Available online at www.sciencedirect.com
ScienceDirect
Systematic and Applied Microbiology 30 (2007) 171–179
SYSTEMATIC AND APPLIED MICROBIOLOGY
www.elsevier.de/syapm

Phylogenetic position of *Salinibacter ruber* based on concatenated protein alignments

Victor Soria-Carrasco[b,1], Maria Valens-Vadell[a,1], Arantxa Peña[c], Josefa Antón[c], Rudolf Amann[d], Jose Castresana[b], Ramon Rosselló-Mora[a,*]

[a]*Institut Mediterrani d'Estudis Avançats (CSIC-UIB), E-07190 Esporles, Mallorca, Spain*
[b]*Department of Physiology and Molecular Biodiversity, Institute of Molecular Biology of Barcelona, CSIC, Jordi Girona 18, 08034 Barcelona, Spain*
[c]*División de Microbiología, Departamento de Fisiología, Genética y Microbiología, Universidad de Alicante, E-03080 Alicante, Spain*
[d]*Max Planck Institute for Marine Microbiology, D-28359 Bremen, Germany*

Received 11 July 2006

Abstract

A total of 22 genes from the genome of *Salinibacter ruber* strain M31 were selected in order to study the phylogenetic position of this species based on protein alignments. The selection of the genes was based on their essential function for the organism, dispersion within the genome, and sufficient informative length of the final alignment. For each gene, an individual phylogenetic analysis was performed and compared with the resulting tree based on the concatenation of the 22 genes, which rendered a single alignment of 10,757 homologous positions. In addition to the manually chosen genes, an automatically selected data set of 74 orthologous genes was used to reconstruct a tree based on 17,149 homologous positions. Although single genes supported different topologies, the tree topology of both concatenated data sets was shown to be identical to that previously observed based on small subunit (SSU) rRNA gene analysis, in which *S. ruber* was placed together with *Bacteroidetes*. In both concatenated data sets the bootstrap was very high, but an analysis with a gradually lower number of genes indicated that the bootstrap was greatly reduced with less than 12 genes. The results indicate that tree reconstructions based on concatenating large numbers of protein coding genes seem to produce tree topologies with similar resolution to that of the single 16S rRNA gene trees. For classification purposes, 16S rRNA gene analysis may remain as the most pragmatic approach to infer genealogic relationships.

Introduction

The central goal of taxonomy and systematics is the achievement of a classification system that accurately reflects natural relationships among living organisms. Although there are profound discrepancies about what may or may not be a 'natural' classification [33], it is unquestionable that one of the most revolutionary developments in prokaryotic taxonomy was the reconstruction of the first 16S rRNA gene phylogenies [12]. Since then and not without heated debates [33], genealogies based on sequence comparisons of the small subunit (SSU) rRNA have been consolidated as the

*Corresponding author. Tel.: +34 971 611 826; fax: +34 971 611 761.
E-mail address: rossello-mora@uib.es (R. Rosselló-Mora).
[1]Both authors contributed equally to this work.

0723-2020/$ - see front matter © 2006 Elsevier GmbH. All rights reserved.
doi:10.1016/j.syapm.2006.07.001

standard criterion for establishing a unique backbone for the classification of prokaryotes [26]. The selection of the SSU rRNA gene as a molecular chronometer [41] was based on numerous objective reasons that at the time led to the belief that such a single gene would reflect organismal phylogenies. However, and despite more than a quarter of a century of a fairly satisfactory application in the systematics of prokaryotes, the use of such a molecule as a phylogenetic marker is still heavily questioned. The main objections to its use are that a single gene may not be sufficiently representative of a whole genome [34], it is suspected that even this gene could experience interspecific horizontal exchange and/or recombination [13,36,42], and that its non-protein codifying nature could hamper the achievement of a proper alignment, thus rendering distorted phylogenies [11,21]. It seems, however, that this latter pitfall could be minimized by the availability of comprehensive databases of aligned sequences, which include information of higher order structure, enhanced by the extensive current database [26]. Another limitation of the use of a SSU rRNA gene as a phylogenetic marker is the fact that genomes can contain multiple copies of genes or operons with a significant divergence [2,26,28,29].

In parallel to the generalized use of the SSU rRNA gene sequence as a phylogenetic marker, alternative genes have been tested in order to validate and/or resolve conflicting phylogenies [24–26]. Phylogenetic reconstructions based on genes as the large subunit (LSU) rRNA, elongation factors, RNA polymerases, ATPases and recA proteins, among others, have rendered highly comparable topologies. However, the resolution power of each individual gene is restricted by the number of informative positions (i.e. size or number of homologous sites), and only the LSU rRNA has been shown to be a good alternative single gene phylogenetic marker [26].

