J Mol Evol (1995) 41:366–375

JOURNAL OF MOLECULAR EVOLUTION

Evolution According to Large Ribosomal Subunit RNA

Peter De Rijk, Yves Van de Peer, Ilse Van den Broeck, Rupert De Wachter

Departement Biochemie, Universiteit Antwerpen (UIA), Universiteitsplein 1, B-2610 Antwerpen, Belgium

Received: 15 December 1994 / Accepted: 28 February 1995

Abstract. Evolutionary trees were constructed, by distance methods, from an alignment of 225 complete large subunit (LSU) rRNA sequences, representing Eucarya, Archaea, Bacteria, plastids, and mitochondria. A comparison was made with trees based on sets of small subunit (SSU) rRNA sequences. Trees constructed on the set of 172 species and organelles for which the sequences of both molecules are known had a very similar topology, at least with respect to the divergence order of large taxa such as the eukaryotic kingdoms and the bacterial divisions. However, since there are more than ten times as many SSU as LSU rRNA sequences, it is possible to select many SSU rRNA sequence sets of equivalent size but different species composition. The topologies of these trees showed considerable differences according to the particular species set selected.

The effect of the dataset and of different distance correction methods on tree topology was tested for both LSU and SSU rRNA by repetitive random sampling of a single species from each large taxon. The impact of the species set on the topology of the resulting consensus trees is much lower using LSU than using SSU rRNA. This might imply that LSU rRNA is a better molecule for studying wide-range relationships. The mitochondria behave clearly as a monophyletic group, clustering with the Proteobacteria. Gram-positive bacteria appear as two distinct groups, which are found clustered together in very few cases. Archaea behave as if monophyletic in most cases, but with a low confidence.

Key words: Large ribosomal subunit RNA — Small ribosomal subunit RNA — Archaea — Bacteria — Eucarya — Plastids — Mitochondria

Introduction

More than 3,000 sequences of small subunit ribosomal RNA (SSU rRNA) are now available (Van de Peer et al. 1994b). This molecule has been used extensively for study: of relationships among all different life forms (Cedergren et al. 1988; Van de Peer et al. 1990) and of evolution within the bacterial and archaeal domains (Woese 1987; Olsen et al. 1994), the eukaryotic domain (e.g., Hendriks et al. 1988; Sogin 1989), and within eukaryotic kingdoms (e.g., Field et al. 1988; Wilmotte et al. 1993). Its functional constancy and universal occurrence make it exceptionally suitable as a molecular clock (Woese 1987). It shows an alternation of conserved and variable sequence areas, which in principle makes it appropriate for analyses on different evolutionary time scales. Its chain length is sufficiently large for the calculation of statistically meaningful distances.

Large ribosomal subunit RNA (LSU rRNA) shares these advantages with SSU rRNA. Moreover it has the added bonus of an even longer chain length, and thus might be more reliable for inferring phylogenies. However, the smaller number of known complete LSU sequences and the larger differences in length and variability have thwarted acquisition of a good alignment encompassing all sequences. Therefore LSU rRNA has been used more often to find branching patterns between closely related species (e.g., Baroin-Tourancheau et al.

Abbreviations: LSU rRNA, large subunit ribosomal RNA; SSU rRNA, small subunit ribosomal RNA; JC, Jukes and Cantor; JN, Jin and Nei

Correspondence to: R. De Wachter

1992) than to get a general overview of the biological evolution. An exception is the tree published by Cedergren et al. (1988), which covers all forms of life. However, it was limited to 41 species, and was based on a partial alignment restricted to 619 positions.

The number of known LSU sequences has been growing steadily since. As more sequences have become available, the knowledge about the secondary structure of the molecule has also gradually improved. By now we have a reliable alignment of the sequences for 225 different species covering Bacteria, Archaea, Eucarya, plastids, and mitochondria (De Rijk et al. 1994). At several points hypervariable stretches occur in a more conserved core sequence. These are responsible for the increased size of eukaryotic LSU rRNAs with respect to their bacterial equivalents. Depending on the author, they are called D regions (Hassouna et al. 1984) or expansion segments (Clark et al. 1984). The structure and function of these areas might very well be specific for different groups of organisms. For some of these areas a satisfactory general alignment could not be found. However, the part where the alignment is considered reliable still covers 2,411 positions, which should provide enough data for the construction of meaningful evolutionary trees. For 172 of the species represented in the LSU rRNA alignment, SSU rRNA sequences are also known. This makes a comparison between these two molecular clocks possible.

Methods

Alignments and Positions Used. The LSU rRNA sequence alignment used in this study is essentially the one that has been reported (De Rijk et al. 1994), and it can be obtained by anonymous ftp. It was constructed by means of the program DCSE (De Rijk and De Wachter 1993), using an automatic alignment approach, with manual checking and editing. Homology in primary as well as secondary structure was taken into account in the alignment process. The secondary structure model is illustrated in Fig. 1 with *Saccharomyces cerevisiae* LSU rRNA. A few sequences, which might possibly be of doubtful quality judging from their alignment, were omitted for this study. Also omitted were the kinetoplast sequences, which bear little resemblance to the sequences found in other mitochondria. Conversely, a few sequences that became available since publication of the alignment have been added. When more than one sequence for a given species is known, only one was used. The resulting LSU rRNA dataset contains sequences from 38 Eucarya, 16 Archaea, 64 Bacteria, 31 plastids, and 76 mitochondria.

Regions that could not be satisfactorily aligned in some of the species used were not used for tree construction. They are indicated in Fig. 1 on the sequence of *Saccharomyces cerevisiae*. These regions correspond mainly to the highly variable regions. The remaining area, defined as area 2, is considered to be reliably aligned. It contains a large part of the core structure found in almost all eukaryotic and prokaryotic LSU rRNAs. It covers 2,411 alignment positions, and corresponds to 65% of the *Escherichia coli* sequence. Several mitochondrial sequences have deletions even in this core. Therefore a smaller area 1 was also defined, consisting of positions occupied by a nucleotide in nearly every sequence in the alignment. This area, also indicated in Fig. 1, spans 1,075 positions.

