Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 34 (2005) 371–381
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Reconstructing evolutionary relationships from functional data: a consistent classification of organisms based on translation inhibition response

Carlos Briones[a,*], Susanna C. Manrubia[a], Ester Lázaro[a], Antonio Lazcano[b], Ricardo Amils[a,c,*]

[a] *Centro de Astrobiología, Carretera de Ajalvir Km. 4, 28850 Torrejón de Ardoz, Madrid, Spain*
[b] *Facultad de Ciencias, UNAM, Apdo. Postal 70-407, Cd. Universitaria, 04510 México D. F., Mexico*
[c] *Centro de Biología Molecular 'Severo Ochoa,' Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain*

Received 29 October 2003; revised 6 September 2004

Abstract

The last two decades have witnessed an unsurpassed effort aimed at reconstructing the history of life from the genetic information contained in extant organisms. The availability of many sequenced genomes has allowed the reconstruction of phylogenies from gene families and its comparison with traditional single-gene trees. However, the appearance of major discrepancies between both approaches questions whether horizontal gene transfer (HGT) has played a prominent role in shaping the topology of the Tree of Life. Recent attempts at solving this controversy and reaching a consensus tree combine molecular data with additional phylogenetic markers. Translation is a universal cellular function that involves a meaningful, highly conserved set of genes: both rRNA and r-protein operons have an undisputed phylogenetic value and rarely undergo HGT. Ribosomal function reflects the concerted expression of that genetic network and consequently yields information about the evolutionary paths followed by the organisms. Here we report on tree reconstruction using a measure of the performance of the ribosome: antibiotic sensitivity of protein synthesis. A large database has been used where 33 ribosomal systems belonging to the three major cellular lineages were probed against 38 protein synthesis inhibitors. Different definitions of distance between pairs of organisms have been explored, and the classical algorithm of bootstrap evaluation has been adapted to quantify the reliability of the reconstructions obtained. Our analysis returns a consistent phylogeny, where archaea are systematically affiliated to eukarya, in agreement with recent reconstructions which used information-processing systems. The integration of the information derived from relevant functional markers into current phylogenetic reconstructions might facilitate achieving a consensus Tree of Life.

Keywords: Phylogeny; rRNA trees; Functional evolution; Ribosomes; Protein synthesis inhibitors; Antibiotics

1. Introduction

Phylogenetic analysis of small ribosomal subunit RNA (SSU rRNA) genes has led to an unprecedented ordering of all extant organisms into groups on the basis of their evolutionary relationships (Woese, 1987; Woese et al., 1990). This approach, which is to a considerable extent the outcome of the combined development of both molecular biology and computer science, has had a major impact on our understanding not only of the classification but also of the evolution of organisms. By revealing the existence of three major domains of cell descent, SSU rRNA-based phylogenies have led to

* Corresponding authors. Fax: +34 91 5201074 (C. Briones), +34 91 4978087 (R. Amils).
E-mail addresses: brioneslc@inta.es (C. Briones), ramils@cbm.uam.es (R. Amils).

1055-7903/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2004.10.020

relevant insights into early stages of cellular evolution. Nonetheless, these analyses have also resulted in major, ongoing, taxonomic debates (Bromham and Penny, 2003; Doolittle, 1999; Huynen et al., 1999; Mayr, 1998; Woese, 2000; Wolf et al., 2002). At the base of the controversy lies the fact that tree reconstruction is affected by horizontal gene transfer, which is currently acknowledged as a driving force of prokaryotic evolution (Clarke et al., 2002; Kurland et al., 2003; Nesbo et al., 2001; Ochman et al., 2000) and may even involve lateral transfer of rRNA operons (Perez-Luz et al., 1998; Yap et al., 1999). Other discrepancies are related to the constancy of the rate of the molecular clock among different genes and organisms, and to the influences that population size and fixation of mutations with a selective effect have on the divergence of sequences (Bromham and Penny, 2003). Moreover, the release of complete genome sequences from over 200 organisms has fueled this debate by showing that phylogenies constructed with many universally distributed genes may exhibit important differences with the rRNA-based universal tree.

Sequence comparison plays a prominent role in reconstructing the evolution of genes. Nevertheless, since each gene may have its own evolutionary history, such phylogeny does not necessarily reflect that of the organisms carrying those genes. This has led to claim that new evolutionary paradigms are required, where genes and organisms are all considered in concert (Doolittle, 1999; Moritz and Hillis, 1990; Vandamme et al., 1996). The current discussions have underlined the need for more integrative approaches such as genome trees based on gene content. The resulting trees, which are not phylogenies in the classical sense since they are the outcome of a simple hierarchical classification, are nevertheless remarkably similar to the overall topology of SSU rRNA trees. These dendrograms are based on the overall resemblance among species, and are constructed using measures of similarity of the genomes (Fitz-Gibbon and House, 1999; Snel et al., 1999; Tekaia et al., 1999), the distribution of orthologs and protein folds (Lin and Gerstein, 2000; Wolf et al., 2001), and comparisons of metabolic and information transfer pathways (Podani et al., 2001). Their topology, close to that of SSU rRNA phylogenies, indicates the conservation of a strong evolutionary signal within genomes regardless of the complexity of their evolutionary histories in which gene duplication, gene loss, and lateral transfer must have taken place (Bromham and Penny, 2003; Fitz-Gibbon and House, 1999; Lin and Gerstein, 2000; Podani et al., 2001; Snel et al., 1999; Tekaia et al., 1999; Wolf et al., 2002).

Whole-genome assessments are useful approaches with which to understand evolutionary patterns, allowing important inferences about ancestral states to be made. Such a method can be extended to some phenotypic characters (morphological, physiological or biochemical ones) if they have a clear genetic basis, and if the data are collected in such a way that phylogenetic properties can be inferred from them. Individually, many of these phenotypic characters are not relevant as parameters for genetic relatedness, and their use was generally limited to checking the biological consistency of the classifications obtained using genotypic data (Hillis, 1987; Wayne et al., 1987). However, there are some phenotypic properties that correspond to universally distributed functions and structures, with clear genetic bases which are independent of environmental effects or growth conditions. Therefore, these features are good candidates to test evolutionary hypotheses.

