dnaJ is a useful phylogenetic marker for alphaproteobacteria

Ana Alexandre,¹ Marta Laranjo,¹ J. Peter W. Young² and Solange Oliveira¹

Correspondence
Solange Oliveira
ismo@uevora.pt

¹Laboratório de Microbiologia do Solo – Instituto de Ciências Agrárias Mediterrânicas (ICAM), Universidade de Évora, 7002-554 Évora, Portugal

²Department of Biology, University of York, York, UK

In the past, bacterial phylogeny relied almost exclusively on 16S rRNA gene sequence analysis. More recently, multilocus sequence analysis has been used to infer organismal phylogenies. In this study, the *dnaJ* chaperone gene was investigated as a marker for phylogeny studies in alphaproteobacteria. Preliminary analysis of G + C contents and G + C3s contents (the G + C content of the synonymous third codon position) showed no clear evidence of horizontal transfer of this gene in proteobacteria. *dnaJ*-based phylogenies were then analysed at three taxonomic levels: the *Proteobacteria*, the *Alphaproteobacteria* and the genus *Mesorhizobium*. Dendrograms based on DnaJ and 16S rRNA gene sequences revealed the same topology described previously for the *Proteobacteria*. These results indicate that the DnaJ phylogenetic signal is able to reproduce the accepted relationships among the five classes of the *Proteobacteria*. At a lower taxonomic level, using 20 alphaproteobacteria, the 16S rRNA gene-based phylogeny is distinct from the one based on DnaJ sequence analysis. Although the same clusters are generated, only the topology of the DnaJ tree is consistent with broader phylogenies from recent studies based on concatenated alignments of multiple core genes. For example, the DnaJ tree shows the two clusters within the *Rhizobiales* as closely related, as expected, while the 16S rRNA gene-based phylogeny shows them as distantly related. In order to evaluate the phylogenetic performance of *dnaJ* at the genus level, a multilocus analysis based on five housekeeping genes (*atpD*, *gapA*, *gyrB*, *recA* and *rplB*) was performed for ten *Mesorhizobium* species. In contrast to the 16S rRNA gene, the DnaJ sequence analysis generated a tree similar to the multilocus dendrogram. For identification of chickpea mesorhizobium isolates, a *dnaJ* nucleotide sequence-based tree was used. Despite different topologies, 16S rRNA gene- and *dnaJ*-based trees led to the same species identification. This study suggests that the *dnaJ* gene is a good phylogenetic marker, particularly for the class *Alphaproteobacteria*, since its phylogeny is consistent with phylogenies based on multilocus approaches.

INTRODUCTION

The phylum *Proteobacteria* is composed of five classes, the *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, *Deltaproteobacteria* and *Epsilonproteobacteria* (Garrity *et al.*, 2005; Stackebrandt *et al.*, 1988). By the end of 2008, more than 380 complete genomes of proteobacteria were available in public databases (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>). Alphaproteobacteria exhibit an enormous diversity in their morphological and

metabolic characteristics and the class is presently recognized solely as a clade in the 16S rRNA gene-based phylogeny (Stackebrandt *et al.*, 1988). Based on 16S rRNA gene trees, the *Alphaproteobacteria* has been divided into seven orders: *Caulobacteriales*, *Rhizobiales*, *Rhodobacterales*, *Rhodospirillales*, *Rickettsiales*, *Sphingomonadales* and *Parvularculales* (Kerstens *et al.*, 2006). The *Alphaproteobacteria* includes important bacteria that are widely studied, including, for example, the most important genera of soil bacteria able to live in symbiosis with leguminous plants (order *Rhizobiales*). This interaction between plants and rhizobia leads to the formation of nodules, special structures on plant roots where biological nitrogen fixation takes place, an important process in global nutrient cycling.

Evolutionary relationships among bacteria have been estimated by 16S rRNA gene sequence comparisons, since

Abbreviation: HGT, horizontal gene transfer.

The GenBank/EMBL/DDBJ accession numbers for the *atpD*, *dnaJ*, *gapA*, *gyrB*, *recA*, *rplB* and 16S rRNA gene sequences determined in this study are given in bold in Table 1.

Details of PCR primers and conditions, graphical representations of G + C content analysis and a *dnaJ*-based ML tree for *Mesorhizobium* are available as supplementary material with the online version of this paper.

this was originally considered to reflect, more or less, organismal phylogeny (Olsen & Woese, 1993). More recently, it has been shown that rRNA genes may show sequence heterogeneity and undergo horizontal transfer and genetic recombination (Acinas *et al.*, 2004). Furthermore, because of its very high sequence conservation, the 16S rRNA gene has limited usefulness in resolving closely related species. Another disadvantage of this gene is that most bacteria harbour several copies of the 16S rRNA gene and, in some cases, different copies have different evolutionary histories. For example, the 16S rRNA gene copies of *Escherichia coli* are divergent (Cilia *et al.*, 1996). For that reason, the search for other genes able to tell an evolutionary story for bacterial species has led to the use of several housekeeping genes, including *recA* (Eisen, 1995), *gyrB* (Yamamoto & Harayama, 1995), *rpoD* (Yamamoto *et al.*, 2000), *atpD* (Gaunt *et al.*, 2001) and *glnA* (Turner & Young, 2000), as alternative phylogenetic markers. The 16S rRNA gene phylogeny is not always in agreement with phylogenies based on housekeeping loci (Xiao *et al.*, 2007). In most cases, 16S rRNA gene and housekeeping gene phylogenies are similar, but phylogenies generated from protein-coding gene sequences frequently show higher resolution and reliability than the corresponding 16S rRNA gene-based phylogeny (Thompson *et al.*, 2004).

Recent phylogenetic studies based on multiple protein sequences from completely sequenced genomes are believed to reveal the 'true' evolutionary history of bacteria. For example, Fukami-Kobayashi *et al.* (2007) showed a tree of life comprising 167 organisms belonging to the *Archaea*, *Bacteria* and *Eukarya*, based on domain organization of all proteins encoded in each species genome; Ciccarelli *et al.* (2006) showed a global phylogeny comprising 191 species (*Archaea*, *Bacteria* and *Eukarya*) based on 31 universal proteins. Gupta & Sneath (2007) studied just proteobacteria, using the amino acid sequences of 10 protein-coding genes. A larger number of proteins (648) was used by Young *et al.* (2006) to generate a phylogeny that shows the relationship of *Rhizobium leguminosarum* and its close relatives with completely sequenced genomes. These studies indicate that a large amount of core gene sequence information produces strongly supported and more resolved phylogenies, frequently in disagreement with 16S rRNA gene-based phylogenies. Single gene-based phylogenies may now be validated by comparison with presumed organismal phylogeny obtained from multigenic approaches.