We are now at the end of the first decade of microbial genomics, an emerging field that has led to enormous expectations such as, for example, a holistic understanding of the microbial gene pool of the planet [35]. Despite the fact that it is too soon to evaluate the impact of genomics in microbial taxonomy, the over 200 (361 published, including 41 from eukaryotes) hitherto sequenced genomes have opened a door that has allowed the search for alternative measurements of organismal relationships. Thus, with the current available data set, different treeing approaches are being evaluated, such as phylogenies based on gene content [37], gene order, evolutionary distances between orthologs, or multiple tree reconciliation approaches [40]. However, to compare the hitherto established SSU rRNA phylogenies, a straightforward approach to overcome the pitfalls of protein phylogenies due to the lack of enough informative residues may be by concatenating genes in a single alignment. Concatenates can range from a small set of selected proteins (4 different genes; [6]), to whole sets of gene families (e.g. 57 or 49 genes; [7] and [38], respectively). However, to avoid incongruent molecular phylogenies a good number of genes (e.g. 20 genes in a study of yeast species; [32]) seem to be necessary. Moreover, a recent automatic reconstruction based on the complete genome data set resulted in good agreement with previously established relationships [10]. The final selection of the 31 genes to be concatenated resulted from the exclusion of all such genes that rendered incongruent phylogenies and were thus believed to be susceptible to horizontal gene transfer.

Salinibacter ruber is the first extremely halophilic member of the *Bacteria* domain with demonstrated ecological relevance [4]. This organism, which was discovered by the use of culture independent molecular techniques, was brought to pure laboratory culture and classified as a new genus and species [5]. To date, members of *S. ruber* have been isolated from a large number of widely separated salterns, but all of them have shown an enormous degree of taxonomic resemblance [30]. Phylogenetic reconstructions based on the 16S rRNA gene [5], and on the inter-spacer region between the 16S and 23S rRNA genes [30] affiliated *S. ruber* with the phylum *Bacteroidetes,* and its closest related cultured organism was *Rhodothermus marinus*. However, in all cases, and with any algorithm used, the clade comprising *R. marinus* and *S. ruber* appeared as a deep branch within the phylum, placed close to the node of bifurcation of the superphylum that comprises *Bacteroidetes* and *Chlorobi* [26]. Besides the extremophilic nature of the cultured members of both species, the isolates appear to be aerobic and heterotrophic with more metabolic resemblance to the members of *Bacteroidetes* than to the green sulfur phototrophic members of *Chlorobi*. However, due to the tight branching order of *S. ruber*, *Bacteroidetes* and *Chlorobi* and to the accessibility of the genome information, we were encouraged to test the phylogenetic reconstruction by concatenating genes.

Our group has recently undertaken a low coverage genome-sequencing program of our model organism *S. ruber* [3]. The sequences retrieved had been used for data mining purposes, and constituted excellent raw material for initiating a genetic diversity survey of the species. The sequenced portion represented more than 70% of the M31 genome, and was comprised of over 3000 automatically annotated ORFs. In addition to our incomplete genome-sequencing program [3], the complete genome sequence of the identical strain M31 of *S. ruber* has just recently been published [27]. This database has been an excellent tool to confirm our results and the sequences of the selected genes. Here, we present the phylogenetic analysis of *S. ruber* M31 with respect to its closest relatives for which the complete

genome sequence is available. The selection of the genes has been restricted to those with as low paralogy occurrence as possible, essentiality, and presence in the genomes of the species selected for phylogenetic reconstruction. Here, we present the comparison of 22 single gene phylogenies with a concatenated alignment of the same set of genes. In addition, the results were also compared with the tree resulting from concatenating 74 genes automatically selected from among the 13 representative genomes. Altogether, the results confirm the robustness of the SSU rRNA molecule for phylogenetic analyses, and that a large number of protein coding genes are necessary to achieve a similar resolution.

Material and methods

S. ruber M31 genomic data

The low coverage genome sequencing survey on the type strain of the species M31 [3] constituted the raw material to select genes for the phylogenetic analysis. However, and due to the recent publication of the same genome by Mongodin et al. [27], nucleotide sequences of the selected genes were confirmed from the completely annotated genome available on the Institute for Genomic Research website.

Selection of genomes to be compared

The genomes of the strains listed in Table 1 were selected from those publicly available and completely sequenced. All representative genomes of the phylum *Bacteroidetes* and *Chlorobi* were used for the comparative studies. Representatives of close phyla were chosen as outgroup genomes.

Table 1. Selected species for which a complete genome sequence is available and a homologous copy of all analyzed genes had been found

Phylum	Species
Proteobacteria	*Geobacter sulfurreducens* PCA *Nitrosomonas europaea* ATCC 19718
Firmicutes	*Oceanobacillus iheyensis* HTE831
Planctomyces	*Rhodopirellula baltica* SH1
Spirochaeta	*Treponema denticola* ATCC35405
Bacteroidetes	*Bacteroides thetaiotaomicron* VPI-5482 *Bacteroides fragilis* YCH46 *Porphyromonas gingivalis* W83 *Prevotella intermedia* 17 *Cytophaga hutchinsonii*
Chlorobi	*Chlorobium tepidum* TLS[a] *Chlorobium chlorochromatii* CaD3

[a]now reclassified as *Chlorobaculum tepidum*.