One example of such phenotypic properties is the antibiotic sensitivity of protein synthesis. Although antibiotic sensitivities were used to settle the eubacterial origin of mitochondria (Tzagoloff et al., 1979) and strongly hinted at the relationship between some components of the translation machineries of nucleated cells and the archaea (Zillig, 1987), their use as phylogenetic markers has been relatively limited. During the last two decades our group has been involved in the study of the protein synthesis machinery of organisms in order to assess its phylogenetic role (Amils and Sanz, 1986; Sanz et al., 1994). The advantages of the translational apparatus over other macromolecular assemblies are manifold, and are based on the outstanding amount of structural, functional, and genetic information available for the ribosomal system. Since ribosomes are complexes containing three to four rRNA molecules and more than 50 proteins, it is highly unlikely that the whole set of genes involved undergoes horizontal transfer. In addition, rRNAs are the main traditionally accepted molecular clocks (Pace, 1997; Woese, 1987; Woese et al., 1990) and it has been shown that the trees made of concatenated alignments of ribosomal protein sequences carry a strong phylogenetic signal (Wolf et al., 2001). This clearly reveals the intrinsic phylogenetic value of the ribosomal particle and its constituents.

Our study of the ribosomal systems has been developed using protein synthesis inhibitors with different taxonomic and functional specificity. The cell-free in vitro analysis performed avoids problems related to transport, inactivation, or the pleiotropic effects of antibiotics. It also allows the optimization of the ionic conditions for every system, the realization of the required control experiments, and the suitable collection and standardization of the data (Cammarano et al., 1985; Sanz et al., 1992). Such datasets are amenable to quantitative analysis, and may be used to derive a classification of organisms based on their response to protein synthesis inhibitors. Preliminary studies suggested the inherent phylogenetic value of the sensitivity of ribosomes to antibiotics, although the simple approach

initially used only considered a very limited part of the available information, and inhibition data from a representative number of eukaryotes was not included (Amils et al., 1989, 1990; Sanz et al., 1994). A further analysis was carried out to study the fractal dimension of the dataset and, despite its complexity, a meaningful clustering of the organisms analyzed was obtained (Briones et al., 1998).

Here we present a method based on cluster analysis to extract most of the information contained in the whole set of the inhibition curves for 33 representative ribosomal systems belonging to the domains Archaea, Bacteria, and Eucarya. A simple procedure to define distances between pairs of organisms has allowed the construction of phenograms that closely resemble those of SSU rRNA sequence comparison, and whose robustness can be measured by bootstrap analysis. The results obtained reveal the possibility of reconstructing evolutionary relationships from the quantification of functions that are subjected to selective processes.

2. Methods

2.1. Biological data collection

Thirty-three ribosomal systems from organisms belonging to the three major lineages were analyzed. They were tested in optimized in vitro translation systems against 38 protein synthesis inhibitors belonging to the three groups of specificity: inhibitors of bacterial ribosomes (I), inhibitors of eukaryotic ones (II), and universal inhibitors (III). It should be emphasized that this taxonomic classification of antibiotics (Vázquez, 1979) was developed prior to the recognition of Archaea as a distinct cellular lineage (Woese et al., 1990). However, no protein synthesis inhibitor specific for archaea has been characterized up to now. Ribosomal systems analyzed are listed in Table 1, and the antibiotics used are shown in Table 2. Further details on organisms and protein synthesis inhibitors have been discussed elsewhere (Briones et al., 1998). The preparation of the different cell-free systems and the experimental conditions for in vitro protein synthesis assays were previously described (Altamura et al., 1995; Amils and Sanz, 1986; Cammarano et al., 1985; Casquero, 1992; Elhardt and Böck, 1982; Oliver et al., 1987; Sánchez et al., 1994; Sanz et al., 1994, 1988). Examples of the inhibition curves obtained with the antibiotics tested were shown elsewhere (Amils et al., 1989; Sanz et al., 1994).

2.2. Data description

Up to six measures were taken for each ribosome-inhibitor pair. They correspond to six different concentrations (c) of the antibiotic in the in vitro translation system: 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, and 10^{-3} M. Let us call $x_i^\alpha(c)$ the functional response of organism i to antibiotic α at concentration c. The expression $x_i^\alpha(c)$ stands for the percentage of inhibition in the protein synthesis assay with respect to a control experiment without antibiotic. This quantity varies from negative values (corresponding to activation instead of inhibition produced by the antibiotic, a situation sometimes found at the lowest concentrations of the antibiotic) to 100 (complete inhibition). This allows a direct comparison of the responses for different ribosomal systems and antibiotics. In spite of the exhaustive experimental analysis developed, which involved 33 organisms and 38 protein synthesis inhibitors, not all the combinations (i,α) were available. A relevant example of this limitation was the impossibility of using aminoglycoside antibiotics against ribosomes from extreme halophilic archaea, due to the competition for binding sites between such cationic inhibitors and the high concentration of cations required for the maintenance of the functional structure of halophilic ribosomes (Amils et al., 1990; Sanz et al., 1992). In other ribosome-inhibitor pairs, only a subset of the six possible antibiotic concentrations had been assayed. We have designed a procedure that takes into account as many data points as possible. Four different definitions of the distance between organisms have been compared; in all cases the introduction of additional data through interpolation or similar procedures has been avoided.

2.3. Distances between the functional response of the ribosomal systems

Each organism i in the system is defined by an ordered vector $x_i^\alpha(c)$ composed of 228 elements: 38 antibiotics times six possible concentrations.

There are many ways to define differences between vectors and thus to get a distance $d(i,j)$ between a pair of organisms i and j. Several proposals can be found elsewhere (Hartigan, 1975; StatSoft, 2002). Here we introduce the following general form:

$$d(i,j) = \frac{1}{A(i,j)} \sum_\alpha \frac{1}{K^\alpha(i,j)} \sum_c |x_i^\alpha(c) - x_j^\alpha(c)|^p. \qquad (1)$$

We have studied four different possible distances on the basis of this definition:

- *Distance 1*. The absolute value of the distance is used and all the antibiotics are equally weighted. This amounts to setting $p = 1$, and defining $K^\alpha(i,j)$ as the number of common measures for antibiotic α in organisms i and j, and $A(i,j)$ as the number of antibiotics for which at least one concentration value was assayed for both organisms. The first sum runs over

Table 1
Ribosomal systems tested in this study, and their corresponding SSU rRNA sequences used for the molecular phylogeny analysis