Genes coding for molecular chaperones have been used as alternative phylogenetic markers. The most commonly used are *groEL* (Baldo *et al.*, 2006) and *dnaK* (Vitorino *et al.*, 2007). Chaperones are known for their role as folding modulators, in sequestering and stabilizing a wide range of polypeptides when the wrong conformational structure is presented (Frydman, 2001). Under standard growth conditions, most of these proteins are expressed constitutively, because they are essential in the folding of nascent

polypeptide chains. The bacterial DnaK/J system is a widely studied example of such chaperones.

The DnaJ protein has several domains: the highly conserved J domain (N-terminal), the G/F-rich domain, a conserved region of four repeats with the consensus sequence CxxCxGxG, and a non-conserved C-terminal region of variable length (for review see Siegenthaler *et al.*, 2004). In most genomes, *dnaK* is a single copy gene, often found in the same operon as *dnaJ*. In *Escherichia coli* (*Gammaproteobacteria*), the *dnaK* operon is bicistronic, comprising the *dnaK* (1917 bp) and *dnaJ* (1131 bp) genes (Saito & Uchida, 1978). The *dnaJ* gene is sometimes found in more than one copy, as in the high-G+C-content Gram-positive actinobacteria. It has been reported that these different copies may have different physiological functions and distinct evolutionary paths (Ventura *et al.*, 2005).

Its high degree of sequence conservation, functional preservation and universal distribution among bacteria are key features that make the *dnaJ* gene suitable for inferring phylogenetic relationships. *dnaJ* has already been used successfully for discrimination of species and subspecies, even within genera in which the 16S rRNA gene has insufficient resolution, as in pathogenic bacteria such as *Mycobacterium* (Morita *et al.*, 2004), *Streptococcus* (Itoh *et al.*, 2006) and more recently *Staphylococcus* (Shah *et al.*, 2007) and *Vibrio* (Nhung *et al.*, 2007). To our knowledge, only one study has reported the use of *dnaJ* as a phylogenetic marker within the *Proteobacteria*, with the aim of discriminating *Legionella pneumophila* serogroups (Liu *et al.*, 2003).

Our aim was to find a single core gene with a phylogenetic signal that resembled multilocus phylogenies at higher taxonomic levels (*Proteobacteria* and *Alphaproteobacteria*) and with a good phylogenetic signal at a lower taxonomic level, namely for the genus *Mesorhizobium*. We analysed the phylogenetic signal of the *dnaJ* gene from 19 proteobacteria of all five classes and then from 20 alphaproteobacteria and compared it with the 16S rRNA gene-based phylogeny, as well as with phylogenies from other studies based on multiple proteins. We also used the *dnaJ* gene to infer the phylogeny of isolates and type strains within the genus *Mesorhizobium* (*Alphaproteobacteria*). The *dnaJ*-based phylogeny was, in this case, compared to that of five concatenated genes (*atpD*–*gapA*–*gyrB*–*recA*–*rplB*) and 16S rRNA gene-based phylogenies.

METHODS

Bacterial strains and growth conditions. Rhizobial isolates nodulating chickpea were obtained from two soil samples collected in Portugal using trap plants (Somasegaran & Hoben, 1994). Five isolates were selected for further study: C-1-Coimbra, C-27b-Coimbra, V-15b-Viscu, V-18-Viscu and V-20-Viscu. Four additional isolates obtained previously and identified by 16S rRNA gene analysis (Laranjo *et al.*, 2001, 2002, 2004) were used.

PCR amplifications and sequencing. Bacterial genomes frequently harbour more than one gene annotated as *dnaJ*. To prevent amplification of multiple copies in the same strain and of different copies in different species, a forward primer in the *dnaK* gene (see Supplementary Table S1, available in IJSEM Online) was used for PCR amplification (*dnaK* is a single-copy gene, with very few exceptions), guaranteeing that the target sequence is the *dnaJ* copy found in the *dnaK/J* operon. *dnaJ* amplification by PCR was performed for nine chickpea rhizobia isolates and ten *Mesorhizobium* type strains. Primers *dnaK*-F and *dnaJ*-R were used for PCR amplification. The 16S rRNA gene was amplified by PCR according to Laranjo *et al.* (2004). The *atpD* gene, which codes for ATP synthase subunit β , was amplified using primers *atpD*-F and *atpD*-R (Gaunt *et al.*, 2001). The *gapA* gene, which codes for glyceraldehyde-3-phosphate dehydrogenase, was amplified using primers *gap*-for and *gap*-rev. The *gyrB* gene, which codes for DNA gyrase subunit β , was amplified using primers *gyrB*-for_new and *gyrB*-rev_new. The *recA* gene, which codes for DNA recombinase A, was amplified using primers *recA*-F and *recA*-R (Gaunt *et al.*, 2001). The *rplB* gene, which codes for the 50S ribosomal protein L2 involved in translation, was amplified using primers L2-for and L2-rev. Primer details are given in Supplementary Table S1.

Total DNA was extracted as described previously (Rivas *et al.*, 2001). All PCR mixtures except that for amplification of the *recA* gene contained 1 U FidelityTaq DNA polymerase (USB), 1 $\times$ reaction buffer (with 1.5 mM $MgCl_2$), 0.2 mM of each dNTP (Invitrogen), 5 % DMSO (Duchefa), 15–25 pmol of each primer and 7–10 μ l DNA. *recA* PCRs were prepared with 2 U Taq DNA polymerase (Fermentas), 1 $\times$ reaction buffer, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP (Invitrogen), 0.004 % BSA (Promega), 25 pmol of each primer and 3 μ l DNA. Amplification conditions are outlined in Supplementary Table S1.