Manual gene selection

The preliminary decision on the type of genes used to reconstruct a concatenated phylogeny was made using several different criteria: (i) the genes should represent functional metabolisms understood as housekeeping features, and when possible homologous to those previously recommended in similar studies [14,16,32,43]; (ii) the genes should be spread and distant from each other through the whole genome of *S. ruber*; (iii) the genes should be widely distributed within the selected bacterial genomes; (iv) the genes must have an homologous copy in at least 10 genomes of the representative strains selected for the study (Table 1); (v) orthology assignment of the genes should be unambiguous upon inspection of initial trees (to facilitate this inspection, gene families with many duplications in the different genomes were previously eliminated); (vi) genes should represent different protein families responsible for different metabolic features rather than genes with an expected synchronized evolution, such as, for example, those involved in the translational apparatus [7]. In total, we found 22 genes matching these criteria (Table 2).

Automatic gene selection

The *S. ruber* M31 genome was blasted by the use of the Blastp algorithm [1] against the 12 remaining selected genomes. Genes present in at least 10 genomes (with an E-value $<1e^{-10}$) and with less than 52 hits were selected. We arbitrarily selected this number of hits to avoid genes with an average of more than 4 copies per genome, which we considered an excessive number of genes to make reliable orthology assignments. Sequences giving hits were aligned with Mafft [20] and cleaned using Gblocks 0.91 [9] with relaxed conditions in order to preserve as much information as possible: ''Minimum number of sequences for a conserved position'' and ''Minimum number of sequences for a flank position'' were both half the number of sequences, ''Maximum number of contiguous nonconserved positions'' was 10, ''Minimum length of a block'' was 5 and ''Allowed gap positions'' was ''With half''. An initial tree was estimated by maximum likelihood using Phyml [15] with the JTT model of protein evolution [18] and a gamma rate distribution using 4 rate categories and a proportion of invariable sites. The gamma rate and the proportion of invariable sites were estimated from the data. In the resulting trees, two topology filters were applied to try to avoid the use of

sequences with a paralogous relationship. First, all trees where multiple sequences of the same species did not appear monophyletic were discarded. When multiple sequences of the same species were monophyletic, only the sequence with the shortest branch in the tree was selected. This filter eliminated gene families where the risk of selecting paralogues was too high, giving a remainder of 152 genes. However, even upon applying this filter some obvious non-orthologues remained, and particularly some affecting *S. ruber*. Thus, a second filter was applied in which gene families with trees where *Bacteroidetes*, *Chlorobium* and *S. ruber* did not appear monophyletic were eliminated. These three groups could appear in any order in the selected families. The rationale for the application of this filter is that genes that are not able to recover an obvious clade, as supported by previous studies [25], may not be good for resolving the more difficult trichotomy of interest. In addition, this procedure selects genes sharing a common history, most surely the history of the organism. However, the use of the families that did not pass the second filter produced the same topology, indicating that this was not essential to find the phylogenetic position of *Salinibacter* (not shown). To be stricter in the selection of genes, we applied the two filters, which gave rise to 88 genes. Cleaning the alignments of these genes with Gblocks using stringent parameters (everything default except that ''Allowed gap positions'' was ''With half''), which are better for gene concatenation, eliminated all positions from 14 alignments. Thus, the final automatically selected set was composed of 74 genes.

Tree construction

Each individual gene alignment was used to reconstruct its phylogeny by using the maximum likelihood algorithm with Phyml [15] and the same conditions as for the initial trees of the automatic procedure (see above).

To reconstruct the concatenated trees, the genes that fulfilled the requisites proposed for the two data sets were concatenated, after applying Gblocks with stringent conditions, into two single alignments, respectively. From each alignment, a maximum likelihood tree was constructed with Phyml using the same parameters as above, except that 6 rate categories were used for the gamma distribution.

Bootstrap support in concatenates with different numbers of genes

To measure the bootstrap support as a function of the number of concatenated genes, sets of 4–16 genes were randomly chosen from the manually selected genes and concatenated. For each set, a total of 20 random concatenates were generated. The support of the clade uniting *S. ruber* with *Bacteroidetes* for each individual sample was measured by calculating 100 bootstraps with Phyml, and taking from the consensus tree the bootstrap support of this clade. For those few alignments where the consensus tree showed a different branching pattern, the percent number of times in which *S. ruber* appeared together with *Bacteroidetes* was calculated from the 100 bootstrap trees. Finally, the average of the 20 bootstrap values of each gene set was calculated.

Results and discussion

In the present study 22 genes were manually selected to perform the main phylogenetic analysis. Only those genes that fulfilled the strict criteria listed in the Material and methods section were used. The complete initial selection of the genes was comprised of 39 genes, 17 of which were removed from the study because they either presented paralogy in *S. ruber* M31, or showed irregular alignments with a small number of conserved homologous positions after being filtered with the Gblocks program [9]. The remaining 22 genes represented different metabolic mechanisms and appeared well scattered within the genome (Table 2).