The SSU rRNA alignment used in this study has also been published and made available by anonymous ftp (Van de Peer et al. 1994b). As in the case of LSU rRNA, only the reliably aligned areas were used for tree construction. These correspond to the following nucleotides in the sequence of Saccharomyces cerevisiae SSU rRNA: 4–59, 89–127, 139–178, 290–307, 380–509, 546–634, 860–1043, 1069–1215, 1262–1343, 1376–1393, 1399–1486, 1492–1669, and 1729–1795. The number of species common to the LSU and SSU rRNA datasets amounts to 172.

Tree Construction Methods. All trees were constructed by distance methods, since the large size of the datasets does not lend itself to analysis by parsimony methods. Dissimilarity values were computed on the basis of the LSU and SSU rRNA alignments and converted into evolutionary distances. Different equations are available for correcting for multiple mutations per site and thus converting dissimilarity into evolutionary distance. The equation of Jukes and Cantor (1969) assumes that all substitutions are equally likely and all sites equally variable. The equation of Kimura (1980) also assumes equal variability of sites but accounts for transversions and transitions separately. In practice, however, trees constructed from rRNA sequence alignments have very similar topologies regardless of whether distances are obtained according to Jukes and Cantor (1969) or Kimura (1980). A more serious error may be introduced by the assumption that the probability of substitution is the same at all sites, whereas it has been shown (Van de Peer et al. 1993) in the case of SSU rRNA that the substitution rate of the most variable sites is about 1,000 times the rate of the least variable sites. Golding (1983) has shown that applying the Jukes-Cantor correction to sequences composed of sites with unequal evolutionary rates leads to an underestimation of large evolutionary distances with respect to smaller ones. As a result, distant species may seem to be closer to each other than they actually are and this can cause artificial clustering of long branches (Olsen 1987). This artefact can be avoided by converting dissimilarities into evolutionary distances according to Jin and Nei (1990). These authors assume that there is a gamma distribution of substitution rates over the sequence positions. In the present work a set of trees was constructed on the basis of distances computed according to Jukes and Cantor (1969), while another set was constructed on the basis of distances computed by the equation (Rzhetsky and Nei 1994)

$$d = \frac{3}{4} a \left[\left(1 - \frac{4}{3} f \right)^{-1/a} - 1 \right]$$

where f is the fraction of different nucleotides observed on comparison of two sequences, and d is the estimated number of nucleotide substitutions per site that gave rise to this dissimilarity. The parameter a was set to 1 (Nei 1991). Insertions and deletions were not taken into account in the calculation of evolutionary distances.

Computation of distance matrices, construction of evolutionary trees, bootstrap analysis (Felsenstein 1985), and the computation of consensus trees were done with the software package TREECON (Van de Peer and De Wachter 1993, 1994).

Results

The Complete LSU Tree

Evolutionary trees were constructed for all different species in the LSU rRNA dataset with distances computed according to Jin and Nei (1990) (JN). This was done for two alignments, corresponding to areas 1 and 2 as defined in the Methods section and Fig. 1. The tree based