No.	Organism		SSU rRNA	
	Genus	Species	GenBank Accession No.	Length (nt)
	Archaea			
	Halophilic archaea			
1	*Haloarcula*	*californiae*	—	—
2	*Haloarcula*	*marismortui*	X61688	1472
3	*Haloarcula*	*sinaiiensis*	D14129	1471
4	*Halobacterium*	*salinarium*	AJ002947	1458
5	*Halobacterium*	*halobium*	M38280	1473
6	*Halococcus*	*morrhuae*	X00662	1475
7	*Haloferax*	*gibbonsii*	D13378	1470
8	*Haloferax*	*mediterranei*	D11107	1472
9	*Natromonas*	*pharaonis*	D87971	1465
10	*Natronococcus*	*occultus*	Z28378	1464
	Sulfur-dependent thermophiles			
11	*Acidianus*	*brierleyi*	X90477	1460
12	*Acidianus*	*infernus*	X89852	1428
13	*Desulfurococcus*	*mobilis*	M36474	1495
14	*Metallosphaera*	*sedula*	X90481	1416
15	*Sulfolobus*	*solfataricus*	X03235	1493
16	*Thermoplasma*	*acidophilum*	M38637	1471
17	*Thermoproteus*	*tenax*	M35966	1504
	Methanogens			
18	*Methanobacterium*	*formicicum*	M36508	1476
19	*Methanobacterium*	*thermoautotrophicum*	Z37156	1483
	Bacteria			
	Cyanobacteria			
20	*Anabaena*	sp.	X59559	1489
21	*Prochlorothrix*	*hollandica*	AJ007907	1408
22	*Synechococcus*	sp.	AJ000716	1452
23	Chloroplast—*Spinacia*	*oleracea*	AJ400848	1491
	Proteobacteria			
24	*Chromatium*	*vinosum*	M26629	1524
25	*Escherichia*	*coli*	J01695	1542
26	*Rhodobacter*	*sphaeroides*	D16425	1389
27	*Salinivibrio*	*costicola*	X74699	1485
	Eucarya			
28	*Chlamydomonas*	*reinhardtii*	M32703	1791
29	*Neurospora*	*crassa*	X04971	1795
30	*Rattus*	sp.	K01593	1874
31	*Saccharomyces*	*cerevisiae*	J01353	1798
32	*Tetrahymena*	*thermophila*	M10932	1753
33	*Triticum*	*aestivum*	AJ272181	1764

the different values of the concentration (up to six, typically less), and the second sum runs over the different inhibitors, $\alpha = 1$–38.

- *Distance 2*. The absolute value of the distance is used and all data pairs are equally weighted (antibiotics with more concentrations assayed contribute more). Here $p = 1$, $K^{\alpha}(i,j) = 1$, and $A(i,j)$ is the total number of non-zero data pairs for organisms i and j.
- *Distance 3*. The squared difference between the functional responses is used and all antibiotics are equally weighted: $p = 2$, other factors as in Distance 1.
- *Distance 4*. The squared difference is used and all data pairs are equally weighted: $p = 2$ and other factors as in Distance 2.

2.4. Cluster analysis

Different multivariate analytical approaches were used for representing the matrices obtained, including the distance-based cladistic method neighbor-joining (NJ) (Saitou and Nei, 1987) and the unweighted pair

Table 2
Protein synthesis inhibitors used in this study

No.	Group and antibiotic	Functional specificity		
		TCI	PTI	TI
	Group I			
1	Althiomycin*		+	
2	Carbomycin-A*		+	
3	Gentamycin	+		
4	Griseoviridin*		+	
5	Kanamycin	+		
6	Neamine	+		
7	Neomycin	+		
8	Paromomycin	+		
9	Ribostamycin	+		
10	Streptomycin	+		
11	Thiostrepton*	+		
12	Tobramycin	+		
13	Tylosin		+	
14	Viomycin*			+
15	Virginiamycin-M		+	
	Group II			
16	α-Sarcin*	+		
17	Anisomycin*		+	
18	Cryptopleurine*			+
19	Cycloheximide*			+
20	Haemanthamine		+	
21	Harringtonine	+	+	
22	Mitogillin	+		
23	Narciclasine		+	
24	Restrictocin	+		
25	Streptimidone			+
26	Streptovitacin-A			+
27	Toxin-T2		+	
28	Tubulosine*			+
29	Tylophorine			+
	Group III			
30	Amicetin		+	
31	Anthelmycin*		+	
32	Blasticidin-S		+	
33	Edeine-A1*,[a]		+	
34	Fusidic acid*	+		
35	Hygromycin-B	+		+
36	Puromycin		+	
37	Sparsomycin*		+	
38	Tetracycline*	+		

Functional specificity shows the step of the elongation process inhibited by the antibiotic: ternary complex formation inhibitors (TCI), peptidyl-transferase inhibitors (PTI), and translocation inhibitors (TI). Fifteen antibiotics selected as a subset for developing a functional clustering are marked with asterisks.

[a] Edeine-A1 also inhibits the initiation of protein synthesis (Gale et al., 1972; Vázquez, 1979).

group method with arithmetic mean (UPGMA or 'average linkage') (Sokal and Sneath, 1973; Swofford and Olsen, 1990). SSU rRNA sequences (GenBank accession numbers and lengths are shown in Table 1) were aligned using the program CLUSTAL W (Thompson et al., 1994), and evolutionary distances were estimated by the program DNADIST in PHYLIP v3.5 package (Felsenstein, 1991) using the Kimura 2-parameter correction method that weights transitions and transversions at 2:1. The programs GROWTREE (Devereux et al., 1984), NEIGHBOR in PHYLIP, and S-Plus 2000 (Mathsoft, 2000) were used to generate functional phenograms and phylogenetic trees of the organisms.

