrentis*, which causes louse-borne relapsing fever worldwide, and *Borrelia*

Fig. 1. Scatter plot of the relationship between *rpoB* gene sequence similarity and DDH (%).

duttonii, which causes East African tick-borne relapsing fever (Cutler *et al.*, 1999). Also, this criterion agreed with DDH criteria that place *Escherichia coli* and *Shigella flexneri* in the same species as well as placing *Salmonella enterica* serovar Typhimurium and *Salmonella enterica* serovar Typhi in the same species, *Bacillus anthracis* and *Bacillus cereus* in the same species and *Mycobacterium tuberculosis* complex isolates in the same species. *Mycobacterium tuberculosis* complex isolates have been regarded as variants of a single clone (Sreevatsan *et al.*, 1997) and would become clones Tuberculosis, Africanum, Bovis and Microti (Lan & Reeves, 2001). Also, *Corynebacterium glucuronolyticum* and *Corynebacterium seminale* were placed in the same species, with a DDH value of 78 %, *rpoB* sequence similarity of 99.7 % and 16S rRNA gene sequence similarity

of 99.9 % (Devriese *et al.*, 2000; Khamis *et al.*, 2004). A complete *rpoB* gene sequence similarity >99.4 % also agreed with the previous proposition that the genus *Brucella* contains only one species (Verger *et al.*, 1985, 1987), with DDH >90 % among all nomenspecies (Marianelli *et al.*, 2006). Our data agreed with the re-elevation of *Pseudomonas syringae* pv. *tomato* to a novel species (Gardan *et al.*, 1999) because it displayed only 41–46 % DDH and 95.4 % *rpoB* sequence similarity but 99.0 % 16S rRNA gene sequence similarity with *Pseudomonas syringae* pv. *syringae*.

There are a few more complicated cases with respect to speciation in our dataset. Discrepancies in species delineation of isolates with *rpoB* sequence similarities >97.7 %, 16S rRNA gene sequence similarities >99 % and DDH values <70 % were observed in pairwise comparisons of

Fig. 2. Comparison of *rpoB* gene sequence similarities (>88.5 %) and DDH values. Interspecies comparisons are indicated by filled diamonds, whereas intraspecies comparisons are indicated by open squares.

Fig. 3. Comparison of 16S rRNA gene sequence similarities (>98 %) and DDH values. Interspecies comparisons are indicated by filled diamonds, whereas intraspecies comparisons are indicated by open squares.

Mycobacterium fortuitum with *Mycobacterium houstonense* and *Mycobacterium senegalense* with *Mycobacterium houstonense*, suggesting the re-examination of *Mycobacterium houstonense* (Adékambi & Drancourt, 2004; Adékambi *et al.*, 2003; Schinsky *et al.*, 2004). As for *Mycobacterium ulcerans* and *Mycobacterium marinum*, a 99.6 % *rpoB* sequence similarity and >98 % whole genome sequence similarity (Stinear *et al.*, 2007) contrasted with a DDH value <59 % (Yip *et al.*, 2007). The presence of a 174 kb virulence plasmid, two bacteriophages, multiple copies of IS2404 and multiple DNA deletions and rearrangements in *Mycobacterium ulcerans* may explain, at least in part, the strikingly low value for DDH (Stinear *et al.*, 2007; Yip *et al.*, 2007). These genome sequence data led to the conclusion that *Mycobacterium ulcerans*, a mycolactone-producing mycobacterium, had recently evolved via lateral gene transfer and reductive evolution from the environmental, mycolactone-non-producing species *Mycobacterium marinum* to become a niche-adapted specialist (Stinear *et al.*, 2007).

Borrelia bisetii and *Borrelia burgdorferi* exhibited 95.7 % *rpoB* sequence similarity but 99.7 % 16S rRNA gene sequence similarity and a 72.8 % DDH value. However, caution should be taken in interpreting this exception because a reciprocal DDH value was not available (Masuzawa *et al.*, 2001) and the DDH value was close to the 70 % cut-off. Prokaryotic species are considered to be groups of isolates that share 50–70 % DDH and 5–7 % difference in thermal stability between the homologous and heterologous duplexes (Ursing *et al.*, 1995). Further experimental DDH data are therefore required in order to interpret this exception.