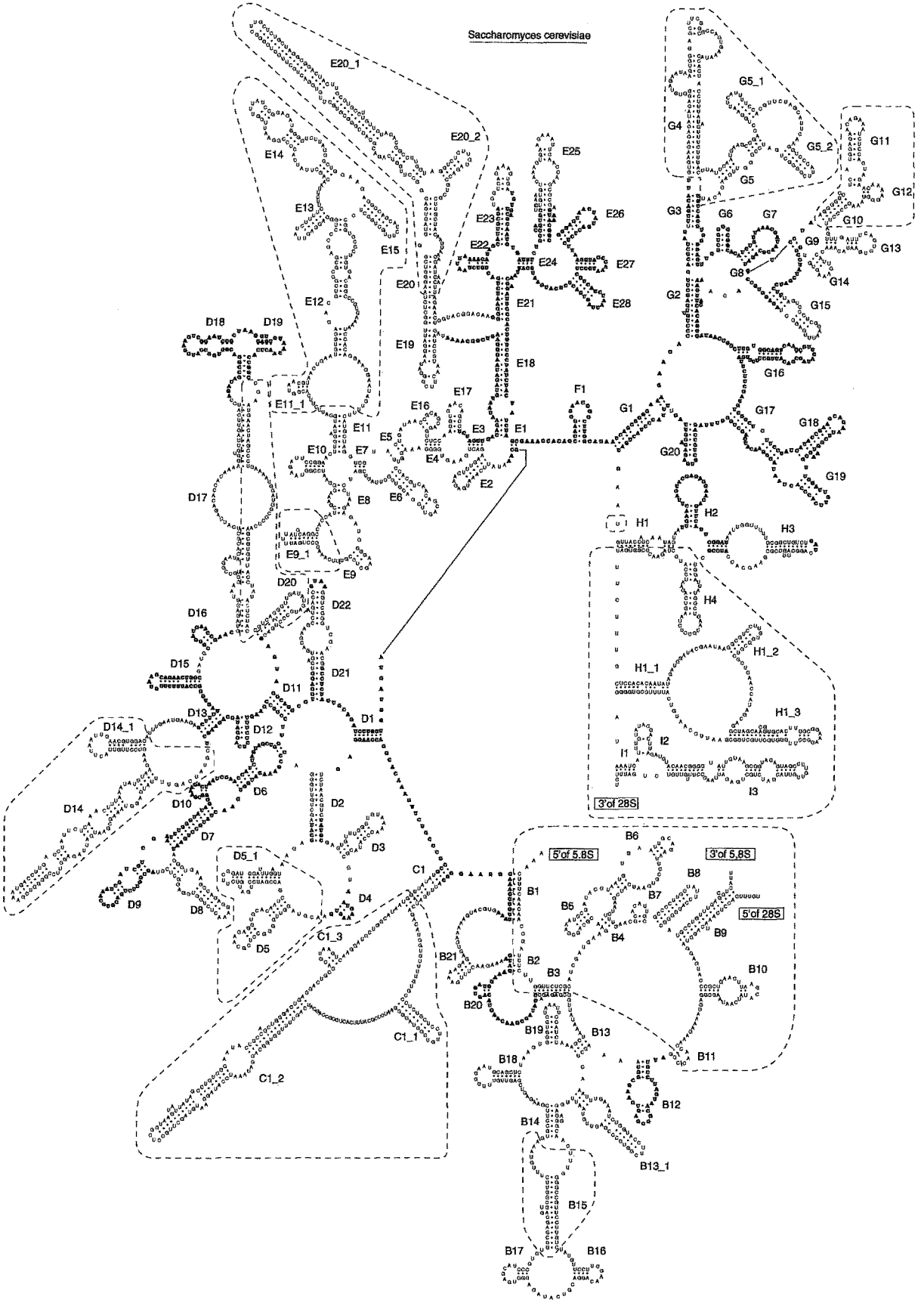

Fig. 1. Secondary structure model for *Saccharomyces cerevisiae* LSU rRNA according to De Rijk et al. (1994). Areas enclosed in a *broken line* were not used in the present study, since they might not be properly aligned in some of the species used. The remaining area was used for phylogenetic tree construction and is referred to as area 2. Area 1 comprises all characters shown in a *bold typeface.*

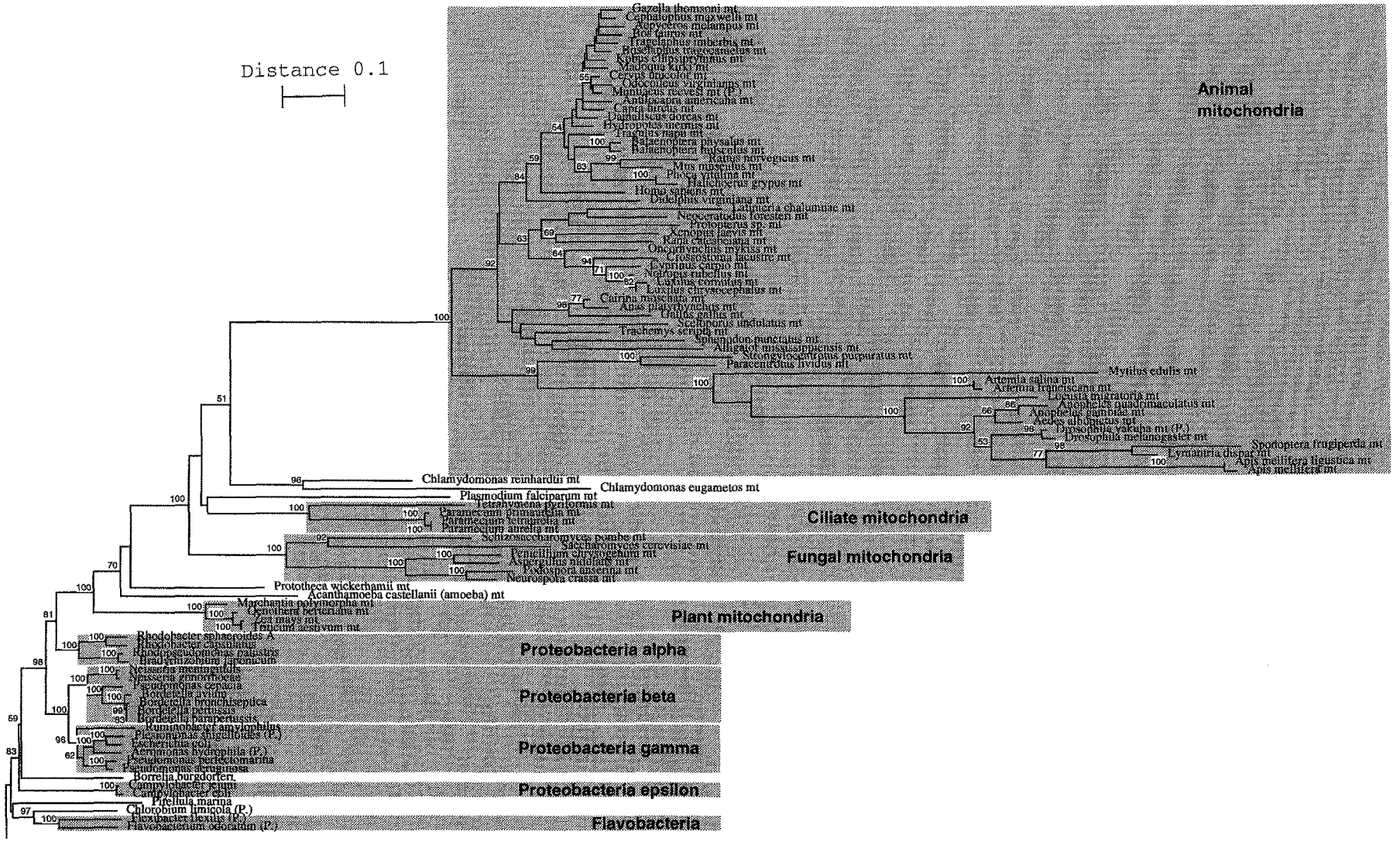

Fig. 2. LSU tree constructed on area 2 of the entire set of 225 LSU rRNA sequences using the Jin and Nei distance correction method. The major groups used in the resampling tests (Fig. 5) are indicated by a *gray area* with the name of the group. Archaea and Eucarya subtrees are indicated as *triangles*. They are shown in more detail in Figs. 3 and 4.