PCR products were purified using a GFX PCR DNA and gel band purification kit (GE Healthcare) or ExoSAP-It (USB) following the manufacturers' instructions. Sequencing reactions were performed by Macrogen (Korea). For 16S rRNA gene sequencing, two extra primers, IntF and IntR, were used as internal primers for double-stranded sequencing (Laranjo *et al.*, 2004). For *dnaJ* sequencing, a forward primer located in the 5'-end of the *dnaJ* gene was used (*dnaJ*-F; 5'-GCTGGGCGTGCAAAAGGG-3') together with the primer *dnaJ*-R. For the remaining genes, the sequencing primers were those used for PCR amplification.

Data analysis. Sequences generated in this study were edited using BioEdit 7.0.5.3 (Hall, 1999) and aligned using CLUSTAL W (Thompson *et al.*, 1994). The sequences were checked manually for correct alignment. Gene sequences from completely sequenced genomes were retrieved from the NCBI complete microbial genomes database.

In order to detect possible events of horizontal gene transfer (HGT) of the *dnaJ* gene, the G+C content and codon usage frequency were analysed for all proteobacteria included in the study (Eisen, 2000). G+C content data for the chromosome were retrieved from the NCBI database. G+C contents of individual genes were calculated using BioEdit 7.0.5.3 (Hall, 1999). In addition, the G+C content of the synonymous third codon position (G+C3s) was also calculated for the chromosome and the *dnaJ* gene using CodonW (Peden, 2000). Codon usage frequency analysis of the *dnaJ* gene was performed with Graphical Codon Usage Analyser 2.0 (<http://gcua.schoedl.de/>), using the codon usage table of each species available at the Codon Usage Database (<http://www.kazusa.or.jp/codon/>).

Molecular phylogeny was reconstructed using PHYLIP version 3.67 (Felsenstein, 2006) by the maximum-likelihood (ML) method. Neighbour-joining (NJ) (Saitou & Nei, 1987) phylogenies were generated using MEGA4 (Tamura *et al.*, 2007). Evolutionary distances

were calculated by Kimura's two-parameter model for nucleotide sequence alignments (Kimura, 1980) whereas, for amino acid sequence alignments, the Poisson correction (Nei & Kumar, 2000) was applied.

For a multilocus approach to the mesorhizobia phylogeny, five protein-coding genes were chosen: *atpD*, *gapA*, *gyrB*, *recA* and *rplB*. The incongruence length difference (ILD) test (Bull *et al.*, 1993; Cunningham, 1997) was performed to check whether trees for the different genes were sufficiently similar to allow data combination. The ILD test was performed using PAUP*4.0 (Swofford, 2003).

RESULTS AND DISCUSSION

In the phylogenetic studies at higher taxonomic levels, sequences from 19 proteobacteria representing all five classes of the *Proteobacteria* were used. For further phylogenetic analysis of alphaproteobacteria, 20 strains, from all orders except the *Parvularculales*, were selected. Finally, at a lower taxonomic level, ten type strains and nine isolates from the genus *Mesorhizobium* were used. Bacterial strains used in this study are listed in Table 1.

In all proteobacteria used in the present study, the *dnaJ* gene was found to be encoded adjacent to *dnaK*, except in epsilonproteobacteria, for which no *dnaK/J* operon was found among the 11 completely sequenced genomes. In the chromosome of each of the proteobacteria used, only one *dnaJ* gene copy was found that had the expected size and included the four characteristic domains. Other orthologues annotated as *dnaJ*, often dispersed in the chromosome, were found to lack at least one of the characteristic *dnaJ* domains. In two species (*Ensifer meliloti* and *Rhizobium etli*), a *dnaJ* homologue was found to be encoded on a plasmid; however, these homologues were partial *dnaJ* sequences.

Prior to the use of *dnaJ* as a phylogenetic marker, analysis of G+C content and codon usage frequency was performed, in order to detect possible HGT events that would compromise the phylogenetic analysis. The G+C content of the *dnaJ* gene was compared with the G+C content of the entire chromosome for the alphaproteobacteria (see Supplementary Fig. S1a, available in IJSEM Online) and other proteobacteria used in this study (data not shown). *dnaJ* G+C contents are similar to those of the chromosome, while the 16S rRNA gene G+C content shows little variation across species (ranging from 50 to 57 mol%), regardless of the G+C content of the chromosome (ranging from 29 to 69 mol%). For example, in *Rickettsia prowazekii*, the 16S rRNA gene G+C content (51 mol%) is clearly higher than that of the chromosome (29 mol%), whereas the G+C content of the *dnaJ* gene (36 mol%) is closer to the chromosomal value. Core genes, such as *recA* or *dnaK*, present a G+C content similar to that of the whole chromosome. No evidence of HGT events was provided by this analysis for the alphaproteobacteria or the other proteobacteria used (data not shown). Nevertheless, the possibility of *dnaJ* HGT occurring between organisms with similar G+C contents cannot be excluded.

Table 1. Strains of the *Proteobacteria* with corresponding accession numbers or gene IDs of sequences used in this study

Accession numbers for sequences resulting from this study are shown in bold.