The phylogeny of all genes was individually reconstructed with the maximum likelihood algorithm, as implemented in the Phyml program [15]. The 22 different reconstructions fell into the three different categories as shown in Fig. 1. In the most frequent tree topology, with 54% of the cases, M31 affiliated with the members of the phylum *Bacteroidetes* (genes: *ileS, pyrG, rpsC, S5, rpoC, rpoB, gyrB, thrS, mfd, ftsY, tuf, uvrA-2*; Fig. 1a). These branching orders were followed by a topology where M31 affiliated with the single representative of the phylum *Chlorobi*, in 27% of the cases (genes: *ffh, glyA, recN, ruvB, recG, rho*; Fig. 1b). The third most frequent topology was that where M31 affiliated outside both phyla, but still with the superphylum comprising *Chlorobi* and *Bacteroidetes*. This last case included 18% of the genes (genes: *groEL, recA, uvrA, valS*; Fig. 1c). It is remarkable here that *recA* did not support the phylogeny based on SSU rRNA [4,5], nor the global results after concatenation. *recA* has been used for identification and classification purposes (e.g. [19,31]), despite the fact that it has been shown to lack resolving power and to have undetected paralogy effects [25]. Additionally, some of the genes initially analyzed (e.g. *lig, pgk, moxR, pckA, proA*) showed tree topologies that most probably corresponded to paralogous sequences. Some studies have explained these phenomena in terms of horizontal gene transfer [10,13,27]. However, as the genomic information increases, it becomes more

Table 2. Genes used in the present work for complete phylogenetic analysis

Gene	Protein	Length[a]	% after Gblocks[b]	Affiliation[c]	Gene position[d]	GO[e]
recG	ATP-dependent DNA helicase	700	70% (553/785)	Chlo	361774…363876	0004003
recN	Recombination/replication protein	572	78% (462/585)	Chlo	1015436…1017154	0006310
rho	Transcription termination factor	472	48% (371/770)	Chlo	2769096…2770514	0003715
ffh	Signal recognition protein	461	82% (412/502)	Chlo	2025515…2026900	0048501
ruvB	Holliday junction helicase subunit A	344	80% (311/384)	Chlo	1441721…1442755	0016055
glyA	Serine hydroxymethyl transferase	432	72% (384/531)	Chlo	986911…988209	0004372
mfd	Transcription-repair coupling factor	1142	59% (815/1364)	Bact	61566…64994	0006281
rpoB	DNA directed RNA polymerase beta subunit	1335	68% (1120/1629)	Bact	2173447…2177454	0003899
ileS	Isoleucyl t-RNA synthetase	1080	64% (940/1459)	Bact	611236…614478	0004822
FtsY	signal recognition particle-docking protein FtsY	339	79% (299/376)	Bact	2687022…2687996	0005786
rpoC	RNA polymerase beta subunit	1448	79% (1314/1652)	Bact	2168989…2173335	0003899
rpsC	Ribosomal protein S3	236	75% (196/260)	Bact	1323483…1324193	0009283
S5	Ribosomal protein S5	175	88% (157/178)	Bact	1328851…1329378	0009283
thrS	Threonyl t-RNA synthetase	694	72% (618/848)	Bact	3458760…3458833	0004829
tuf	Translation elongation factor G	396	95% (385/403)	Bact	1317874…1319064	0003746
uvrA-2	Excinuclease ABC subunit A-2	1016	35% (822/2295)	Bact	2971341…2974391	0009381
gyrB	DNA gyrase subunit B-2	652	70% (587/832)	Bact	3003154…3005112	0003918
pyrG	CTP synthase	567	84% (500/592)	Bact	724839…726536	0003883
valS	Valyl t-RNA synthetase	901	62% (718/1155)	Bact/Chlo	2530916…2528211	0004832
recA	Recombination protein	353	73% (313/428)	Bact/Chlo[f]	1907949…1909010	0006310
uvrA	Excinuclease ABC subunit A	981	38% (781/2028)	Bact/Chlo	601430…604375	0009381
groEL	Chaperonin Hsp 60 1	570	76% (510/666)	Bact/Chlo	2622845…2624527	0005515

[a]Length of the original protein in M31.
[b]Percentage of the original protein remaining after filtering with Gblocks, in brackets the absolute number of the remaining homologous positions, and the original length of the alignment.
[c]Affiliation of M31 after the individual reconstructions.
[d]Gene position in the complete genome of M31 [27].
[e]GO_numbers of each gene.
[f]In its tree, *R. baltica* branches off between *S. ruber* and the *Bacteroidetes*/*Chlorobium* group.

evident that there may be other genetic forces responsible for phylogenetic discrepancies rather than gene exchange [22,23,25]. Such incongruencies can be attributed to the difficulty of obtaining orthologous sequences, different mutation rates that distort the phylogenetic results, lack of resolution power due to small alignments, choice of sampled taxa and/or assumptions in the election of the substitution model [8,32,39]. Most of the tree topologies placed *Salinibacter* M31 with the superphylum *Chlorobi-Bacteroidetes*, a phylogenetic relationship that agreed with the 16S rRNA gene sequence analysis [4,5]. However, due to the relatively high percentage affiliation with *Chlorobi* (27%) in relation to *Bacteroidetes* (54%) one could conclude that the branch comprising the genus *S. ruber* must have diverged at a very early stage, close to the divergence of the two main phyla of the superphylum. Actually, it cannot be discarded that the branch comprising the two genera *S. ruber* and *R. marinus* represents a new phylum within the domain *Bacteria*.