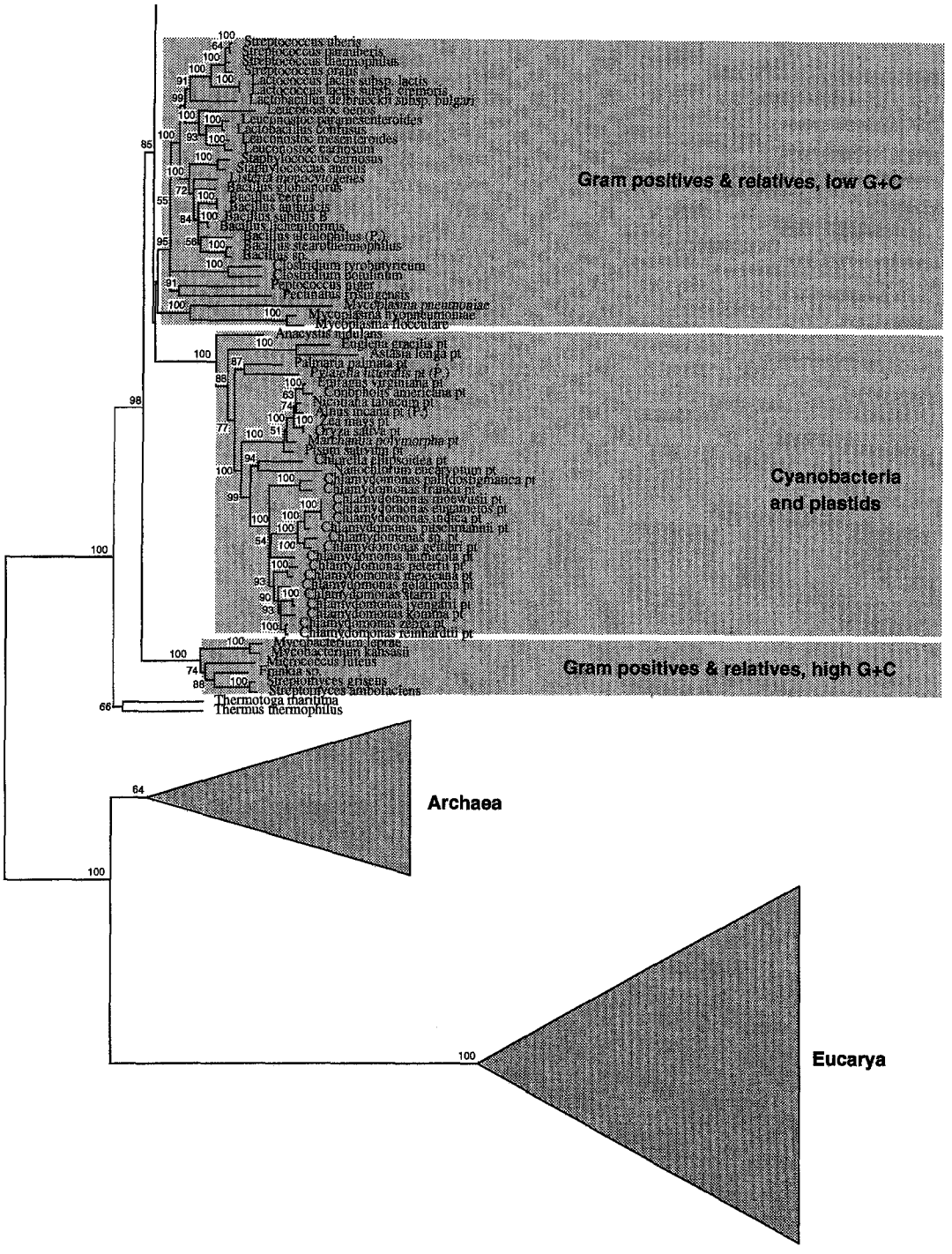

Fig. 2. Continued.

on area 2 is shown in Fig. 2. The Archaea and Eucarya are summarised as triangles. The root was placed, somewhat arbitrarily, between the Bacteria and Archaea/Eucarya since there is some evidence that Archaea and Eucarya may be sister groups (Iwabe et al. 1989). The tree based on area 1, which is not shown, has a very similar topology, but distances between branching points are generally shorter and bootstrap values are often lower. Differences in topology between the two trees are mainly in branching patterns poorly supported by bootstrap analysis. Separate trees made for the Archaea and Eucarya are discussed next.

A striking property of the complete LSU tree (Fig. 2) is the large difference in branch lengths. While branches are relatively short in Bacteria and Archaea, they are longer in the Eucarya, and very long in the Mitochondria. The insect mitochondria show extremely long branches. These differences properly illustrate the difference in evolutionary rate in these organisms and organelles.

In both the tree shown in Fig. 2 and the tree con-

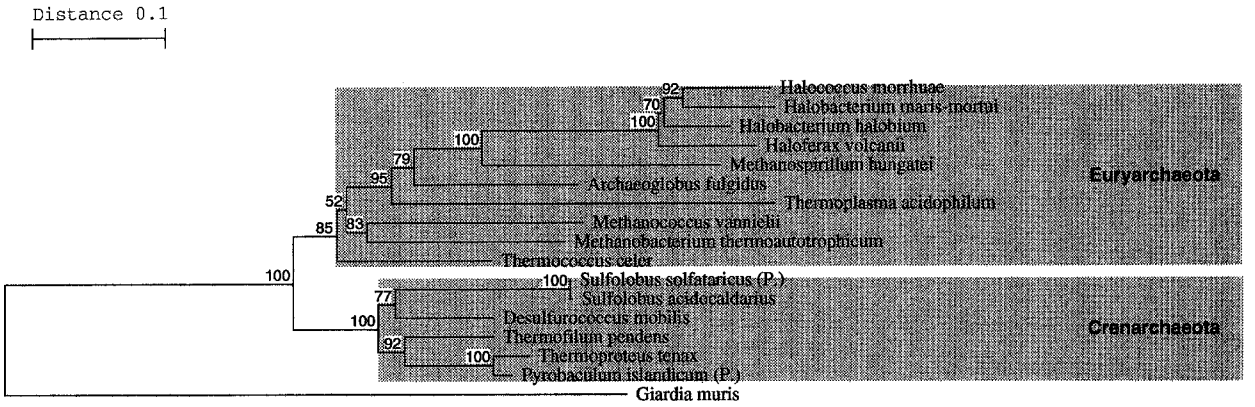

Fig. 3. The archaeal tree was constructed using the Jin and Nei distance correction method and using the eukaryote *Giardia muris* (Diplomonada) as an outgroup.