Strain	Genome	16S rRNA gene	<i>dnaJ</i>	<i>dnaK</i>	<i>atpD</i>	<i>gapA</i>	<i>gyrB</i>	<i>recA</i>	<i>rplB</i>
<i>Alphaproteobacteria</i>									
<i>Agrobacterium tumefaciens</i> C58	NC_003062	3244993	1137637	1137638					
<i>Bradyrhizobium japonicum</i> USDA 110	NC_004463	1055154	1049300	1049096	1054445	1055273	1047796	1048547	1051246
<i>Brucella suis</i> 1330 ^T	NC_004310	1167560	1167829	1167828					
<i>Caulobacter crescentus</i> CB15	NC_002696	943247	944076	944075					
<i>Ensifer meliloti</i> 1021	NC_003047	1234653	1231814	1234653	1234718	1234439	1231637	1233464	1233011
<i>Erythrobacter litoralis</i> HTCC2594	NC_007722	3868981	3870763	3870762					
<i>Gluconobacter oxydans</i> 621H	NC_006677	3248602	3250315	3250314					
<i>Granulibacter bethesdensis</i> CGDNIH1 ^T	NC_008343	4276889	4274426	4274427					
<i>Maricaulis maris</i> MCS10	NC_008347	4285688	4283983	4283981					
Genus <i>Mesorhizobium</i>									
<i>M. amorphae</i> ACCC 19665 ^T		AF041442	EF504296		AY493453	AM072544	AM076341	AY494816	AM076350
<i>M. chacoense</i> LMG 19008 ^T		AJ278249	EU273806		AY493460	AM072546	AM076343	AY494825	AM076352
<i>M. ciceri</i> UPM-Ca7 ^T		DQ444456	EF504297		AJ294395	AM072545	AM076342	AJ294367	AM076351
<i>M. ciceri</i> 27-Beja			EF504304						
<i>M. huakuii</i> CCBAU 2609 ^T		D12797	EF504298		AJ294394	AM072547	AM076344	AJ294370	AM076353
<i>M. loti</i> LMG 6125 ^T		X67229	EU053202		EU039868	EU273807	EU273810	EU039875	EU273813
<i>M. loti</i> MAFF 303099	NC_002678	3205748	1228214	1228215					
<i>M. loti</i> 75-Elvas			EF504306						
<i>M. mediterraneum</i> UPM-Ca36 ^T		L38825	EF504299		AM418768	AM072549	AM076346	AJ294369	AM076355
<i>M. mediterraneum</i> 29-Beja			EF504305						
<i>M. plurifarium</i> ORS 1032 ^T		Y14158	EF504300		AM076366	AM072550	AM076347	AY494824	AM076356
<i>M. septentrionale</i> HAMBI 2582 ^T		AF508207	EF504301		DQ659498	EU273808	EU273811	DQ444304	EU273814
<i>M. temperatum</i> HAMBI 2583 ^T		AF508208	EF504302		DQ659499	EU273809	EU273812	DQ444305	EU273815
<i>M. tianshanense</i> A-1B5 ^T		AF041447	EF504303		AM076367	AM072551	AM076348	AJ294368	AM076357
<i>M. tianshanense</i> 93-Évora			EF504307						
<i>Mesorhizobium</i> sp. C-1-Coimbra			EF504308						
<i>Mesorhizobium</i> sp. C-27b-Coimbra			EF504309						
<i>Mesorhizobium</i> sp. V-15b-Viseu			EF504310						
<i>Mesorhizobium</i> sp. V-18-Viseu			EF504311						
<i>Mesorhizobium</i> sp. V-20-Viseu			EF504312						
<i>Nitrobacter winogradskyi</i> Nb-255 ^T	NC_007406	3676957	3676654	3676653					
<i>Novosphingobium aromaticivorans</i> DSM 12444 ^T	NC_007794	3917048	3917700	3917701					
<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> 3841	NC_008380	4403801	4402721	4402722					
<i>Rhizobium etli</i> CFN42 ^T	NC_007761	3893174	3890921	3890922	AJ294404	3891389	3892150	AJ294375	3892587
<i>Rhodobacter sphaeroides</i> 2.4.1 ^T	NC_007493	3718805	3718165	3718166					
<i>Rhodopseudomonas palustris</i> CGA009	NC_005296	2690886	2691660	2689800					
<i>Rhodospirillum rubrum</i> ATCC 11170 ^T	NC_007643	3833695	3835440	3837010					
<i>Rickettsia prowazekii</i> Madrid E	NC_000963	883924	883882	883879					

Table 1. cont.

Strain	Genome	16S rRNA gene	<i>dnaJ</i>	<i>dnaK</i>	<i>atpD</i>	<i>gapA</i>	<i>gyrB</i>	<i>recA</i>	<i>rplB</i>
<i>Roseobacter denitrificans</i> Och 114 ^T	NC_008209	4198019	4198416	4198415					
<i>Silicibacter pomeroyi</i> DSS-3 ^T	NC_003911	3196347	3196511	3195579					
Betaproteobacteria									
<i>Azoarcus</i> sp. EbN1	NC_006513	3180819	3179956						
<i>Bordetella pertussis</i> Tohama I	NC_002929	3131280	2666530						
<i>Burkholderia ambifaria</i> AMMD ^T	NC_008390	4308934	4309294						
<i>Nitrospira multififormis</i> ATCC 25196 ^T	NC_007614	3784866	3784939						
<i>Ralstonia solanacearum</i> GMI1000	NC_003295	1220256	1221481						
Gammaproteobacteria									
<i>Aeromonas hydrophila</i> subsp. <i>hydrophila</i> ATCC 7966 ^T	NC_008570	4488614	4488918						
<i>Escherichia coli</i> K-12 MG1655	NC_000913	948332	944753						
<i>Legionella pneumophila</i> Paris	NC_006368	3116667	3118141						
<i>Pseudomonas aeruginosa</i> PAO1	NC_002516	3240252	881760						
<i>Vibrio cholerae</i> O1 bv. <i>eltor</i> N16961	NC_002505	2614447	2614523						
Deltaproteobacteria									
<i>Geobacter sulfurreducens</i> PCA ^T	NC_002939	2685615	2688568						
<i>Pelobacter carbinolicus</i> DSM 2380 ^T	NC_007498	3725044	3722901						
Epsilonproteobacteria									
<i>Campylobacter jejuni</i> RM1221	NC_003912	3232519	3231902						
<i>Helicobacter pylori</i> 26695	NC_000915	899682	898772						

In order to detect *dnaJ* recently acquired by HGT, the G + C3s content was also analysed (Sharp *et al.*, 2005). The G + C3s of the chromosome shows the average G + C3s of the chromosomal protein-coding genes. Highly expressed genes display higher codon biases than the average, as is the case for genes coding for some chaperones, such as *dnaK* (Karlin *et al.*, 2004). A gene with a G + C3s content significantly different from the average of the chromosome and from that of highly expressed genes is probably a recently acquired gene (Karlin *et al.*, 2004). In the alphaproteobacteria (Supplementary Fig. S1b), the *dnaJ* gene has, in most cases, a G + C3s content between that of the chromosome and that of *dnaK* (a highly expressed gene). The only two species that show a G + C3s for *dnaJ* clearly below the average of the chromosome genes are *Agrobacterium tumefaciens* and *Roseobacter denitrificans*. Therefore, *dnaJ* from these two species could have been acquired by HGT. For the remaining proteobacteria (data not shown), the G + C3s values for *dnaJ* are often very similar to those of the corresponding chromosome and, in some cases, similar to the value for *dnaK* (for example, in *Burkholderia ambifaria*), so no evidence was found for HGT of the *dnaJ* gene.