To avoid stochastic problems with single genes, the best approach to validate the 16S rRNA gene phylogeny of *S. ruber* can be obtained from a comparison with the concatenation of essential and genetically, metabolically, and genomically unlinked genes. As in similar analyses [7,8,10,17,32], we concatenated the 22 selected genes into a single alignment with 10,757 homologous positions. The resulting tree shown in Fig. 2 was obtained after performing maximum likelihood analysis as implemented in the program Phylml. This tree showed a topology identical to that observed for the SSU rRNA gene analysis, in which *S. ruber* affiliates with *Bacteroidetes* [4,5]. Branches with low bootstrap values (i.e. the branching of *Rhodopirellula baltica* and *Treponema denticola*) also rendered irresolvable branching orders in the 16S rRNA gene analysis. It can be assumed that a tree based on an alignment nearly 10 times larger than that of the SSU rRNA may show major robustness in its results. Actually, the branching order of this tree (Fig. 2) is very strongly supported by the high bootstrap values, in most of the cases with 100% frequency. The tree is based on selected genes that are distantly placed within the genome, and represent different metabolisms of the cell. One can thus assume that the concatenated tree may reflect the organism's phylogeny. Given that the tree branching pattern is identical to that of the SSU rRNA we can conclude that previously calculated trees [4,5] may also reflect the phylogeny of the organism and