The 16S rRNA gene sequences of *Burkholderia mallei*, *Burkholderia pseudomallei* and *Burkholderia thailandensis* are more than 99 % similar. Comparison of complete *rpoB* gene sequences derived from their genome sequences

showed that *Burkholderia pseudomallei* and *Burkholderia thailandensis* shared 97.1 % similarity (DDH value <47 %) (Yabuuchi *et al.*, 2000) whereas *Burkholderia pseudomallei* and *Burkholderia mallei* shared 99.8 % similarity (DDH value >76 %) (Rogul *et al.*, 1970). This finding was in agreement with MLSA results based on seven housekeeping genes, *ace*, *gltB*, *gmhD*, *lepA*, *lipA*, *narK* and *ndh* (Godoy *et al.*, 2003). This was also true for *Burkholderia cenocepacia* and *Burkholderia vietnamiensis*, which shared 97.7 % *rpoB* sequence similarity. These examples illustrate the contribution of the *rpoB* gene sequence approach to the resolution of taxonomic ambiguities.

Furthermore, the intraspecies *rpoB* sequence similarity of 440 isolates belonging to 83 species, 44 genera and 14 phyla ranged from 98.2 to 100 % (Supplementary Table S3). However, there is overlapping *rpoB* similarity between isolates of *Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringiensis*, *Mycobacterium tuberculosis* and *Mycobacterium bovis*, *Burkholderia mallei* and *Burkholderia pseudomallei*, *Escherichia coli* and *Shigella* species, *Yersinia pestis* and *Yersinia pseudotuberculosis*, and *Brucella* species.

Correlation between *rpoB* gene sequence similarity and ANI

The plot of 134 values derived from 85 bacterial genomes (Supplementary Table S2) shows that ANI correlates more strongly ($r^2=0.93$ for linear model) with *rpoB* gene sequence similarity (Fig. 4) than with 16S rRNA gene sequence identity ($r^2=0.79$ for linear model) (Konstantinidis & Tiedje, 2005a) and shows that *rpoB* gene sequence similarity can resolve areas where the 16S rRNA gene is inadequate, such as the species level. The 97.7 % *rpoB* gene sequence similarity cut-off for speciation of closely related bacterial strains corresponds to 94.3 % ANI (Fig. 4). The latter value is similar to the ANI of 95 ± 0.5 % recently found by Goris *et al.* (2007) to

Fig. 4. Scatter plot of the relationship between *rpoB* gene sequence similarities and ANI (%).

correspond to the recommended cut-off point of 70 % DDH for species delineation. A borderline 94.3 % ANI measurement has been found between *Neisseria gonorrhoeae* and *Neisseria meningitidis* (Konstantinidis & Tiedje, 2005a).

Genus delineation cut-off

Despite the fact that the 16S rRNA gene sequence similarity cut-off for genus delineation is different in different bacterial clades, the published cut-off of <95 % 16S rRNA gene sequence similarity has been used to delineate genera otherwise delineated by other methods (Clarridge, 2004; Schloss & Handelsman, 2004). Using this value as a standard, complete *rpoB* gene sequence similarity <85.5 % delineated genera in 33 of 35 cases (94.3 %) (Fig. 1). For example, [*Haemophilus*] *ducreyi* exhibited only 81.1 % *rpoB* sequence similarity with *Haemophilus influenzae*, 80.5 % with *Pasteurella multocida* and 83.7 % with *Mannheimia haemolytica* but 86.6 % with *Actinobacillus pleuropneumoniae*. It is therefore a member of the same genus as *Actinobacillus pleuropneumoniae*. These data are in agreement with a previous analysis based on five housekeeping genes (Kuhnert & Korczak, 2006) that suggested the transfer of [*Haemophilus*] *ducreyi* to the genus *Actinobacillus*. The *rpoB* sequence similarities of >93 % among *Escherichia*, *Shigella* and *Salmonella* species support the suggestion that they belong to a single genus and should be considered highly related (Brenner *et al.*, 1969, 1972; Crosa *et al.*, 1973). The genus *Yersinia* was found to be at the borderline of the enteric group, with *rpoB* gene sequence similarity <85.5 % (Supplementary Table S1). An *rpoB* gene sequence similarity <85.5 % was found for species of the family *Pasteurellaceae*, for which the taxonomy of the genera, species and subspecies is not well settled (Blackall *et al.*, 2007; Christensen *et al.*, 2007; Kuhnert & Korczak, 2006).