Another approach to look for evidence of HGT is to compare the codon usage table of each organism with the codon usage frequency of a given gene. If the *dnaJ* gene includes a large number of codons that are uncommonly used by the bacterium, it is likely that this gene was acquired by HGT. Codon usage analysis did not reveal any discrepancies that could clearly indicate HGT events (data not shown). Nonetheless, the *dnaJ* gene of *Agrobacterium tumefaciens* shows the highest percentage of uncommonly used codons. This discrepancy in codon usage together

with the G + C3s result raises doubts about the origin of the *dnaJ* gene in *Agrobacterium tumefaciens*, suggesting that it might have been acquired by HGT. Thus, the position of *Agrobacterium tumefaciens* in the *dnaJ*-based phylogeny should be regarded with caution.

In general, our analysis of G + C and G + C3s contents suggests that *dnaJ* is a core gene and is not commonly subject to HGT between species. These features are important in a good phylogenetic marker, so we used it for phylogenetic analysis of members of the *Proteobacteria*, *Alphaproteobacteria* and *Mesorhizobium*.

Phylogenetic analyses were performed using both the ML and NJ methods. In general, the two methods generated trees with the same topology.

The *Proteobacteria*

The *Proteobacteria* were represented in this study by 19 strains with completely sequenced genomes. The topology of the proteobacterial maximum-likelihood tree obtained with DnaJ amino acid sequences (Fig. 1a) is identical to that based on 16S rRNA gene sequences (Fig. 1b). The five classes of the *Proteobacteria* are well defined and their branching order is the same in both trees: the *Betaproteobacteria* is closer to the *Gammaproteobacteria*, and these two classes group first with the *Alphaproteobacteria*, then with the *Deltaproteobacteria* and finally with the *Epsilonproteobacteria*.

The branching order of all classes of the *Proteobacteria*, in both DnaJ and 16S rRNA gene phylogenies (Fig. 1), is in agreement with several recent studies that combine data from a large number of protein-coding genes, although

Fig. 1. Phylogeny of 19 members of the *Proteobacteria* from all five classes (indicated by Greek letters) based on analysis of the DnaJ amino acid sequence (a) and the 16S rRNA gene sequence (b). Trees were generated by ML. The first bootstrap percentage indicated on internal branches corresponds to the ML method (100 replicates) and the second to the NJ method (1000 replicates); dots indicate that nodes were not resolved using that method. Bars, 0.1 substitutions per site (ML).

some different species were used (Ciccarelli *et al.*, 2006; Fukami-Kobayashi *et al.*, 2007; Gupta & Sneath, 2007). It is noteworthy that some of these broader phylogenies showed that the *Deltaproteobacteria* and *Epsilonproteobacteria* may form one group, apart from the remaining classes of the *Proteobacteria* (Ciccarelli *et al.*, 2006; Fukami-Kobayashi *et al.*, 2007; Gupta, 2000).

Although the same branching order of classes of the *Proteobacteria* is generated in both phylogenies, the relationships between species within each cluster are different in a few cases. For example, relationships among the gammaproteobacteria *Aeromonas hydrophila*, *Vibrio cholerae* and *Escherichia coli* in the 16S rRNA gene-based tree are consistent with the phylogeny shown in a previous study using 31 concatenated protein sequences and comprising a large group of gammaproteobacteria (Seshadri *et al.*, 2006), but the DnaJ tree shows a different relationship.

Despite such minor discrepancies, the global congruence found between the 16S rRNA gene and the DnaJ phylogeny and the agreement with other phylogenies based on multilocus data indicate a good performance of *dnaJ* in reconstructing proteobacteria phylogeny.

The Alphaproteobacteria

To study the phylogenetic signal of the *dnaJ* gene at a lower taxonomic level, 20 strains belonging to the *Alphaproteobacteria* with completely sequenced genomes were used. In contrast to the 16S rRNA gene phylogeny (Fig. 2b), the phylogeny based on DnaJ sequences (Fig. 2a) is in agreement with the currently accepted relationships among genera and species within this group of bacteria obtained in other studies based on large amounts of

sequence information (Ciccarelli *et al.*, 2006; Gupta, 2005; Gupta & Sneath, 2007).

Both the DnaJ and 16S rRNA gene phylogenies (Fig. 2) show *Rhodospirillum rubrum*, *Gluconobacter oxydans* and *Granulibacter bethesdensis* (cluster A) and *Rickettsia prowazekii* in a distant position from the other clusters (the letters denote the clusters generated in the DnaJ phylogeny). However, clusters B, C, D and E are differently related in the two phylogenies. In the DnaJ phylogeny, cluster B (*Agrobacterium tumefaciens*, *Brucella suis*, *Ensifer meliloti*, *Mesorhizobium loti*, *Rhizobium etli* and *Rhizobium leguminosarum*) is closely related to cluster C (*Bradyrhizobium japonicum*, *Nitrobacter winogradskyi* and *Rhodopseudomonas palustris*) and then these clusters group first with cluster D (*Caulobacter crescentus*, *Erythrobacter litoralis*, *Maricaulis maris* and *Novosphingobium aromaticivorans*) and finally with cluster E (*Rhodobacter sphaeroides*, *Roseobacter denitrificans* and *Silicibacter pomeroyi*). In contrast to the DnaJ-based phylogeny, the 16S rRNA gene phylogeny shows cluster B closer to cluster E, then these two clusters group with two species of cluster D and finally with cluster C. Furthermore, in the 16S rRNA gene phylogeny, cluster D from the DnaJ phylogeny is dissolved and *Maricaulis maris* groups with cluster E while *Caulobacter crescentus* groups with cluster C species.

The relationships among different species inferred from DnaJ sequence comparisons are consistent with those inferred in previous studies, namely the phylogeny based on concatenated sequences of 10 proteins (Gupta & Sneath, 2007) and the tree of life generated from the concatenated alignment of 31 universal protein families (Ciccarelli *et al.*, 2006). In the DnaJ-based phylogeny, all members of the *Rhizobiales* cluster together (branch B+C), thus *Bradyrhizobium japonicum* (cluster C) is closer to

Fig. 2. Phylogeny of 20 members of the *Alphaproteobacteria*, based on analysis of the DnaJ amino acid sequence (a) and the 16S rRNA gene sequence (b). Trees were generated by ML. Two species of the *Epsilonproteobacteria* were used as an outgroup. The first bootstrap percentage indicated on internal branches corresponds to the ML method (100 replicates) and the second to the NJ method (1000 replicates); dots indicate that nodes were not resolved using that method. Bar, 0.1 substitutions per site (ML). Letters denote clusters generated in the DnaJ phylogeny.