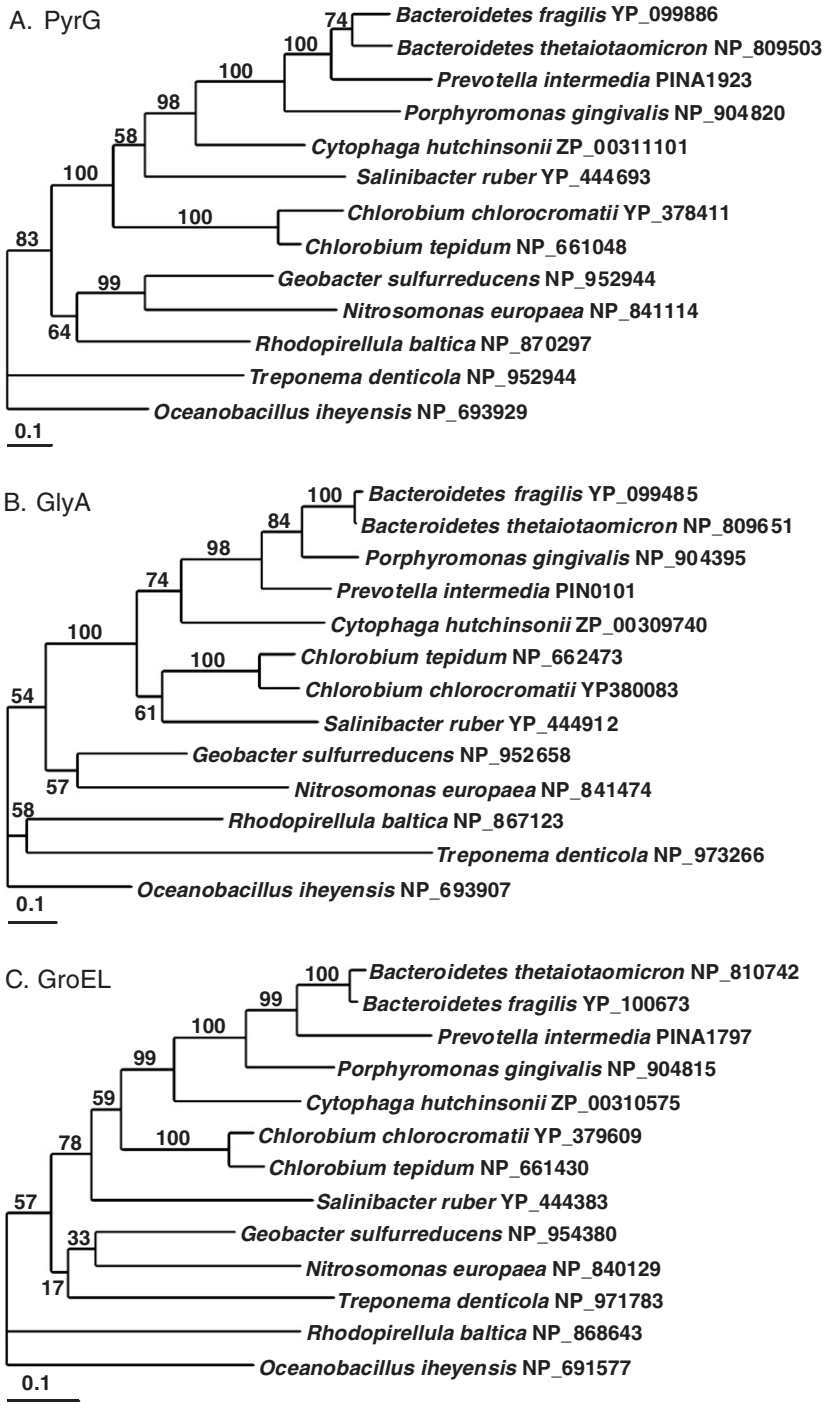

Fig. 1. Phylogenetic reconstructions of three selected gene products based on maximum likelihood as implemented in the program Phyml. (a) CTP synthase (*pyrG*), affiliates with *Bacteroidetes*; (b) *glyA*, affiliates with *Chlorobi*; (c) Chaperonin Hsp 60 (*groEL*), affiliates with the superphylum *Bacteroidetes-Chlorobi*, but outside both phyla. Numbers above the branches represent bootstrap support from 100 replicas. The scale bar represents 0.1 substitutions/position.

not only the genealogy of an insignificant portion of the whole genome.

In order to validate the manually aligned tree, we automatically selected a total of 74 orthologous genes present in at least 10 of the 13 full genomes analyzed in the current study. The final alignment contained 17,149 homologous positions. The tree topology observed with this latter approach was identical to that of the 22 genes, and also linked *S. ruber* to *Bacteroidetes* with 100% support (not shown). A similar analysis using 152 genes selected after the application of less strict filters to avoid paralogous sequences, also reproduced the same

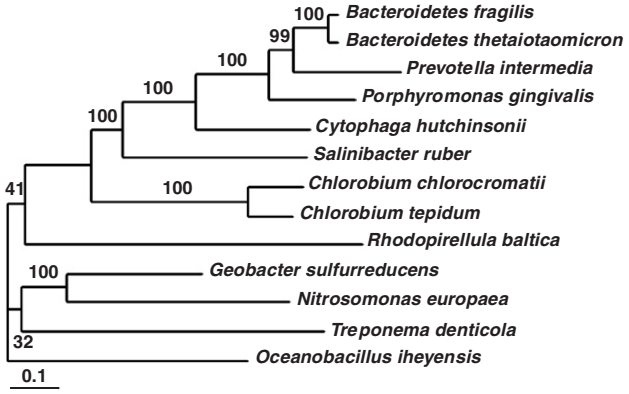

Fig. 2. Phylogenetic reconstruction based on the concatenation of the 22 selected genes. The analysis was based on an alignment of 10,757 homologous positions. The tree shown is based on the maximum likelihood algorithm as implemented in the program Phyml. Numbers above the branches represent bootstrap support from 100 replicas. The scale bar represents 0.1 substitutions/position.

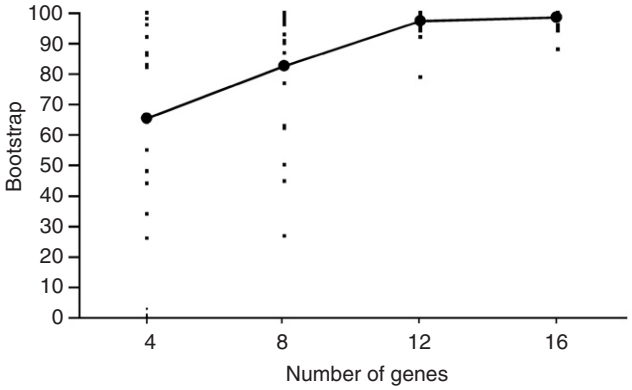

Fig. 3. Bootstrap support for the affiliation of *S. ruber* with *Bacteroidetes* in samples with different numbers of genes. The line connects the average of 20 samples analyzed for each data set with different number of genes.

topology regarding the grouping of *S. ruber* and *Bacteroidetes*. In this case this grouping showed a bootstrap support of 87% (not shown).

Finally, we wanted to know how many genes were necessary to obtain such good support for the phylogenetic placement of *S. ruber*. To test this, we proceeded with the analysis of different data sets based on the 22 manually aligned genes. The data sets were composed of concatenations that ranged from 4 to 16 randomly selected genes (with 20 samples each) and they were used to analyze the bootstrap support for the affiliation of *S. ruber* and *Bacteroidetes* as a function of the size of the concatenated alignments. As can be seen in Fig. 3, the increase of the number of genes in the analysis resulted in the stabilization of the branching orders, which is reflected by increased bootstrap values. From this plot, it can be deduced that a stable phylogeny for *S. ruber* (i.e. >90% bootstrap support) can only be achieved by the use of at least 12 independent genes. Reconstructions performed with smaller sets of genes rendered trees with much lower support.

Being pragmatic, and in the light of the present study, it can be concluded that for classification purposes the SSU rRNA gene phylogenetic reconstruction may be a very convenient option. Results like this reinforce the fact that the use of the SSU rRNA for microbial taxonomy purposes may still be the approach of choice when a given strain is to be classified and no genome data is available. In addition, as already stated by Rokas et al. [32], our results also show that a robust tree made from protein sequences should contain a large number of protein alignments. Smaller alignments may suffer from greater influence of each single gene and the branching orders observed may lack stability.