structed on the smaller set of positions of area 1, the mitochondria form a monophyletic cluster with the Proteobacteria subgroup α. Distinct clusters can be seen of plant, fungal, and animal mitochondria, but the protoctist mitochondria are not monophyletic. The plant mitochondria branch first, followed by those of *Acanthamoeba castellanii* and the alga *Prototheca wickerhamii.* The following branches lead to the fungal mitochondria, the ciliate mitochondria, *Chlamydomonas* mitochondria, and animal mitochondria. In the tree on the area 1 alignment, the ciliate mitochondria branch before the fungal mitochondria, but with very low bootstrap values. As expected, the plastids cluster with the cyanobacterium *Anacystis nidulans.*

The bacterial species form clusters corresponding to the major bacterial taxa described by Woese (1987). However, the low G + C and high G + C subdivisions of the Gram-positive bacteria do not form together a monophyletic cluster. The mycoplasmas cluster with the Gram positives of low G + C content, though with a very low bootstrap value. This low value may be an artefact though, caused by the larger evolutionary rate of the mycoplasmas (Woese 1987). The deepest bacterial branch is formed by species belonging to the Thermotogales and the radioresistant micrococci, the latter represented by *Thermus thermophilus.* These species cluster together in the tree based on area 2, while they do not in the tree based on area 1. The order of divergence of the other taxa cannot be deduced reliably from the trees, since the internodal branches are very short and the bootstrap values low.

The Archaea and Eucarya Subtrees

A tree for the Archaea is shown in Fig. 3. It was reconstructed from the area 2 alignment, on a dataset containing only the Archaea and a single eukaryotic outgroup organism. The reason is that topological distortions due to differences in evolutionary rate within a cluster can be exacerbated by the use of very distant outgroup species, as is the case in the complete LSU tree. The topology of the Fig. 3 tree indeed differs somewhat from that found for the Archaea in the complete LSU tree. However, shifts in topology are restricted to branches poorly supported by bootstrap analysis. The archaeal tree shows two clusters representing the major archaeal groups: the Crenarchaeota and the Euryarchaeota (Olsen et al. 1994). The topology of the tree based on the shorter area 1 alignment for the same set of species (not shown) is identical.

Figure 4 shows the eukaryotic tree reconstructed on the area 2 alignment with a single archaeal outgroup organism. In this tree as well, a number of branches are shifted to a position different from the one they occupy in the eukaryotic subtree of the complete LSU tree. In particular, the species *Dictyostelium discoideum, Caenorhabditis elegans, Drosophila melanogaster,* and *Aedes albopictus* seem extremely susceptible to displacement when distant species are included in the dataset, as is the case in the complete LSU tree of Fig. 2. Due to this behavior, which is also observed with SSU rRNA, these species are often omitted in evolutionary studies although their rRNA sequences are available.

Animals, plants, and fungi appear as monophyletic clusters. The protoctists, on the contrary, are separated into several clusters. The alveolates diverge within the radiation of fungi, plants, and animals, while other protoctists form early branches in the eukaryotic tree. The branching order between all major clusters is poorly supported by bootstrap analysis. Only the very early branching of the genus *Giardia* is manifestly supported. The eukaryotic tree based on the area 1 alignment (not shown) has the same topology except for the position of two species–viz., *Euglena gracilis* and *Entamoeba histolytica.*

Comparison of LSU rRNA with SSU rRNA as a Molecular Clock

Since the LSU trees based on alignment area 2 showed better resolution and higher bootstrap values than those

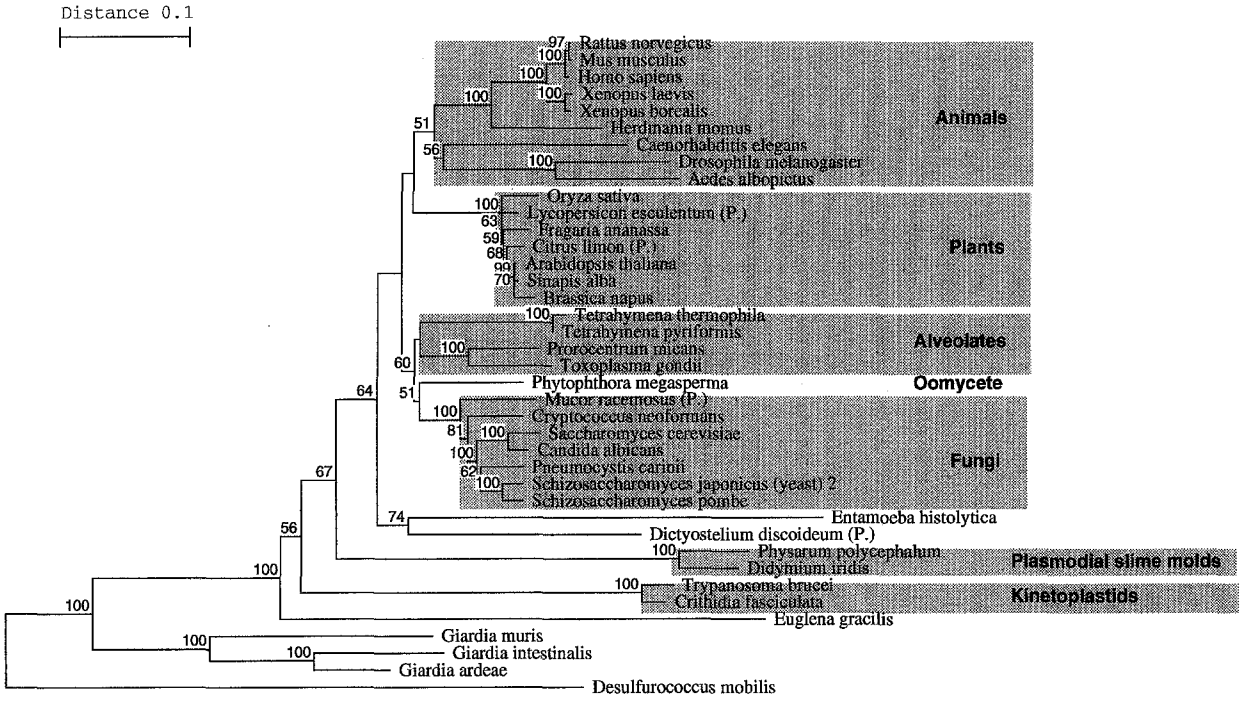

Fig. 4. The tree of the Eucarya was constructed using the Jin and Nei distance correction method and using the archaebacterium *Desulfurococcus mobilis* as an outgroup.