Mesorhizobium loti (cluster B) than to *Caulobacter crescentus* (*Caulobacterales* in cluster D), as reported before (Ciccarelli *et al.*, 2006; Gupta & Sneath, 2007). In contrast, the 16S rRNA gene-based phylogeny shows *Bradyrhizobium japonicum* (cluster C) close to *Caulobacter crescentus* (cluster D) and distant from *Mesorhizobium loti* (cluster B). Moreover, within cluster B, the *DnaJ*-based phylogeny shows *Ensifer meliloti* closer to *Agrobacterium tumefaciens* and *Mesorhizobium loti* closer to *Brucella suis*, which is in full agreement with the phylogeny proposed by Young *et al.* (2006) based on the concatenated sequences of 638 proteins (the present study uses *Brucella suis* rather than *Brucella melitensis*) and with the multilocus tree of life presented by Ciccarelli *et al.* (2006), and not concordant with the 16S rRNA gene phylogeny. Therefore, the possible HGT origin of *dnaJ* from *Agrobacterium tumefaciens* (cluster B) suggested by the G+C content analysis was not confirmed.

The coherence between our results using *dnaJ* and those from broader phylogenies using multiple protein-coding genes (Ciccarelli *et al.*, 2006; Gupta & Sneath, 2007) shows that *dnaJ* has a better phylogenetic signal than the 16S rRNA gene in reconstructing the phylogeny of alphaproteobacteria.

The genus *Mesorhizobium*

The high level of sequence conservation of the 16S rRNA gene represents a limitation on the use of this gene for closely related bacteria. In the genus *Mesorhizobium*, other genes have been used for phylogenetic purposes, such as *dnaK* (Stepkowski *et al.*, 2003), *atpD* and *recA* (Vinuesa *et al.*, 2005). The non-coding 16S–23S rRNA intergenic spacer (ITS) has also been used (Rivas *et al.*, 2007). Relationships between mesorhizobial species are different depending on the gene used. Due to this unclear positioning and to the lack of a multilocus analysis focusing on mesorhizobia, there is no generally accepted phylogeny of this genus.

In order to obtain a reliable phylogenetic tree for the genus *Mesorhizobium* that could be compared to the *dnaJ*- and 16S rRNA gene-based phylogenies, five core genes (*atpD*, *gapA*, *gyrB*, *recA* and *rplB*) from ten mesorhizobia type strains were partially sequenced. The ILD test was applied to find out which genes could be combined, and the results allow the concatenation of the amino acid sequences of all genes. The ML tree based on the concatenated alignment of the five genes (approx. 770 amino acids long), shown in Fig. 3(a), reveals the putative organismal phylogeny among *Mesorhizobium* species. Both the *DnaJ* tree (Fig. 3b) and the 16S rRNA gene tree (Fig. 3c) were compared with the hypothetical organismal phylogeny to evaluate the accuracy of the phylogenetic signal of each single gene.

The 16S rRNA gene tree shows a different topology from the multilocus tree. According to the multilocus analysis, *Mesorhizobium huakuii* groups with *Mesorhizobium ciceri*

Fig. 3. Phylogeny based on the concatenated amino acid sequence alignment of AtpD–GapA–GyrB–RecA–RplB (a), the *DnaJ* sequence (b) and the 16S rRNA gene sequence (c) for the genus *Mesorhizobium*. Trees were generated by ML. The first bootstrap percentage indicated on internal branches corresponds to the ML method (100 replicates) and the second to the NJ method (1000 replicates); dots indicate that nodes were not resolved using that method. Bars, 0.1 substitutions per site (ML).

and *Mesorhizobium loti* in the deeper branching of the dendrogram. In contrast to this, the 16S rRNA gene-based phylogeny shows *Mesorhizobium ciceri* and *Mesorhizobium*

loti apart from the remaining eight mesorhizobia type strains. The concatenated tree shows *Mesorhizobium chacoense* and *Mesorhizobium plurifarum* as the most divergent species. The only two groups generated in both trees are *Mesorhizobium mediterraneum*/*Mesorhizobium temperatum*/*Mesorhizobium tianshanense* and *Mesorhizobium amorphae*/*Mesorhizobium septentrionale*. Among the single-gene trees, the DnaJ (Fig. 3b) and GyrB (data not shown) trees show the topologies most consistent with that of the multilocus analysis, although they have some low bootstrap values. The similarity between the DnaJ-based tree and the putative *Mesorhizobium* phylogeny derived from the multilocus analysis suggests that this gene is not commonly subject to HGT between *Mesorhizobium* species.

The present multilocus phylogenetic analysis may contribute to the clarification of the phylogenetic relationships among *Mesorhizobium* species, namely the proximity of *Mesorhizobium ciceri*, *Mesorhizobium loti* and *Mesorhizobium huakuii*. Interestingly, analysis of phenotypic data from a large set of chickpea mesorhizobia isolates also supported a closer relationship between *Mesorhizobium loti*/*Mesorhizobium ciceri* and *Mesorhizobium huakuii* isolates (Alexandre *et al.*, 2006).

In order to evaluate the suitability of *dnaJ* sequences for identifying native chickpea rhizobia isolates, a *dnaJ*-based tree including nine isolates was generated (Supplementary Fig. S2). At this taxonomic level, nucleotide sequences were used, since the amino acid sequence-based tree showed low resolution and low bootstrap support (data not shown). Although the relative positioning of some type strains was different from the concatenated tree (Fig. 3a) and the 16S rRNA gene-based tree (Fig. 3c), identification of isolates based on *dnaJ* sequences is consistent with identification based on the 16S rRNA gene (Laranjo *et al.*, 2004). The present phylogenetic analysis of native rhizobia reveals a high diversity of species able to nodulate chickpea, namely isolates close to *Mesorhizobium ciceri*, *Mesorhizobium huakuii*, *Mesorhizobium loti*, *Mesorhizobium mediterraneum*, *Mesorhizobium temperatum* and *Mesorhizobium tianshanense*, confirming previous analyses of chickpea mesorhizobia (Laranjo *et al.*, 2004; Rivas *et al.*, 2007).