Acknowledgments

The authors' research is founded by the following projects: BOS-2003-05198-C02-01 and -02 of the Spanish Ministry of Science and Education, a research grant in bioinformatics from the Fundación BBVA (Spain), the grant GRUPOS03/137 from the Generalitat Valenciana, and the network of excellence Marine Genomics Europe GOCE-CT-505403 of the European Union.

References

[1] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[2] G. Amann, K.O. Stetter, E. Llobet-Brossa, R. Amann, J. Antón, Direct proof for the presence and expression of two 5% different 16S rRNA genes in individual cells of *Haloarcula marismortui*, Extremophiles 4 (2000) 373–376.
[3] J. Antón, A. Peña, M. Valens, F. Santos, F.O. Glöckner, M. Bauer, J. Dopazo, J. Herrero, R. Rosselló-Mora, R. Amann, *Salinibacter ruber*: genomics and biogeography, in: N. Gunde-Cimerman, A. Oren, A. Plemenitas (Eds.), Adaptation to Life at High Salt Concentrations in *Archaea*, *Bacteria*, and *Eukarya* (Cellular Origin, Life in Extreme Habitats and Astrobiology), Kluwer Academic Publishers BV, Dordrecht, 2005, pp. 257–266.
[4] J. Antón, R. Rosselló-Mora, F. Rodríguez-Valera, R. Amann, Extremely Halophilic *Bacteria* in crystallizer ponds from solar salterns, Appl. Environ. Microbiol. 66 (2000) 3052–3057.
[5] J. Antón, A. Oren, S. Benlloch, F. Rodríguez-Valera, R. Amann, R. Rosselló-Mora, *Salinibacter ruber* gen. nov., sp. nov., a novel, extremely halophilic member of the *Bacteria* from saltern crystallizer ponds, Int. J. Syst. Evolut. Microbiol. 52 (2002) 485–491.
[6] S. Baldauf, A.J. Roger, I. Wenk-Siefert, W.F. Doolittle, A kingdom-level phylogeny of Eukaryotes based on combined protein data, Science 290 (2000) 972–977.