Our analysis suggested that *Acidovorax avenae* and '*Verminephrobacter eiseniae*' may belong to the same genus, because they shared 87.2 % *rpoB* sequence similarity and 95 % 16S rRNA gene sequence similarity. *Geobacter sulfurreducens* and *Geobacter metallireducens* may belong to different genera (*rpoB* sequence similarity 82 %, 16S rRNA gene sequence similarity <95 %), as may *Pelobacter carbinolicus* and *Pelobacter propionicus* (*rpoB* sequence similarity 63.6 %, 16S rRNA gene sequence similarity <90 %) and *Desulfovibrio vulgaris* and *Desulfovibrio desulfuricans* (*rpoB* sequence similarity 78 %, 16S rRNA gene sequence similarity <90 %).

The species in the genera *Afipia*, *Bradyrhizobium*, *Rhodopseudomonas* and *Nitrobacter* may belong to the same genus (*rpoB* sequence similarity >85.8 %, 16S rRNA gene sequence similarity >98 %). Similarly, species in the genera *Rhizobium*, *Sinorhizobium* and *Agrobacterium* (*rpoB* sequence similarity >87 %, 16S rRNA gene sequence similarity >95 %) may belong to the same genus. However, the genera *Bosea* and *Mesorhizobium* were

distinct but closely related to these genera, as found previously (Khamis *et al.*, 2003). Also, *Anabaena* and *Nostoc* species may belong to the same genus (94.2 % *rpoB* sequence similarity, 98.9 % 16S rRNA gene sequence similarity).

The situation was more complex in the genus *Mycoplasma* using data derived from complete genome sequences. In the *Mycoplasma pneumoniae* group (Kim *et al.*, 2003), *Mycoplasma pneumoniae* and *Mycoplasma genitalium* shared 97.8 % 16S rRNA gene sequence similarity and only 75 % *rpoB* sequence similarity, *Mycoplasma pneumoniae* and *Mycoplasma gallisepticum* shared 59.5 % *rpoB* and 89.5 % 16S rRNA gene sequence similarity, *Mycoplasma pneumoniae* and *Mycoplasma penetrans* shared 57.1 % *rpoB* and 81.4 % 16S rRNA gene sequence similarity, *Mycoplasma genitalium* and *Mycoplasma gallisepticum* shared 61.3 % *rpoB* and 89.5 % 16S rRNA gene sequence similarity, *Mycoplasma genitalium* and *Mycoplasma penetrans* shared 60.7 % *rpoB* and 81.3 % 16S rRNA gene sequence similarity and *Mycoplasma gallisepticum* and *Mycoplasma penetrans* shared 64.2 % *rpoB* and 82.8 % 16S rRNA gene sequence similarity. This group was distant from the *Mycoplasma hominis* group (Kim *et al.*, 2003), with which it shared <50 % *rpoB* and <76 % 16S rRNA gene sequence similarity. In the latter group, *Mycoplasma synoviae* and *Mycoplasma pulmonis* shared only 66.2 % *rpoB* and 82.9 % 16S rRNA gene sequence similarity, whereas *Mycoplasma capricolum* and *Mycoplasma mycoides* shared 95.8 % *rpoB* and 99.2 % 16S rRNA gene sequence similarity and 40 % DDH. This large divergence suggested that the current genus *Mycoplasma* includes members that could be classified into different genera. These data suggest that more taxonomic work is needed to reconcile speciation within this genus in order to provide an improved basis for *Mycoplasma* taxonomy (Gasparich *et al.*, 2004; van Passel *et al.*, 2005).