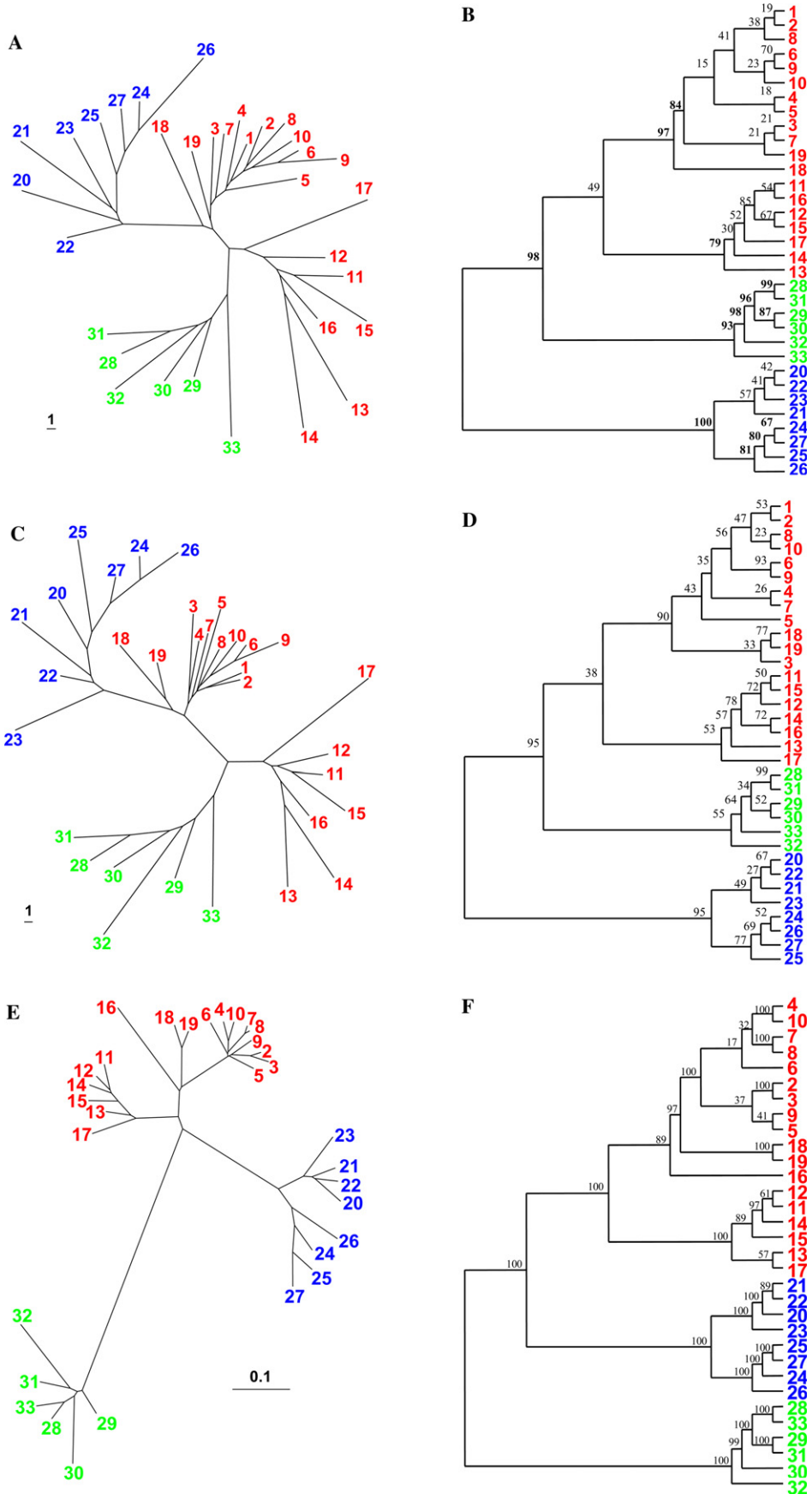
2.5. Bootstrapping coefficients

The bootstrap is a statistical technique that permits a non-parametric estimation of error (Efron, 1982; Felsenstein, 1985). The characteristics of each organism (often sequence data, here response to antibiotics) are resampled randomly with replacement. A dendrogram is constructed from each random data set and the frequency of appearance of each clade in the reference tree (derived from the original data set) is calculated. The higher the frequency of appearance of a group, the more reliable is its classification (Hedges, 1992). We have designed a procedure to assign bootstrap values to the reconstructed functional phenograms. Our method is based on Felsenstein's bootstrap test (Felsenstein, 1985), which relies on Efron's (1982) bootstrap resampling technique (see also Nei and Kumar, 2000). First, we reconstruct the UPGMA clustering with the whole data set as specified. This yields the reference phenogram. Then, for each pair of organisms, 228 data pairs are randomly chosen (that is, on the average, each data pair is chosen once, but with this resampling we effectively include each pair a number of times randomly chosen from a Poisson distribution of average unity). The corresponding clustering is reconstructed and compared with the reference phenogram. Nodes present in both of them are positively counted to compute the final bootstrap value. In SSU rRNA trees, bootstrap coefficients have been determined using the program SEQBOOT in PHYLIP v3.5 (Felsenstein, 1991). In all cases (functional and sequence analysis) each bootstrap value has been obtained as an average over 10^4 independently reconstructed clusterings, and it is given in percent.

3. Results

3.1. Tree construction

Functional clusterings (NJ and UPGMA) obtained using Distance 1, together with the calculated bootstrap values, are shown in Figs. 1A and B. Fig. 2 shows the UPGMA clusterings corresponding to Distances 2, 3, and 4, and the associated bootstrap coefficients at each node. The three major cellular lineages are always recognized, and Eucarya is linked to Archaea with an average bootstrap ranging from 98 (Distance 1) to 72% (Distance 4). Only Eucarya are reliably classified up to the last nodes, with high bootstrap coefficients maintained for all distances (the lowest value for any branch within Eucarya is 75%). In Archaea, the group formed by ex-


A
B
C
D
E
F
0.1

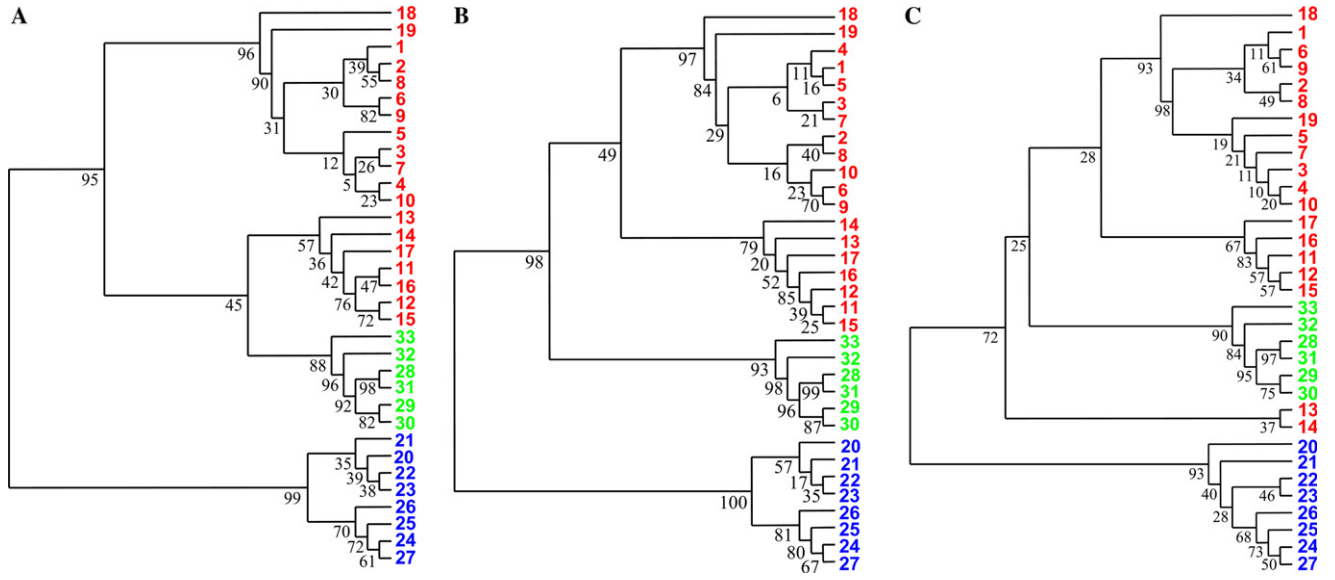


Fig. 2. UPGMA dendrograms and bootstrap coefficients corresponding to Distances 2 (A), 3 (B), and 4 (C) (see Eq. (1)).