For decades, the 16S rRNA gene has been the most ubiquitous gene used for accessing bacterial phylogeny, but this has been changing. Many other core genes are now used as phylogenetic markers in order to infer species phylogeny and evolution. Bacterial *dnaJ* is a core gene, as confirmed by its chromosomal location and a G+C content similar to that of the chromosome. This gene is present as a single 'true' copy, usually in the same operon as *dnaK*, which allows its specific amplification by PCR. No evidence that *dnaJ* had been subject to a HGT event was found in the proteobacteria that were examined, and the phylogenetic signal of this gene was consistent with studies based on multiple protein sequences. In contrast to 16S rRNA gene trees, the alphaproteobacterial DnaJ phylogenetic tree showed a global topology that resembled the

presumed organismal phylogeny inferred from multiple protein sequences from complete genomes. For these reasons, we suggest that a single core gene, such as *dnaJ*, can be used as a phylogenetic marker for proteobacteria at the level of phylum (*Proteobacteria*) and class (*Alphaproteobacteria*). At the level of the genus *Mesorhizobium*, the *dnaJ* gene can be used for identification of isolates. This study highlights the usefulness of the *dnaJ* gene as a single alternative phylogenetic marker for alphaproteobacteria.

ACKNOWLEDGEMENTS

This work was supported by the Fundação para a Ciência e Tecnologia (FCT) (POCTI/BME/44140/2002) and co-financed by EU-FEDER. A. A. acknowledges a PhD fellowship (SFRH/BD/18162/2004) and M. L. acknowledges a post-doctoral fellowship (SFRH/BPD/27008/2006) from FCT. The contribution of J. P. W. Y. was funded by the Natural Environment Research Council, UK.

REFERENCES

- Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V. & Polz, M. F. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *J Bacteriol* **186**, 2629–2635.
- Alexandre, A., Laranjo, M. & Oliveira, S. (2006). Natural populations of chickpea rhizobia evaluated by antibiotic resistance profiles and molecular methods. *Microb Ecol* **51**, 128–136.
- Baldo, L., Bordenstein, S., Wernegreen, J. J. & Werren, J. H. (2006). Widespread recombination throughout *Wolbachia* genomes. *Mol Biol Evol* **23**, 437–449.
- Bull, J. J., Huelsenbeck, J. P., Cunningham, C. W., Swofford, D. L. & Waddell, P. J. (1993). Partitioning and combining data in phylogenetic analysis. *Syst Biol* **42**, 384–397.
- Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B. & Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283–1287.
- Cilia, V., Lafay, B. & Christen, R. (1996). Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Mol Biol Evol* **13**, 451–461.
- Cunningham, C. W. (1997). Can three incongruence tests predict when data should be combined? *Mol Biol Evol* **14**, 733–740.
- Eisen, J. A. (1995). The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. *J Mol Evol* **41**, 1105–1123.
- Eisen, J. A. (2000). Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. *Curr Opin Genet Dev* **10**, 606–611.
- Felsenstein, J. (2006). PHYLIP (phylogeny inference package). Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, USA.
- Frydman, J. (2001). Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu Rev Biochem* **70**, 603–647.
- Fukami-Kobayashi, K., Minezaki, Y., Tateno, Y. & Nishikawa, K. (2007). A tree of life based on protein domain organizations. *Mol Biol Evol* **24**, 1181–1189.
- Garrrity, G. M., Bell, J. A. & Lilburn, T. (2005). Phylum XIV. *Proteobacteria* phyl. nov. In *Bergey's Manual of Systematic*