rpoB gene sequence-based prediction of DDH values for fastidious or uncultivated endosymbionts

When applying a <85.5 % *rpoB* similarity cut-off for the delineation of bacterial genera to as-yet-uncultivated organisms for which experimental DDH and biochemical parameters are therefore lacking, we found that '*Candidatus* Blochmannia floridarius' and '*Candidatus* Blochmannia pennsylvanicus' may belong to different genera, as they shared 80.6 % *rpoB* and 93.1 % 16S rRNA gene sequence similarity. This was also true for *Anaplasma marginale* and *Anaplasma phagocytophilum*, which shared 76.3 % *rpoB* sequence similarity and 92.3 % 16S rRNA gene sequence similarity, and for *Buchnera aphidicola* strains APS, Sg, Bp and Cc, which exhibited *rpoB* sequence similarities of 74.1–85.0 % and 16S rRNA gene sequence similarity of 88.7–94.1 %. Because strains of *Buchnera aphidicola* are known simply by the host aphid that they infect (Munson *et al.*, 1991), naming species based on host

association would lead to unmanageable proliferation of species names (Lo *et al.*, 2007).

Sodalis glossinidius (an enteric symbiont of tsetse flies) and *Wigglesworthia glossinidia* (an endocellular obligate symbiont of tsetse flies) have been classified within the *Enterobacteriaceae*. Our analysis showed that *Sodalis glossinidius* was indeed more closely related to species of the *Enterobacteriaceae* (82.9–85.0 % *rpoB* sequence similarity, <93 % 16S rRNA gene sequence similarity) than was *Wigglesworthia glossinidia* (63.7–66.0 % *rpoB* sequence similarity, <87 % 16S rRNA gene sequence similarity). Thus, the two organisms shared the name of their host species but were evolutionarily unrelated. Furthermore, most of the as-yet-uncultured gammaproteobacteria clustered separately from members of the *Enterobacteriaceae* in the maximum-likelihood tree, with a bootstrap value of 100 % (Supplementary Fig. S1). This confirmed that the as-yet-uncultured gammaproteobacterial lineage was clearly different from that of other recognized gammaproteobacteria. This was in agreement with the new composite measure termed 'genome conservation', where the phylogenetic analysis was not affected by the number of species incorporated in the dataset (Kunin *et al.*, 2005).

Our analysis also confirmed that *Wolbachia pipientis* wMel (intracellular bacterium of *Drosophila melanogaster*) (Wu *et al.*, 2004) and *Wolbachia* sp. wBm (endosymbiont of *Brugia malayi*) (Foster *et al.*, 2005) belonged to the same genus but not to the same species (87.6 % *rpoB* sequence similarity and 97.3 % 16S rRNA gene sequence similarity), with a predicted DDH value of about 7.7 %. This finding is in agreement with a recent proposal to re-examine taxonomic relationships of *Wolbachia pipientis* strains (Lo *et al.*, 2007).

Conclusions

Phenotypic, genotypic (DDH) and phylogenetic (16S rRNA gene) information is commonly used to determine a consensus taxonomy of bacteria (Vandamme *et al.*, 1996). While the 16S rRNA gene sequence often lacks resolution when compared with DDH, we found that the *rpoB* gene sequence correlated with ANI, a genomic sequence-derived parameter, and with DDH, an experimental comparison of whole genomes. *rpoB* gene sequence comparison may then be used as a first, preliminary indication of species and genus delineation. However, the lack of universal primers for *rpoB* gene amplification and sequencing due to the saturation of third codon positions over a long evolutionary timescale may limit such an application. Nonetheless, analysis of *rpoB* gene sequences derived from on-going genomic and metagenomic sequencing projects could allow estimation of DDH values for even fastidious and as-yet-uncultured organisms. We therefore propose *rpoB* gene sequence similarity as a suitable estimate of DDH for bacterial species and genus delineation.