treme halophiles plus methanogens is extremely reliable as well, with bootstraps of 96 to 97% (distances 1,2 and 3). Within this group, *Methanobacterium formicicum* is systematically classified as outgroup, with an 84 to 90% reliability. All bacteria are grouped together with an average bootstrap of 93 to 100%. The phyla *Cyanobacteria* and *Proteobacteria* are clearly identified, though the bootstrap coefficient for *Proteobacteria* is significantly higher. Distance 4 also recognizes the latter group, but does not classify *Cyanobacteria* as a group in itself (see Fig. 2C).

The most interesting fact concerns the classification of sulfur-dependent thermophiles. All of them are grouped together (except by Distance 4, which places the pair *Methallosphaera sedula* and *Desulfurococcus mobilis* as an outgroup of the rest of archaea plus eucariotes) but the position of the whole group in the phenogram changes. This can be expected from the low bootstrap coefficient that clusters this group to other. Thus, using Distances 1 or 3, sulfur-dependent thermophiles are linked to the rest of Archaea, with a bootstrap value of only 49%; with Distance 2 they are grouped with Eucarya with a bootstrap coefficient of 45%; and finally, using Distance 4 the group splits, with the larger part of it remaining linked to the rest of archaea with a bootstrap value of only 28%. Indeed, sulfur-dependent thermophiles show the lowest sensitivity to most of the protein synthesis inhibitors, and also the lowest response to changes in the concentration of the antibiotics, so it is to be expected that their classification is less robust than that for all the other organisms.

We were interested in knowing if it is feasible to define a minimal set of antibiotics that could yield a reliable clustering of organisms. The existence of such a minimal set would have the advantage of eliminating some antibiotics whose specificity or mechanism of action are apparently very similar (Amils et al., 1990; Gale et al., 1972; Vázquez, 1979), and therefore could not significantly increase the functional information or our data. Fifteen antibiotics (see Table 2) were selected to homogeneously represent the whole set. Figs. 1C and D show the functional NJ and UPGMA clusterings obtained with this subset of antibiotics. Although the internal branching order in every domain shows some differences with respect to the dendrograms in Figs. 1A and B, the main topological features are maintained. Nevertheless, bootstrap coefficients are lower than those obtained with the full set of data.

The comparison between our functional results and the clusterings obtained through molecular phylogeny techniques was carried out via SSU rRNA sequence

Fig. 1. Functional phylogeny (A–D) and phylogenetic reconstruction by means of SSU rRNA sequence comparison (E and F) of organisms. NJ trees (A and C) and UPGMA dendrograms (B and D) obtained for the organisms listed in Table 1 with the functional method developed here. Clusterings (A) and (B) are constructed using all the antibiotics listed in Table 2 and Distance 1. Most robust nodes (those with the highest bootstrap and also obtained when the other definitions of distance are used) are shown in bold face. Clusterings (C) and (D) are generated from 15 selected protein synthesis inhibitors (marked in Table 2). Note the lower absolute values of the bootstrap (see also next figure and main text). Scale bar represents one unit of functional distance. Phylogenetic reconstruction by SSU rRNA sequence comparison is shown through a NJ tree (E) and a UPGMA clustering (F); scale bar represents 0.1 nucleotide substitutions per site. Color code is red for Archaea, blue for Bacteria, and green for Eucarya. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

analysis of the selected organisms (Figs. 1E and F). A good agreement between functional clusterings and those obtained with SSU rRNA was observed, although sequence analysis systematically places Archaea closer to Bacteria than to Eukaria. In any case, it must be noticed that both functional and SSU rRNA clusterings obtained here are unrooted phylogenies.

3.2. Limit resolution of functional trees and completeness of data

A relevant question that must be addressed is the completeness of the data set, that is, if the addition of new measurements to the current set would improve the accuracy of the classification. A way to answer this question is to analyze the phenograms and the associated bootstrap coefficients when only a subset of the data is used, and then to extrapolate, if feasible, to more complete data sets. We have thus randomly selected a fraction z of all the data available and calculated the quality of the reconstruction. This defines the average bootstrap $b(z)$. For example, if $z = 0.9$, then 10% of the data available has been eliminated. With the new set, the reference tree has been reconstructed and its average bootstrap calculated, as specified above. The average bootstrap b is defined as the average over the 31 significant nodes of the reference tree (one less than the number of organisms and without considering the last node, which groups all the species together and always has bootstrap 100). Note that for any value of $z < 1$ there are many possibilities to select a subset of data. Hence, we have repeated the procedure 100 times, each time averaging over 100 clusterings to get the bootstrap for each data subset, such that the final values are averaged over 10^4 dendrograms. This analysis has been carried out for the four different measures of distance discussed, and the results are represented in Fig. 3.

The function $b(z)$ can be numerically estimated by fitting the resulting curves. The question to be answered is then the asymptotic value b^∞ to which it tends when the available data increases. Although the ideal situation would be to get a perfect classification of organisms when the amount of data available diverges (that is, $b(z) \to 100$ if $z \to \infty$), it is possible that b^∞ saturates at a lower value, and that the method here introduced is not precise enough to return a perfect classification (within the limit discussed above). We observe that the average bootstrap steadily increases for the increasing amount of data considered, and it still does not seem to reach saturation. From our numerical results, we make the ansatz that the function $b(z)$ is

$$b(z) - b^\infty = kz^{-\beta}, \tag{2}$$

where b^∞ is the asymptotic value of the function and the exponent β and the constant k have to be fixed with a least-squares fit to the numerical data. It turns out that