- Bacteriology*, 2nd edn, vol. 2, part B, p. 1. Edited by D. J. Brenner, N. R. Krieg, J. T. Staley & G. M. Garrity. New York: Springer.
- Gaunt, M. W., Turner, S. L., Rigottier-Gois, L., Lloyd-Macgilp, S. A. & Young, J. P. W. (2001). Phylogenies of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia. *Int J Syst Evol Microbiol* **51**, 2037–2048.
- Gupta, R. S. (2000). The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol Rev* **24**, 367–402.
- Gupta, R. S. (2005). Protein signatures distinctive of alpha proteobacteria and its subgroups and a model for alpha-proteobacterial evolution. *Crit Rev Microbiol* **31**, 101–135.
- Gupta, R. S. & Sneath, P. H. A. (2007). Application of the character compatibility approach to generalized molecular sequence data: branching order of the proteobacterial subdivisions. *J Mol Evol* **64**, 90–100.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95–98.
- Itoh, Y., Kawamura, Y., Kasai, H., Shah, M. M., Nhung, P. H., Yamada, M., Sun, X., Koyana, T., Hayashi, M. & other authors (2006). *dnaJ* and *gyrB* gene sequence relationship among species and strains of genus *Streptococcus*. *Syst Appl Microbiol* **29**, 368–374.
- Karlin, S., Theriot, J. & Mrazek, J. (2004). Comparative analysis of gene expression among low G+C gram-positive genomes. *Proc Natl Acad Sci U S A* **101**, 6182–6187.
- Kerstens, K., De Vos, P., Gillis, M., Swings, J., Vandamme, P. & Stackebrandt, E. (2006). Introduction to the *Proteobacteria*. In *The Prokaryotes: a Handbook on the Biology of Bacteria*, 3rd edn, vol. 5, pp. 3–37. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer and E. Stackebrandt. New York: Springer.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120.
- Laranjo, M., Rodrigues, R., Alho, L. & Oliveira, S. (2001). Rhizobia of chickpea from southern Portugal: symbiotic efficiency and genetic diversity. *J Appl Microbiol* **90**, 662–667.
- Laranjo, M., Branco, C., Soares, R., Alho, L., Carvalho, M. & Oliveira, S. (2002). Comparison of chickpea rhizobia isolates from diverse Portuguese natural populations based on symbiotic effectiveness and DNA fingerprint. *J Appl Microbiol* **92**, 1043–1050.
- Laranjo, M., Machado, J., Young, J. P. W. & Oliveira, S. (2004). High diversity of chickpea *Mesorhizobium* species isolated in a Portuguese agricultural region. *FEMS Microbiol Ecol* **48**, 101–107.
- Liu, H., Li, Y., Huang, X., Kawamura, Y. & Ezaki, T. (2003). Use of the *dnaJ* gene for the detection and identification of all *Legionella pneumophila* serogroups and description of the primers used to detect 16S rDNA gene sequences of major members of the genus *Legionella*. *Microbiol Immunol* **47**, 859–869.
- Morita, Y., Maruyama, S., Kabeya, H., Nagai, A., Kozawa, K., Kato, M., Nakajima, T., Mikami, T., Katsube, Y. & Kimura, H. (2004). Genetic diversity of the *dnaJ* gene in the *Mycobacterium avium* complex. *J Med Microbiol* **53**, 813–817.
- Nei, M. & Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. New York: Oxford University Press.
- Nhung, P. H., Shah, M. M., Ohkusu, K., Noda, M., Hata, H., Sun, X. S., Iihara, H., Goto, K., Masaki, T. & other authors (2007). The *dnaJ* gene as a novel phylogenetic marker for identification of *Vibrio* species. *Syst Appl Microbiol* **30**, 309–315.
- Olsen, G. J. & Woese, C. R. (1993). Ribosomal RNA: a key to phylogeny. *FASEB J* **7**, 113–123.
- Peden, J. F. (2000). *Analysis of Codon Usage*. Nottingham: University of Nottingham.
- Rivas, R., Velázquez, E., Valverde, A., Mateos, P. F. & Martínez-Molina, E. (2001). A two primers random amplified polymorphic DNA procedure to obtain polymerase chain reaction fingerprints of bacterial species. *Electrophoresis* **22**, 1086–1089.
- Rivas, R., Laranjo, M., Mateos, P. F., Oliveira, S., Martínez-Molina, E. & Velázquez, E. (2007). Strains of *Mesorhizobium amorphae* and *Mesorhizobium tianshanense*, carrying symbiotic genes of common chickpea endosymbiotic species, constitute a novel biovar (*ciceri*) capable of nodulating *Cicer arietinum*. *Lett Appl Microbiol* **44**, 412–418.
- Saito, H. & Uchida, H. (1978). Organization and expression of DnaJ and DnaK genes of *Escherichia coli* K12. *Mol Gen Genet* **164**, 1–8.
- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.
- Seshadri, R., Joseph, S. W., Chopra, A. K., Sha, J., Shaw, J., Graf, J., Haft, D., Wu, M., Ren, Q. & other authors (2006). Genome sequence of *Aeromonas hydrophila* ATCC 7966^T: jack of all trades. *J Bacteriol* **188**, 8272–8282.
- Shah, M. M., Iihara, H., Noda, M., Song, S. X., Nhung, P. H., Ohkusu, K., Kawamura, Y. & Ezaki, T. (2007). *dnaJ* gene sequence-based assay for species identification and phylogenetic grouping in the genus *Staphylococcus*. *Int J Syst Evol Microbiol* **57**, 25–30.
- Sharp, P. M., Bailes, E., Grocock, R. J., Peden, J. F. & Sockett, R. E. (2005). Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res* **33**, 1141–1153.
- Siegenthaler, R. K., Grimshaw, J. P. A. & Christen, P. (2004). Immediate response of the DnaK molecular chaperone system to heat shock. *FEBS Lett* **562**, 105–110.
- Somasegaran, P. & Hoben, H. J. (1994). *Handbook for Rhizobia*. New York: Springer.
- Stackebrandt, E., Murray, R. G. E. & Trüper, H. G. (1988). *Proteobacteria* classis nov., a name for the phylogenetic taxon that includes the “purple bacteria and their relatives”. *Int J Syst Bacteriol* **38**, 321–325.
- Stepkowski, T., Czaplińska, M., Miedzinska, K. & Moulin, L. (2003). The variable part of the *dnaK* gene as an alternative marker for phylogenetic studies of rhizobia and related alpha Proteobacteria. *Syst Appl Microbiol* **26**, 483–494.
- Swofford, D. L. (2003). PAUP*: Phylogenetic analysis using parsimony (* and other methods). Sunderland, MA: Sinauer Associates.
- Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (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.
- Thompson, C. C., Thompson, F. L., Vandemeulebroecke, K., Hoste, B., Dawyndt, P. & Swings, J. (2004). Use of *recA* as an alternative phylogenetic marker in the family Vibrionaceae. *Int J Syst Evol Microbiol* **54**, 919–924.
- Turner, S. L. & Young, J. P. (2000). The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. *Mol Biol Evol* **17**, 309–319.
- Ventura, M., Canchaya, C., Bernini, V., Del Casale, A., Dellaglio, F., Neviani, E., Fitzgerald, G. F. & van Sinderen, D. (2005). Genetic characterization of the *Bifidobacterium breve* UCC 2003 *hrcA* locus. *Appl Environ Microbiol* **71**, 8998–9007.

- Vinuesa, P., Silva, C., Lorite, M. J., Izaguirre-Mayoral, M. L., Bedmar, E. J. & Martínez-Romero, E. (2005). Molecular systematics of rhizobia based on maximum likelihood and Bayesian phylogenies inferred from *rrs*, *atpD*, *recA* and *nifH* sequences, and their use in the classification of *Sesbania* microsymbionts from Venezuelan wetlands. *Syst Appl Microbiol* 28, 702–716.
- Vitorino, L., Chelo, I. M., Bacellar, F. & Zé-Zé, L. (2007). *Rickettsiae* phylogeny: a multigenic approach. *Microbiology* 153, 160–168.
- Xiao, X., Wang, P., Zeng, X., Bartlett, D. H. & Wang, F. P. (2007). *Shewanella psychrophila* sp nov and *Shewanella piezotolerans* sp nov., isolated from west Pacific deep-sea sediment. *Int J Syst Evol Microbiol* 57, 60–65.
- Yamamoto, S. & Harayama, S. (1995). PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. *Appl Environ Microbiol* 61, 1104–1109.
- Yamamoto, S., Kasai, H., Arnold, D. L., Jackson, R. W., Vivian, A. & Harayama, S. (2000). Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of *gyrB* and *rpoD* genes. *Microbiology* 146, 2385–2394.
- Young, J. P., Crossman, L., Johnston, A., Thomson, N. R., Ghazoui, Z. F., Hull, K. H., Wexler, M., Curson, A. R., Todd, J. D. & other authors (2006). The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol* 7, R34.