ACKNOWLEDGEMENTS

We thank Bonnie B. Plikaytis for expert review of the manuscript. Funding was provided by the Unité des Rickettsies, Faculté de Médecine, Université de la Méditerranée, Marseille, France.

REFERENCES

- Adékambi, T. & Drancourt, M. (2004). Dissection of phylogenetic relationships among 19 rapidly growing *Mycobacterium* species by 16S rRNA, *hsp65*, *sodA*, *recA* and *rpoB* gene sequencing. *Int J Syst Evol Microbiol* **54**, 2095–2105.
- Adékambi, T., Colson, P. & Drancourt, M. (2003). *rpoB*-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria. *J Clin Microbiol* **41**, 5699–5708.
- Adékambi, T., Reynaud-Gaubert, M., Greub, G., Gevaudan, M. J., La Scola, B., Raoult, D. & Drancourt, M. (2004). Amoebal coculture of "Mycobacterium massiliense" sp. nov. from the sputum of a patient with hemoptoic pneumonia. *J Clin Microbiol* **42**, 5493–5501.
- Adékambi, T., Berger, P., Raoult, D. & Drancourt, M. (2006a). *rpoB* gene sequence-based characterization of emerging non-tuberculous mycobacteria with descriptions of *Mycobacterium bolletii* sp. nov., *Mycobacterium phocaicum* sp. nov. and *Mycobacterium aubagnense* sp. nov. *Int J Syst Evol Microbiol* **56**, 133–143.
- Adékambi, T., Stein, A., Carvajal, J., Raoult, D. & Drancourt, M. (2006b). Description of *Mycobacterium conceptionense* sp. nov., a *Mycobacterium fortuitum* group organism isolated from a posttraumatic osteitis inflammation. *J Clin Microbiol* **44**, 1268–1273.
- Adékambi, T., Raoult, D. & Drancourt, M. (2006c). *Mycobacterium barrassiae* sp. nov., a *Mycobacterium morioakaense* group species associated with chronic pneumonia. *J Clin Microbiol* **44**, 3493–3498.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.
- Amann, R. I., Ludwig, W. & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**, 143–169.
- Blackall, P. J., Bojesen, A. M., Christensen, H. & Bisgaard, M. (2007). Reclassification of [*Pasteurella*] *trehalosi* as *Bibersteinia trehalosi* gen. nov., comb. nov. *Int J Syst Evol Microbiol* **57**, 666–674.
- Brenner, D. J., Fanning, G. R., Johnson, K. E., Citarella, R. V. & Falkow, S. (1969). Polynucleotide sequence relationships among members of *Enterobacteriaceae*. *J Bacteriol* **98**, 637–650.
- Brenner, D. J., Fanning, G. R., Skerman, F. J. & Falkow, S. (1972). Polynucleotide sequence divergence among strains of *Escherichia coli* and closely related organisms. *J Bacteriol* **109**, 953–965.
- Case, R. J., Boucher, Y., Dahllöf, I., Holmstrom, C., Doolittle, W. F. & Kjelleberg, S. (2007). Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. *Appl Environ Microbiol* **73**, 278–288.
- Christensen, H., Kuhnert, P., Busse, H. J., Frederiksen, W. C. & Bisgaard, M. (2007). Proposed minimal standards for the description of genera, species and subspecies of the *Pasteurellaceae*. *Int J Syst Evol Microbiol* **57**, 166–178.
- Clarridge, J. E., III (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* **17**, 840–862.
- Coenye, T. & Vandamme, P. (2004). Use of the genomic signature in bacterial classification and identification. *Syst Appl Microbiol* **27**, 175–185.
- Crosa, J. H., Brenner, D. J., Ewing, W. H. & Falkow, S. (1973). Molecular relationships among the *Salmonellae*. *J Bacteriol* **115**, 307–315.