based on area 1, only area 2 was used for further work on LSU rRNA. Trees were constructed for the 172 species which are common to the LSU and SSU datasets. The overall topology of the LSU tree constructed on this set was the same as that of the tree in Fig. 2, based on the complete set of 225 sequences. The SSU tree constructed on the 172 sequences of the common set showed the same major clusters as the LSU trees, except for the Archaea, which appeared paraphyletic. The bootstrap values in the SSU tree were generally lower than those in the LSU tree. The relative order of divergence of the large taxa, such as the bacterial divisions and the eukaryotic kingdoms, also differs to some extent between the LSU and SSU trees of the common set. Again, such differences occur mainly in the poorly supported branches. Much to our surprise, the SSU tree also showed a monophyletic lineage of all mitochondria, clustering with the Proteobacteria, although with a very small internodal distance and a low bootstrap value. In earlier trees that we constructed on the basis of SSU rRNA (Van de Peer et al. 1990) the mitochondria usually appeared as polyphyletic, the plant mitochondria clustering with the Proteobacteria, while the divergence of animal, fungal, and protist mitochondria seemed to precede that of the bacterial divisions. This result was in conflict with those of Yang et al. (1985) and Cedergren et al. (1988). A number of experiments were carried out in order to find the reason for the change in perception of the mitochondrial descent. It could be caused by the use of a more sophisticated distance correction method in the present study—viz., the method of Jin and Nei (1990) (JN) rather than that of jukes and Cantor (1969) (JC). Alternatively, it could be due to the choice of the species set or due to improvements made to the SSU rRNA alignment.

LSU and SSU trees were constructed for the common species set, using the JC correction, which we used in earlier publications (e.g., Van de Peer et al. 1990), rather than the JN correction used in the trees of Figs. 2–4. The results are not shown but can be summarized as follows. The JC-corrected LSU tree is practically identical to its JN-corrected counterpart. In the JC-corrected SSU tree there are some differences in branching order relative to the JN-corrected tree, but the mitochondria still appear as monophyletic and still cluster with the Proteobacteria.

Several random samples of species of different sizes were then picked from the complete SSU and LSU databases and trees were made on basis of these samples. Sample size ranged from 60 to 200 species. Most of the SSU trees made using the JN correction showed a monophyletic group of mitochondria, clustered with the Proteobacteria α subdivision. However, in a few trees the mitochondria clustered with other bacterial taxa, and in one tree they behaved as polyphyletic. In several trees made using the JC correction the mitochondria appeared polyphyletic, as found previously (Van de Peer et al. 1990). As for the LSU trees, the mitochondria appeared monophyletic in all of them. Using the JN correction they clustered with the Proteobacteria α subdivision in all LSU trees generated. Using the JC correction, they behaved more often as a sister group of the entire Proteobacteria division. One of the eight LSU trees obtained with the JC correction showed the lineage of mitochondria branching between the Archaea and Bacteria.