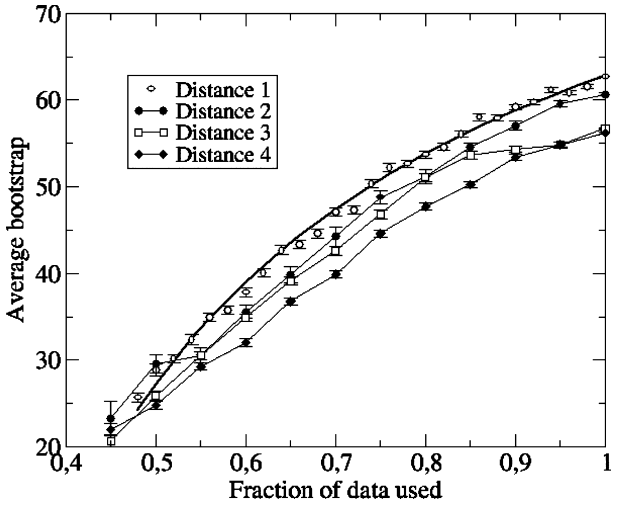


Fig. 3. Average bootstrap of the functional dendrograms when only a fraction z of the data available are used. The curves corresponding to the four different definitions of distance introduced in the text are shown. We observe a comparable performance in their ability to classify the organisms. The bold line is a best fit to the data obtained from Distance 1: $b(z) = 100 - 37/z$. Error bars correspond to the standard deviation over 100 different data subsets with the same value of z. For each of them, the average bootstrap was calculated from 100 independent trees.

the parameters of Eq. (2) that return the best fit (that with the highest correlation coefficient, $r^2 = 0.996$) are $b^\infty = 100 \pm 1$, $k = 37 \pm 1$, and $\beta = 0.97 \pm 0.01$. This curve is shown in Fig. 3 together with our numerical data. Now, by using the approximate function $b(z) = 100 - 37/z$ it is possible to estimate the improvement in $b(z)$ when z increases. For example, if $z = 1$ (the whole set of available data), $b(1) = 63$. If only half of the data were available, then $b(0.5) = 26$, and the classification of the analyzed organisms would be simply impossible. Now, if we doubled our databank to get $z = 2$, the average bootstrap would be expected to rise to $b(2) = 81.5$. Three times as many data would already yield an average bootstrap close to that of the rRNA tree. Therefore, in principle, the functional classification can be as robust as wished.

Any increase in the amount of data available seems thus to translate into a better classification of the organisms. Hence, we conclude that there is little to no redundancy in the antibiotics used, and apparently each of them seems to add new information. This is one reason to claim that there is no 'minimum' set of antibiotics able to retrieve a completely reliable classification, and partly explains the lesser classificatory power of the abovementioned 15 'representative' antibiotics (Figs. 1C and D). The curves $b(z)$ also reveal, on the one hand, the equivalent performance of any of the definitions of distance used (thus the independence of the classification from that definition), and on the other that the limit resolution of the method has not yet been reached.

4. Discussion

It has been shown that the sensitivity of ribosomal systems to antibiotics is an effective marker with which to reconstruct meaningful dendrograms. The overall topology of the clusterings obtained is remarkably similar to those produced by SSU rRNA sequence comparison. The 'universal' phenogram (Figs. 1A and B) reveals a clear separation among the three cellular lineages, in agreement with the clusterings based on ribosomal RNA (Pace, 1997; Woese et al., 1990) and gene content (Fitz-Gibbon and House, 1999; Snel et al., 1999; Tekaia et al., 1999). It is also consistent, in terms of relative distances, with results indicating the sisterhood relationship of Archaea with Eucarya as shown by rooted trees based on paralogous genes (Gogarten et al., 1989; Iwabe et al., 1989) or metabolic pathways (Podani et al., 2001). Moreover, this sisterhood agrees with previous studies involving antibiotic sensitivity, that favored the functional similarities existing between the translational apparatus of eucaryotes and archaea (Zillig, 1987).

The comparison of dendrograms obtained with the whole set of 38 antibiotics (or a 'representative' subset) and four different definitions of distance (Figs. 1A–D and 2) demonstrates the phylogenetic consistency of functional clustering, as quantified by the bootstrap coefficients. A higher robustness is achieved with the use of the entire set of data, in concordance with the analysis performed on the limit resolution of our functional trees (Fig. 3). Among Archaea, sulfur-dependent thermophiles constitute a clade clearly separated from extreme halophiles plus methanogens, in agreement with SSU rRNA trees. Halophilic archaea show the closest branching order in every analysis performed, illustrating how the high sequence similarity displayed by their rRNAs (Briones and Amils, 2000; Kamekura and Dyall-Smith, 1995) correlates with their ribosomal function. Methanogens are not clustered together in Figs. 1A and 2; this inconsistent classification within archaea is probably due to the low number of methanogenic ribosomal systems tested. In Bacteria, a clear separation is observed between the phyla *Cyanobacteria* and *Proteobacteria*, consistent with SSU rRNA data. The inclusion of the *Spinacia oleracea* chloroplast in the cyanobacterial cluster indicates that the ribosomes of the chloroplast maintain the sensitivity pattern of the bacterial group from which they originated by endosymbiosis (Gray, 1992; Margulis, 1976).

At present, the dendrograms here calculated have a lower resolution than those resulting from rRNA cladistic analysis. However, the similarity of the results obtained with our functional analysis and the sequence comparison approach clearly shows the value of the inhibition data to study macro- and mesophylogeny of organisms. Moreover, it cannot be ruled out that for a large enough data set, phylogeny at the level of species could also be established in a robust way. To understand the basis of the relationship between SSU rRNA and functional phylogenies one must recall that antibiotics are small molecules in comparison to the ribosomal particle, and that they act as specific functional effectors for the protein synthesis process. The analysis of point mutations in the rRNA which confer resistance to some inhibitors, and the use of footprinting techniques to reveal the interaction sites of antibiotics (Christiansen et al., 1990; Rodríguez-Fonseca et al., 1995), have indicated that most of these target sites are located at the very conserved, taxonomically consistent, unpaired regions of the rRNAs (Dixon and Hillis, 1993). Furthermore, it is reasonable to assume that the ribosomal 'functional space' with which antibiotics interact is not only dependent on these short sequences, but also on the three-dimensional contacts among such 'functional loops' and other regions of the rRNAs, as well as with ribosomal proteins. Moreover, the large amount of antibiotics tested in our study (together with the different steps of the translation process they inhibit, see Table 2) allows to explore a large fraction of the functional space of the ribosome. Therefore, the strong phylogenetic signal here obtained does not result from any single ribosome-inhibitor pair but from the suitable combination of many antibiotic sensitivities. The phenotypic response analyzed should be seen as the outward manifestation of the evolutionary history underlying the evolution of the translational apparatus. This must also be considered when comparing the detailed topology of the dendrograms obtained with molecular and functional data. Neutral substitutions in RNA sequences increase the distance between organisms in molecular trees, but are not reflected in the functional clustering. On the other hand, a single relevant mutation in rRNA (i.e., in the above-mentioned 'functional loops') has a negligible effect on sequence distance, whilst it can produce important changes in ribosomal function and result in a larger functional distance. This could explain the observed internal branching rearrangements within the major groups when comparing functional and molecular dendrograms.