- Cutler, S. J., Akintunde, C. O., Moss, J., Fukunaga, M., Kurtenbach, K., Talbert, A., Zhang, H., Wright, D. J. & Warrell, D. A. (1999). Successful *in vitro* cultivation of *Borrelia duttonii* and its comparison with *Borrelia recurrentis*. *Int J Syst Bacteriol* **49**, 1793–1799.
- De Ley, J., Cattoir, H. & Reynaerts, A. (1970). The quantitative measurement of DNA hybridization from renaturation rates. *Eur J Biochem* **12**, 133–142.
- Devriese, L. A., Riegel, P., Hommez, J., Vaneechoutte, M., de Baere, T. & Haesebrouck, F. (2000). Identification of *Corynebacterium glucuronolyticum* strains from the urogenital tract of humans and pigs. *J Clin Microbiol* **38**, 4657–4659.
- Drancourt, M. & Raoult, D. (1999). Characterization of mutations in the *rpoB* gene in naturally rifampin-resistant *Rickettsia* species. *Antimicrob Agents Chemother* **43**, 2400–2403.
- Ezaki, T., Hashimoto, Y. & Yabuuchi, E. (1989). Fluorometric deoxyribonucleic acid–deoxyribonucleic acid hybridization in microdilution wells as an alternative to membrane filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. *Int J Syst Bacteriol* **39**, 224–229.
- Foster, J., Ganatra, M., Kamal, I., Ware, J., Makarova, K., Ivanova, N., Bhattacharyya, A., Kapatral, V., Kumar, S. & other authors (2005). The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol* **3**, e121.
- Fournier, P. E., Suhre, K., Fournous, G. & Raoult, D. (2006). Estimation of prokaryote genomic DNA G + C content by sequencing universally conserved genes. *Int J Syst Evol Microbiol* **56**, 1025–1029.
- Fox, G. E., Wisotzkey, J. D. & Jurtshuk, P., Jr (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* **42**, 166–170.
- Gardan, L., Shafik, H., Belouin, S., Broch, R., Grimont, F. & Grimont, P. A. D. (1999). DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremiae* sp. nov. and *Pseudomonas cannabina* sp. nov. (ex Sutic and Dowson 1959). *Int J Syst Bacteriol* **49**, 469–478.
- Gasparich, G. E., Whitcomb, R. F., Dodge, D., French, F. E., Glass, J. & Williamson, D. L. (2004). The genus *Spiroplasma* and its non-helical descendants: phylogenetic classification, correlation with phenotype and roots of the *Mycoplasma mycoides* clade. *Int J Syst Evol Microbiol* **54**, 893–918.
- Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., Stackebrandt, E., Van de Peer, Y., Vandamme, P. & other authors (2005). Opinion: re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**, 733–739.
- Godoy, D., Randle, G., Simpson, A. J., Aanensen, D. M., Pitt, T. L., Kinoshita, R. & Spratt, B. G. (2003). Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderia mallei*. *J Clin Microbiol* **41**, 2068–2079.
- Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P. & Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**, 81–91.
- Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.
- Khamis, A., Colson, P., Raoult, D. & La Scola, B. (2003). Usefulness of *rpoB* gene sequencing for identification of *Afipia* and *Bosea* species, including a strategy for choosing discriminative partial sequences. *Appl Environ Microbiol* **69**, 6740–6749.
- Khamis, A., Raoult, D. & La Scola, B. (2004). *rpoB* gene sequencing for identification of *Corynebacterium* species. *J Clin Microbiol* **42**, 3925–3931.
- Kim, K. S., Ko, K. S., Chang, M. W., Hahn, T. W., Hong, S. K. & Kook, Y. H. (2003). Use of *rpoB* sequences for phylogenetic study of *Mycoplasma* species. *FEMS Microbiol Lett* **226**, 299–305.