In order to quantify these observations, the resam-

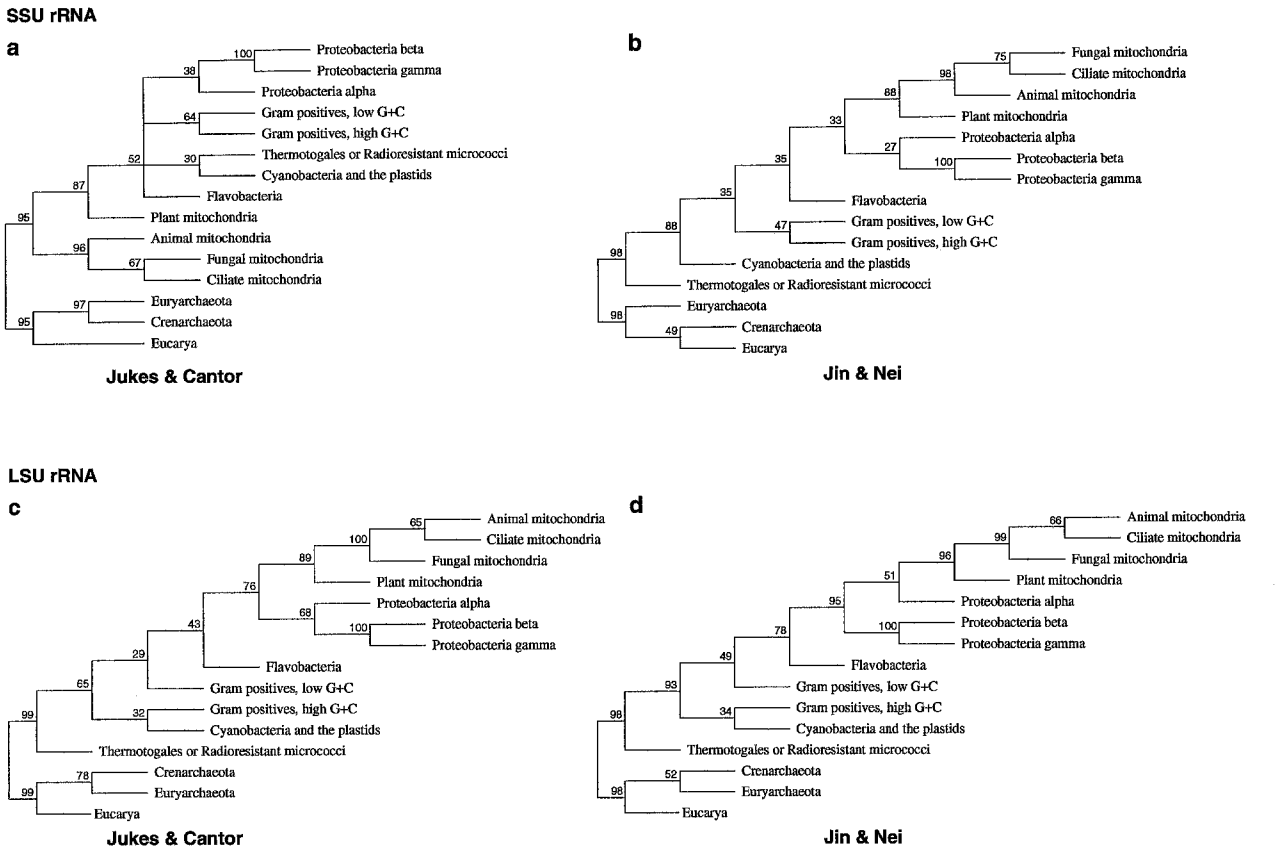

Fig. 5. Consensus trees comparing the results of analyses based on the LSU and SSU rRNA alignments, using the JC and JN distance-correction methods. For each analysis 100 trees were constructed, each containing one representative randomly chosen from each of 15 groups defined in the text. The clusters of size 2 to 14 observed in the 100 trees were ranked in order of frequency of occurrence. The 13 most frequently observed clusters were selected, omitting those incompatible with more frequent clusters, to construct the consensus tree. The frequency of each cluster is indicated at its root in the consensus tree. In tree **A** the divergence order of certain branches could not be decided due to the rejection of mutually incompatible clusters. These branches are drawn as if diverging simultaneously. Horizontal distances are chosen arbitrarily and are unrelated to evolutionary distance.

pling method described previously by Van de Peer et al. (1994a) was used on the set of species common to the LSU and SSU rRNA datasets. One hundred trees were made by randomly taking one species of each of 15 groups which are clearly distinguishable in all the trees constructed. These are: the Eucarya, the Euryarchaeota, the Crenarchaeota, the bacterial divisions or subdivisions represented by at least two genera, and mitochondria from plants, animals, fungi, and ciliates. In order to test the consistency of the early branching of *Thermotoga maritima* and *Thermus thermophilus* within the Bacteria, they were also treated as a group although they belong to different bacterial divisions—viz., the Thermotogales and the radioresistant Micrococci. Since a tree of 15 operational taxonomic units contains 13 clusters (of size 2 to 14), 1,300 clusters are found in the 100 trees. A consensus tree was then constructed by selecting clusters in the order of frequency of their occurrence, rejecting those whose existence is incompatible with previously selected ones (Van de Peer et al. 1994a). This exercise was carried out four times, for both LSU and SSU rRNA, and for both correction methods.

The resulting consensus trees are shown in Fig. 5. With SSU rRNA and using JC correction (Fig. 5a), most (but not all) trees constructed made the mitochondria appear paraphyletic, branching somewhere between the Eucarya, Bacteria, and Archaea. This result shows some similarity with the trees published previously (Van de Peer et al. 1990), based on the same molecule and distance correction method. In the latter trees though, plant mitochondria clustered with the Proteobacteria α subdivision, which is the case in only 30/100 of the random-selection trees. This cluster cannot be seen in Fig. 5a since it conflicts with a more frequently observed cluster (52/100 trees) comprising the entire bacterial domain excluding the plant mitochondria. Among the SSU trees using JN correction (Fig. 5b), 88 out of 100 show a monophyletic cluster of mitochondria, situated somewhere within the Bacteria. Which bacterial division it clusters with is not clear. In 33/100 trees it is with the Proteobacteria, but this value is too low to be significant. The divergence order of the other bacterial groups also varies sharply with the dataset. Only the Thermotogales and radioresistant micrococci seem to keep their early branching position with some confidence. The Archaea behave as paraphyletic, the Euryarchaeota being sepa-

rated from the Eucarya and Crenarchaeota. The two archaeal groups cluster together in only 46/100 trees (not visible in Fig. 5b).

The LSU trees show more consistency, reflected by higher average cluster frequencies. The mitochondria appear monophyletic in 89 cases using JC (Fig. 5c) and in 96 cases using JN correction (Fig. 5d). They cluster with the Proteobacteria in 76/100 trees using the JC correction. The association with the Proteobacteria is even stronger (95/100) in the JN-corrected trees, where they cluster with the α subdivision in 51/100 trees. Either the Thermotogales or the radioresistant micrococci form the first bacterial branch in both trees, although with lower confidence using JC correction. The branching order of the other bacterial groups is the same using both correction methods, but their cluster frequencies are too low to be decisive. The Archaea behave as monophyletic in both LSU trees, but only in 52 cases when the JN correction was applied.

Discussion

After surveying all the trees presented here, one would like to know what the real phylogeny is. The major groupings within bacteria and Eucarya stand rather firm, as they are found in nearly all trees. The divergence order of these groups and the branching order within these groups often remain problematic. Some of these problems are addressed in this study.

According to the LSU trees, the mitochondria are monophyletic, clustering with the Proteobacteria. When the JN correction is used, a relationship with the subgroup α is weakly supported (Figs. 2, 5d). But the SSU trees do not agree on this issue. However, using the more realistic JN-correction method, a monophyletic lineage of mitochondria clustering within the bacteria could also be demonstrated. The group of bacteria this lineage clusters with is less clear as it varies with the dataset too much (Fig. 5b). The findings based on LSU are in agreement with other evidence (Dickerson 1980; Roise and Maduke 1994) indicating the Proteobacteria subgroup α as the closest relatives of the mitochondria.

In agreement with an earlier study (Van de Peer et al. 1994a), the Gram-positive bacteria clearly behave as polyphyletic using LSU rRNA. Only in 6 out of 100 and 3 out of 100 resampled trees were the low G + C and high G + C species found as one cluster using JC and JN correction, respectively. The SSU resampling trees are again less clear on this subject. A binary cluster of low G + C and high G + C Gram positives is seen in 64/100 trees made using JC correction (Fig. 5a), and in 47/100 trees using JN correction (Fig. 5b). However, the SSU trees on resampled datasets presented previously (Van de Peer et al. 1994a), based on a much larger number of species, and containing taxa not present in the LSU rRNA dataset, found this binary cluster in a much smaller number of trees. Combining these results with those obtained here with LSU rRNA, we think that the Gram positives may not form one monophyletic taxon.