Protein synthesis inhibitors are acting as functional markers of the quaternary structure of the ribosomes, and the phylogenetically consistent differences in the inhibition patterns reflect the evolution of the ribosomal particle. Such structural–functional history is correlated with the progressive sequence divergence of its constituent macromolecules. The existence of antibiotic-binding sites maintained in all ribosomal systems suggests that the basic components of the translational machinery have been preserved throughout evolution. The present-day spectrum of sensitivities reflect the result of a 'fine tuning' of the ribosomal function in different organisms and therefore constitutes a record of their evolutionary history.

The clusterings derived here are similar in topology not only to the SSU rRNA tree, but also to the genome trees based on whole gene content (Fitz-Gibbon and House, 1999; Snel et al., 1999; Tekaia et al., 1999; Wolf et al., 2002), on the presence of ortholog sequences (Lin and Gerstein, 2000; Wolf et al., 2001), and on metabolic and information transfer pathways (Podani et al., 2001). The congruous results obtained by means of these different methodologies indicate that although horizontal gene transfer has played a major role in cellular evolution, massive lateral transfer events between distant groups have not obliterated the identity of the three major domains of cell descent. The hierarchical organization of species as indicated by their antibiotic sensitivities is the outcome of a branching process from common ancestors shared by the different organisms included in our dataset. Hence, the structural–functional information provided by our analysis exploits that relationship and constitutes a useful tool with which to study the evolution of organisms.

Further improvements in our method could include the expansion of our databank into the eukaryotic branch to show its inter-kingdom resolution ability. Also, the inclusion of data from selected microorganisms could improve resolution within the bacterial phyla. The use of this method should be tested for its ability to extract phylogenetic information that could be stored in other kinds of selected inherited phenotypic (biochemical, morphological or physiological) data that can increase our list of reliable evolutionary and taxonomic markers.

Acknowledgments

The authors thank K. Koroutchev for a preliminary analysis of the data, and E. González-Toril and E. Viguera for helpful comments on the manuscript. Work at the Centro de Astrobiología was supported by the European Union, the Autonomous Government of Madrid, and the Instituto Nacional de Técnica Aerospacial. Work at the Centro de Biología Molecular was supported by Grant BIO99-0184 of the Spanish Ministry of Education and Culture. S.C.M. was supported by a Ramón y Cajal contract of the Spanish Ministry of Science and Technology. Support from DGAPA/UNAM to A.L. is gratefully acknowledged.

References

Altamura, S., Sanz, J.L., Amils, R., Cammarano, P., Londei, P., 1995. The antibiotic sensitivity spectra of ribosomes from the thermoproteales: phylogenetic depth and distribution of antibiotic binding sites. Syst. Appl. Microbiol. 10, 218–225.

Amils, R., Sanz, J.L., 1986. Inhibitors of protein synthesis as phylogenetic markers. In: Hardesty, B., Kramer, G. (Eds.), Structure, Function and Genetics of Ribosomes. Springer-Verlag, New York, NY, pp. 605–620.

Amils, R., Ramírez, L., Sanz, J.L., Marín, I., Pisabarro, A.G., Ureña, D., 1989. The use of functional analysis of the ribosome as a tool to determine archaebacterial phylogeny. Can. J. Microbiol. 35, 141–147.

Amils, R., Ramírez, L., Sanz, J.L., Marín, I., Pisabarro, A.G., Sánchez, E., Ureña, D., 1990. Phylogeny of antibiotic action. In: Hill, W.E., Dahlberg, A., Garret, R.A., Moore, P.B., Schlessinger, D., Warner, J.R. (Eds.), The Ribosome: Structure, Function and Evolution. American Society for Microbiology, Washington, DC, pp. 645–654.

Briones, C., Amils, R., 2000. Nucleotide sequence of the 23S rRNA from *Haloferax mediterranei* and phylogenetic analysis of halophilic archaea based on LSU rRNA. Syst. Appl. Microbiol. 23, 124–131.

Briones, C., Koroutchev, K., Amils, R., 1998. Functional phylogeny: the use of the sensitivity of ribosomes to protein synthesis inhibitors as a tool to study the evolution of organisms. Orig. Life. Evol. Biosph. 28, 571–582.

Bromham, L., Penny, D., 2003. The modern molecular clock. Nat. Rev. Genet. 4, 216–224.

Cammarano, P., Teichner, A., Londei, P., Acca, M., Nicolaus, B., Sanz, J.L., Amils, R., 1985. Insensitivity of archaebacterial ribosomes to protein synthesis inhibitors. Evolutionary implications. EMBO J. 4, 811–816.

Casquero, I., 1992. Filogenia funcional: estudio de sensibilidad a antibióticos de ribosomas eucarióticos y de bacterias y orgánulos fotosintéticos. Ph.D. thesis. Universidad Autónoma de Madrid, Madrid.

Christiansen, J., Egebjerg, J., Larsen, N., Garrett, R.A., 1990. Analysis of rRNA structure: experimental and theoretical considerations. In: Spedding, G. (Ed.), Ribosomes and Protein Synthesis: A Practical Approach. Oxford University Press, Oxford, pp. 229–252.

Clarke, G.D., Beiko, R.G., Ragan, M.A., Charlebois, R.L., 2002. Inferring genome trees by using a filter to eliminate phylogenetically discordant sequences and distance matrix based on mean normalized BLASTP scores. J. Bacteriol. 184, 2072–2080.

Devereux, J., Haeberly, P., Smities, O., 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12, 387–395.

Dixon, M.T., Hillis, D.M., 1993. Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. Mol. Biol. Evol. 10, 256–267.

Doolittle, W.F., 1999. Phylogenetic classification and the universal tree. Science 284, 2124–2129.

Efron, B., 1982. The Jackknife, The Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Elhardt, D., Böck, A., 1982. An in vitro polypeptide synthesizing system from methanogenic bacteria: sensitivity to antibiotics. Mol. Gen. Genet. 188, 128–134.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Felsenstein, J., 1991. PHYLIP (Phylogeny Inference Package). Version 3.5. University of Washington, Seattle, WA.

Fitz-Gibbon, S.T., House, C.H., 1999. Whole genome-based phylogenetic analysis of free-living microorganisms. Nucleic Acids Res. 27, 4218–4422.

Gale, E.F., Cundliffe, E., Reynolds, P.E., Richmond, M.H., Waring, M.J., 1972. The Molecular Basis of Antibiotic Action. Wiley, London.