- Konstantinidis, K. T. & Tiedje, J. M. (2005a). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102**, 2567–2572.
- Konstantinidis, K. T. & Tiedje, J. M. (2005b). Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* **187**, 6258–6264.
- Koonin, E. V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol* **1**, 127–136.
- Korczak, B. M., Stieber, R., Emler, S., Burnens, A. P., Frey, J. & Kuhnert, P. (2006). Genetic relatedness within the genus *Campylobacter* inferred from *rpoB* sequences. *Int J Syst Evol Microbiol* **56**, 937–945.
- Koski, L. B. & Golding, G. B. (2001). The closest BLAST hit is often not the nearest neighbor. *J Mol Evol* **52**, 540–542.
- Kuhnert, P. & Korczak, B. M. (2006). Prediction of whole-genome DNA–DNA similarity, determination of G + C content and phylogenetic analysis within the family *Pasteurellaceae* by multilocus sequence analysis (MLSA). *Microbiology* **152**, 2537–2548.
- Kunin, V., Ahren, D., Goldovsky, L., Janssen, P. & Ouzounis, C. A. (2005). Measuring genome conservation across taxa: divided strains and united kingdoms. *Nucleic Acids Res* **33**, 616–621.
- Lan, R. & Reeves, P. R. (2001). When does a clone deserve a name? A perspective on bacterial species based on population genetics. *Trends Microbiol* **9**, 419–424.
- La Scola, B., Gundi, V. A., Khamis, A. & Raoult, D. (2006). Sequencing of the *rpoB* gene and flanking spacers for molecular identification of *Acinetobacter* species. *J Clin Microbiol* **44**, 827–832.
- Lo, N., Paraskevopoulos, C., Bourtzis, K., O'Neill, S. L., Werren, J. H., Bordenstein, S. R. & Bandi, C. (2007). Taxonomic status of the intracellular bacterium *Wolbachia pipientis*. *Int J Syst Evol Microbiol* **57**, 654–657.
- Marianelli, C., Ciuchini, F., Tarantino, M., Pasquali, P. & Adone, R. (2006). Molecular characterization of the *rpoB* gene in *Brucella* species: new potential molecular markers for genotyping. *Microbes Infect* **8**, 860–865.
- Masuzawa, T., Takada, N., Kudaken, M., Fukui, T., Yano, Y., Ishiguro, F., Kawamura, Y., Imai, Y. & Ezaki, T. (2001). *Borrelia sinica* sp. nov., a Lyme disease-related *Borrelia* species isolated in China. *Int J Syst Evol Microbiol* **51**, 1817–1824.
- Mollet, C., Drancourt, M. & Raoult, D. (1998). Determination of *Coxiella burnetii* *rpoB* sequence and its use for phylogenetic analysis. *Gene* **207**, 97–103.
- Munson, M. A., Baumann, P. & Kinsey, M. G. (1991). *Buchnera* gen. nov. and *Buchnera aphidicola* sp. nov., a taxon consisting of the mycetocyte-associated, primary endosymbionts of aphids. *Int J Syst Bacteriol* **41**, 566–568.
- Rappé, M. S. & Giovannoni, S. J. (2003). The uncultured microbial majority. *Annu Rev Microbiol* **57**, 369–394.
- Riesenfeld, C. S., Schloss, P. D. & Handelsman, J. (2004). Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* **38**, 525–552.
- Rogul, M., Brendle, J. J., Haapala, D. K. & Alexander, A. D. (1970). Nucleic acid similarities among *Pseudomonas pseudomallei*, *Pseudomonas multivorans*, and *Actinobacillus mallei*. *J Bacteriol* **101**, 827–835.
- Rosselló-Mora, R. & Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiol Rev* **25**, 39–67.
- Schinsky, M. F., Morey, R. E., Steigerwalt, A. G., Douglas, M. P., Wilson, R. W., Floyd, M. M., Butler, W. R., Daneshvar, M. I.,