[7] C. Brochier, E. Bapteste, D. Moreira, H. Philippe, Eubacterial phylogeny based on translational apparatus proteins, Trends Genet. 18 (2002) 1–5.
[8] J.R. Brown, C.J. Douady, M.J. Italia, W.E. Marshall, M.J. Stanhope, Universal trees based on large combined protein sequence data sets, Nat. Genet. 28 (2001) 281–285.
[9] J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, Mol. Biol. Evol. 17 (2000) 540–552.
[10] F.D. Ciccarelli, T. Doerks, C. von Mering, C.J. Creevey, B. Snel, P. Bork, Toward automatic reconstruction of a highly resolved tree of life, Science 311 (2006) 1283–1287.
[11] M.T. Dixon, D.M. Hillis, Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis, Mol. Biol. Evol. 10 (1993) 256–267.
[12] G.E. Fox, K.J. Pechman, C.R. Woese, Comparative cataloguing of 16S ribosomal ribonucleic acid: molecular approach to prokaryotic systematics, Int. J. System. Bacteriol. 27 (1977) 44–57.
[13] J.P. Gogarten, W.F. Doolittle, J.G. Lawrence, Prokaryotic evolution in light of gene transfer, Mol. Biol. Evol. 19 (2002) 2226–2238.
[14] E. Griffiths, R.S. Gupta, The use of signature sequences in different proteins to determine the relative branching order of bacterial divisions: evidence that *Fibrobacter* diverged at a similar time to *Chlamydia* and the *Cytophaga-Flavobacterium-Bacteroides* division, Microbiology 147 (2001) 2611–2622.
[15] S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, Syst. Biol. 52 (2003) 696–704.
[16] R.S. Gupta, What are archaebacteria: life's third domain or monoderm prokaryotes related to Gram-positive bacteria? A new proposal for the classification of prokaryotic organisms, Mol. Microbiol. 29 (1998) 695–707.
[17] S. Hansmann, W. Martin, Phylogeny of 33 ribosomal and six other proteins encoded in an ancient gene cluster that is conserved across prokaryotic genomes: influence of excluding poorly alignable sites from analysis, Int. J. Syst. Evol. Microbiol. 50 (2000) 1655–1663.
[18] D.T. Jones, W.R. Taylor, J.M. Thornton, The rapid generation of mutation data matrices from protein sequences, Comput. Appl. Biosci. 8 (1992) 275–282.
[19] S. Karlin, G.M. Weinstock, V. Brendel, Bacterial classifications derived from recA protein sequence comparisons, J. Bacteriol. 177 (1995) 6881–6893.
[20] K. Katoh, K. Kuma, H. Toh, T. Miyata, MAFFT version 5: improvement in accuracy of multiple sequence alignment, Nucleic Acids Res. 33 (2005) 511–518.
[21] K.M. Kjer, Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs, Mol. Phylogenet. Evol. 4 (1995) 314–330.
[22] C.G. Kurland, Paradigm lost, in: J. Sapp (Ed.), Microbial Phylogeny and Evolution, Concepts and Controversies, Oxford University Press, Oxford, 2005, pp. 207–223.
[23] C.G. Kurland, B. Canback, O.G. Berg, Horizontal gene transfer: a critical view, Proc. Natl. Acad. Sci. USA 100 (2003) 9658–9662.
[24] W. Ludwig, K.-H. Schleifer, Phylogeny of *Bacteria* beyond the 16S rRNA standard, ASM News 65 (1999) 752–757.
[25] W. Ludwig, K.-H. Schleifer, Molecular phylogeny of *Bacteria* based on comparative sequence analysis of conserved genes, in: J. Sapp (Ed.), Microbial Phylogeny and Evolution, Concepts and Controversies, Oxford University Press, Oxford, 2005, pp. 70–98.
[26] W. Ludwig, H.P. Klenk, Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematics, in: second edD.J. Brenner, N.R. Krieg, J.T. Staley, G.M. Garrity (Eds.), Bergey's Manual of Systematic Bacteriology, vol. 1, Springer SBM, New York, 2001, pp. 49–54.
[27] E.F. Mongodin, K.E. Nelson, S. Daugherty, R.T. DeBoy, J. Wister, H. Khouri, J. Weidman, D.A. Walsh, R.T. Papke, G. Sánchez Pérez, A.K. Sharma, C.L. Nesbø, D. MacLeod, E. Bapteste, W.F. Doolittle, R.L. Charlebois, B. Legault, F. Rodriguez-Varela, The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea, Proc. Natl. Acad. Sci. USA 102 (2005) 18147–18152.
[28] S. Mylvaganam, P.P. Dennis, Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic archaeabacterium *Haloarcula marismortui*, Genetics 130 (1992) 399–410.
[29] U. Nübel, B. Engelen, A. Felske, J. Snaidr, A. Wieshuber, R.I. Amann, W. Ludwig, H. Backhaus, Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis, J. Bacteriol. 178 (1996) 5636–5643.
[30] A. Peña, M. Valens, F. Santos, S. Buczolits, J. Antón, P. Kämfer, H.J. Busse, R. Amann, R. Rosselló-Mora, Intraspecific comparative analysis of the species *Salinibacter ruber*, Extremophiles 9 (2005) 151–161.
[31] G.W. Payne, P. Vandamme, S.H. Morgan, J.J. LiPuma, T. Coenye, A.J. Weightman, T.H. Jones, E. Mahenthiralingam, Development of a *recA* gene-based identification approach for the entire Burkholderia genus, Appl. Environ. Microbiol. 71 (2005) 3917–3927.
[32] A. Rokas, B.L. Williams, N. King, S.B. Carroll, Genome-scale approaches to resolving incongruence in molecular phylogenies, Nature 425 (2003) 798–804.
[33] R. Rosselló-Mora, Updating prokaryotic taxonomy, J. Bacteriol. 187 (2005) 6255–6257.
[34] S.R. Santos, H. Ochman, Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins, Environ. Microbiol. 6 (2004) 754–759.
[35] S.C. Schuster, G. Gottschalk, Genomics, microbial genomics in its second decade, Curr. Opin. Microbiol. 8 (2005) 561–563.
[36] P.H. Sneath, Evidence from *Aeromonas* for genetic crossing-over in ribosomal sequences, Int. J. Syst. Evol. Microbiol. 43 (1993) 626–629.
[37] B. Snel, P. Bork, M.A. Huynen, Genome phylogeny based on gene content, Nat. Genet. 21 (1999) 108–110.
[38] H. Teeling, T. Lombardot, M. Bauer, W. Ludwig, F.O. Glöckner, Evaluation of the phylogenetic position of the planctomycete '*Rhodopirellula baltica*' SH 1 by means of

concatenated ribosomal protein sequences, DNA-directed RNA polymerase subunit sequences and whole genome trees, Int. J. Syst. Evol. Microbiol. 54 (2004) 791–801.

[39] J.E. Wertz, C. Goldstone, D.M. Gordon, M.A. Riley, A molecular phylogeny of enteric bacteria and implications for a bacterial species concept, J. Evol. Biol. 16 (2003) 1236–1248.

[40] Y.I. Wolf, I.B. Rogozin, N.V. Grishin, E.V. Koonin, Genome trees and the tree of life, Trends Genet. 18 (2002) 472–479.

[41] C.R. Woese, Bacterial evolution, Microbiol. Rev. 51 (1987) 221–271.

[42] W.H. Yap, Z. Zhang, Y. Wang, Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal gene transfer of an entire rRNA operon, J. Bacteriol. 181 (1999) 5201–5209.

[43] D.R. Zeigler, Gene sequences useful for predicting relatedness of whole genomes in bacteria, Int. J. Syst. Evol. Microbiol. 53 (2003) 1893–1900.