Gogarten, J.P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E.J., Bowman, B.J., Manolson, M.F., Poole, R.J., Date, T., Oshima, T., 1989. Evolution of the vacuolar H+-ATPase: implications for the origin of eukaryotes. Proc. Natl. Acad. Sci. USA 86, 6661–6665.

Gray, M., 1992. The endosymbiont hypothesis revisited. Int. Rev. Cytol. 141, 233–357.

Hartigan, J.A., 1975. Clustering Algorithms. Wiley, New York, NY.

Hedges, S.B., 1992. The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. Mol. Biol. Evol. 9, 366–369.

Hillis, D.M., 1987. Molecular versus morphological approaches to systematics. Annu. Rev. Ecol. Syst. 18, 23–42.

Huynen, M., Snel, B., Bork, P., 1999. Lateral gene transfer, genome surveys, and the phylogeny of prokaryotes. Science 286, 1443.

Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., Miyata, T., 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. USA 86, 9355–9359.

Kamekura, M., Dyall-Smith, M.L., 1995. Taxonomy of the family *Halobacteriaceae* and the description of two new genera *Haloru-brobacterium* and *Natrialba*. J. Gen. Appl. Microbiol. 41, 330–350.

Kurland, C.G., Canback, B., Berg, O.G., 2003. Horizontal gene transfer: a critical view. Proc. Natl. Acad. Sci. USA 100, 9658–9662.

Lin, J., Gerstein, M., 2000. Whole-genome trees based on the occurrence of folds and orthologs: implications for comparing genomes on different levels. Genome Res. 10, 808–810.

Margulis, L., 1976. A review: genetic and evolutionary consequences of symbiosis. Exp. Parasitol. 39, 277–349.

MathSoft, Inc. S-Plus, 2000. Data Analysis Products Division. Seattle, WA, 2000. Available from: <http://www.insightful.com>.

Mayr, E., 1998. Two empires or three?. Proc. Natl. Acad. Sci. USA 95, 9720–9723.

Moritz, C., Hillis, D.M., 1990. Molecular systematics: context and controversies. In: Hillis, D.M., Moritz, C. (Eds.), Molecular Systematics. Sinauer Associates, Sunderland, MA, pp. 3–10.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford.

Nesbo, C.L., L'Haridon, S., Stetter, K.O., Doolittle, W.F., 2001. Phylogenetic analyses of two "archaeal" genes in *Thermotoga maritima* reveal multiple transfers between archaea and bacteria. Mol. Biol. Evol. 18, 362–375.

Ochman, H., Lawrence, J.G., Groisman, E.A., 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405, 299–304.

Oliver, J.L., Sanz, J.L., Amils, R., 1987. Inferring the phylogeny of Archaebacteria: the use of ribosomal sensitivity to protein-synthesis inhibitors. J. Mol. Evol. 24, 281–288.

Pace, N.R., 1997. A molecular view of microbial diversity and the biosphere. Science 276, 734–740.

Perez-Luz, S., Rodriguez-Valera, F., Lan, R., Reeves, P.R., 1998. Variation of the ribosomal operon 16S–23S gene spacer region in representatives of *Salmonella enterica* subspecies. J. Bacteriol. 180, 2144–2151.

Podani, J., Oltvai, Z.N., Jeong, H., Tombor, B., Barabási, A.-L., Szathmári, E., 2001. Comparable system-level organization of Archaea and Eukaryotes. Nat. Genet. 29, 54–56.

Rodríguez-Fonseca, C., Amils, R., Garrett, R.A., 1995. Fine structure of the peptidyl transferase center on 23S-like rRNAs deduced from chemical probing of antibiotic-ribosome complexes. J. Mol. Biol. 247, 224–235.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Sánchez, E., Teixidó, J., Guerrero, R., Amils, R., 1994. Hypersensitivity of *Rhodobacter sphaeroides* ribosomes to protein synthesis inhibitors: structural and functional implications. Can. J. Microbiol. 40, 699–704.

Sanz, J.L., Huber, G., Huber, H., Amils, R., 1994. Using protein synthesis inhibitors to establish the phylogenetic relationships of the Sulfolobales order. J. Mol. Evol. 39, 528–532.

Sanz, J.L., Marín, I., Ureña, D., Amils, R., 1992. Functional analysis of seven ribosomal systems from extremely halophilic archaea. Can. J. Microbiol. 39, 311–317.

Sanz, J.L., Marín, I., Balboa, M.A., Ureña, D., Amils, R., 1988. An NH_4^+ dependent protein synthesis cell-free system for halobacteria. Biochemistry 27, 8194–8199.

Snel, B., Bork, P., Huynen, M.A., 1999. Genome phylogeny based on gene content. Nat. Genet. 21, 108–110.

Sokal, R., Sneath, P., 1973. Numerical Taxonomy. Freeman, San Francisco, CA.

StatSoft, Inc, Electronic Statistics Textbook. Tulsa, OK, 2002. StatSoft. Available from: <http://www.statsoft.com/textbook/stathome.html>.

Swofford, D.L., Olsen, G.J., 1990. Phylogeny reconstruction. In: Hillis, D.M., Moritz, C. (Eds.), Molecular Systematics. Sinauer Associates, Sunderland, MA, pp. 411–501.

Tekaia, F., Lazcano, A., Dujon, B., 1999. The genomic tree as revealed from whole proteome comparisons. Genome Res. 9, 550–557.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Tzagoloff, A., Macino, G., Sebald, W., 1979. Mitochondrial genes and translation products. Annu. Rev. Biochem. 48, 419–441.

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.

Vázquez, D., 1979. Inhibitors of Protein Biosynthesis. Springer-Verlag, Berlin.

Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, P., Krichevsky, M.I., Moore, L.H., Moore, W.E.C., Murray, R.G.E., Stackebrandt, E., Starr, M., Trüper, H.G., 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. 37, 463–464.

Woese, C.R., 1987. Bacterial evolution. Microbiol. Rev. 51, 221–271.

Woese, C.R., 2000. Interpreting the universal phylogenetic tree. Proc. Natl. Acad. Sci. USA 97, 8392–8396.

Woese, C.R., Kandler, O., Wheelis, M.L., 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eukarya. Proc. Natl. Acad. Sci. USA 87, 4576–4579.

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

Wolf, Y.I., Rogozin, I.B., Grishin, N.V., Tatusov, R.L., Koonin, E.V., 2001. Genome trees constructed using five different approaches suggest new major bacterial clades. BMC Evol. Biol. 1, 8.

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

Zillig, W., 1987. Eukaryotic traits in Archaebacteria. Could the eukaryotic cytoplasm have arisen from archaebacterial origin?. Ann. N. Y. Acad. Sci. 503, 78–82.