- Brown-Elliott, B. A. & other authors (2004).** Taxonomic variation in the *Mycobacterium fortuitum* third biovariant complex: description of *Mycobacterium boenickei* sp. nov., *Mycobacterium houstonense* sp. nov., *Mycobacterium neworleansense* sp. nov. and *Mycobacterium brisbanense* sp. nov. and recognition of *Mycobacterium porcinum* from human clinical isolates. *Int J Syst Evol Microbiol* **54**, 1653–1667.
- Schloss, P. D. & Handelsman, J. (2004).** Status of the microbial census. *Microbiol Mol Biol Rev* **68**, 686–691.
- Sreevatsan, S., Pan, X., Stockbauer, K. E., Connell, N. D., Kreiswirth, B. N., Whittam, T. S. & Musser, J. M. (1997).** Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A* **94**, 9869–9874.
- Stackebrandt, E. & Ebers, J. (2006).** Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* **4**, 152–155.
- Stackebrandt, E. & Goebel, B. M. (1994).** Taxonomic note: 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.
- Stinear, T. P., Seemann, T., Pidot, S., Frigui, W., Reyssset, G., Garnier, T., Meurice, G., Simon, D., Bouchier, C. & other authors (2007).** Reductive evolution and niche adaptation inferred from the genome of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. *Genome Res* **17**, 192–200.
- Taillardat-Bisch, A.-V., Raoult, D. & Drancourt, M. (2003).** RNA polymerase β -subunit-based phylogeny of *Ehrlichia* spp., *Anaplasma* spp., *Neorickettsia* spp. and *Wolbachia pipientis*. *Int J Syst Evol Microbiol* **53**, 455–458.
- Tourova, T. P. & Antonov, A. S. (1987).** Identification of microorganisms by rapid DNA-DNA hybridization. *Methods Microbiol* **19**, 333–355.
- Ursing, J. B., Rosselló-Mora, R. A., Garcia-Valdes, E. & Lalucat, J. (1995).** Taxonomic note: a pragmatic approach to the nomenclature of phenotypically similar genomic groups. *Int J Syst Bacteriol* **45**, 604.
- van Passel, M. W., Bart, A. & van der Ende, A. (2005).** Default taxonomy and the genomics era. *Microbiology* **151**, 2818–2820.
- Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K. & Swings, J. (1996).** Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* **60**, 407–438.
- Verger, J. M., Grimont, F., Grimont, P. A. D. & Grayon, M. (1985).** *Brucella*, a monospecific genus as shown by deoxyribonucleic acid hybridization. *Int J Syst Bacteriol* **35**, 292–295.
- Verger, J. M., Grimont, F., Grimont, P. A. D. & Grayon, M. (1987).** Taxonomy of the genus *Brucella*. *Ann Inst Pasteur Microbiol* **138**, 235–238.
- Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E. & other authors (1987).** International Committee on Systematic Bacteriology. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37**, 463–464.
- Wu, M., Sun, L. V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J. C., McGraw, E. A., Martin, W., Esser, C. & other authors (2004).** Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* **2**, E69.
- Yabuuchi, E., Kawamura, Y., Ezaki, T., Ikeda, M., Dejsirilert, S., Fujiwara, N., Naka, T. & Kobayashi, K. (2000).** *Burkholderia uboniae* sp. nov., L-arabinose-assimilating but different from *Burkholderia thailandensis* and *Burkholderia vietnamiensis*. *Microbiol Immunol* **44**, 307–317.
- Yamamoto, S., Bouvet, P. J. & Harayama, S. (1999).** Phylogenetic structures of the genus *Acinetobacter* based on *gyrB* sequences: comparison with the grouping by DNA–DNA hybridization. *Int J Syst Bacteriol* **49**, 87–95.
- Yip, M. J., Porter, J. L., Fyfe, J. A., Lavender, C. J., Portaels, F., Rhodes, M., Kator, H., Colorni, A., Jenkin, G. A. & Stinear, T. (2007).** Evolution of *Mycobacterium ulcerans* and other mycolactone-producing mycobacteria from a common *Mycobacterium marinum* progenitor. *J Bacteriol* **189**, 2021–2029.