The early divergence of Thermotogales and radioresistant micrococci seems to be supported strongly by both LSU and SSU rRNA. However, their clustering in the LSU tree in Fig. 2 is probably an artefact caused by the presence of only two early-diverging species with high G + C content in their rRNA.

The monophyletic origin of the Archaea is less clear. Both the monophyletic and the paraphyletic origins of these species have been advocated (Woese 1987; Lake 1988). While most trees made here do show a binary cluster of the two groups of Archaea selected, the supporting values are relatively low, especially when JN correction is used. One consensus tree (Fig. 5b) actually shows them as paraphyletic.

When trying to create evolutionary trees based on molecular evidence, one has to make several choices—concerning the molecule, the species set, the alignment positions used, the tree construction method—all of which influence the final result. Although some of the problems accompanying these choices were addressed here, only distance-matrix tree-construction methods were used. This is because the construction of the trees containing a large number of species is not feasible using parsimony. Furthermore, the rRNAs show different mutational rates in different lineages, in which case simulation has shown (Jin and Nei 1990; Saitou and Imanishi 1989) that the neighbor-joining method (Saitou and Nei 1987) yields better results than parsimony, especially when used with the JN-correction method. JC and JN correction were used in this study—JC correction mainly to compare with previous results. As can be seen in Fig. 5, the use of JN correction leads to more consistent results in the resampled trees and to more similar results when comparing SSU and LSU rRNA.

It is clear from the experiments presented here that the choice of a set of species to be used to construct a phylogenetic tree can have a big influence on the observed topology of taxon divergence. There is no real objective criterion for preferring one set to another. By using a particular set of species, it is possible to obtain a distorted picture of evolution. The resampling method used here tries to address this problem. Many sets are sampled, and the consensus tree gives an indication of which clusters are most likely to be found in phylogenetic trees based on a given molecule and method. While not having the statistical properties of bootstrap analysis, it can give an idea of which topologies are more likely. In this study, only the set of species for which both SSU and LSU rRNA sequences are available was used in order to make a fairer comparison between the two molecules.

LSU trees seem to be less prone to variation due to changes in method and dataset than SSU trees. The LSU consensus trees based on JC and JN correction give nearly identical results, although with a higher confi-

dence in the JN trees. The topology of the mitochondria in the LSU tree is also more in accordance with other evidence. Therefore LSU rRNA seems to be a more reliable molecular clock than SSU rRNA. However, the number of presently known complete or nearly complete LSU rRNA sequences is an order of magnitude lower than those of SSU rRNA. The set of known LSU rRNA sequences is also more biased, containing mainly animal mitochondria and Gram-positive, low-G + C bacteria, and some large taxa are not represented, or only by a small number of species. Furthermore, the alignment of LSU rRNA is more difficult in some areas, often also as a consequence of the lack of species from certain groups. However, the results presented here indicate that LSU rRNA is worth the trouble of sequencing, as it proves to be extremely useful for studying global phylogenetic relationships.

Acknowledgments. Peter De Rijk is research assistant of the National Fund for Scientific Research. Our research was supported by the Programme on Interuniversity Poles of Attraction (contract 23) and the Programme "Health Hazards" of the Federal Office for Scientific, Cultural and Technical Affairs of the Belgian State, and by the Fund for Collective Fundamental Research.

References

Baroin-Tourancheau A, Delgado P, Perasso R, Adoutte A (1992) A broad molecular phylogeny of ciliates: identification of major evolutionary trends and radiations within the phylum. Proc Natl Acad Sci USA 89:9764–9768

Cedergren G, Gray MW, Abel Y, Sankoff D (1988) The evolutionary relationships among known life forms. J Mol Evol 28:98–112

Clark CG, Tague BW, Ware VC, Gerbi SA (1984) *Xenopus laevis* 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications. Nucleic Acids Res 12:6197–6220

De Rijk P, De Wachter R (1993) DCSE, an interactive tool for sequence alignment and secondary structure research. Comput Appl Biosci 9:735–740

De Rijk P, Van de Peer Y, Chapelle S, De Wachter R (1994) Database on the structure of large ribosomal subunit RNA. Nucleic Acids Res 22:3495–3501

Dickerson R (1980) Evolution and gene transfer in purple photosynthetic bacteria. Nature 283:210–212

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

Field KG, Olsen GJ, Lane DJ, Giovannoni SJ, Ghiselin MT, Raff EC, Pace NR, Raff RA (1988) Molecular phylogeny of the animal kingdom. Science 239:748–753

Golding GB (1983) Estimates of DNA and protein sequence divergence: an examination of some assumptions. Mol Biol Evol 1:125–142

Hassouna N, Michot B, Bachellerie J-P (1984) The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucleic Acids Res 12:3563–3583

Hendriks L, Van Broeckhoven C, Vandenberghe A, Van de Peer Y, De Wachter R (1988) Primary and secondary structure of the 18S ribosomal RNA of the bird spider *Eurypelma californica* and evolutionary relationships among eukaryotic phyla. Eur J Biochem 177:15–20

Iwabe N, Kuma K-I, 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

Jin L, Nei M (1990) Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol Biol Evol 7:82–102

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Muntu HN (ed) Mammalian protein metabolism, vol III. Academic press, New York

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

Lake JA (1988) Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. Nature 331:184–186

Nei M (1991) Relative efficiencies of different tree-making methods for molecular data. In: Miyamoto MM, Cracraft JL (eds) Recent advances in phylogenetic studies of DNA sequences. Oxford University press, Oxford

Olsen GJ (1987) Earliest phylogenetic branchings: comparing rRNA-based evolutionary trees inferred with various techniques. Cold Spring Harbor Symp Quant Biol 52:825–837

Olsen GJ, Woese CR, Overbeek R (1994) The winds of (evolutionary) change: breathing new life into microbiology. J Bacteriol 176:1–6

Roise D, Maduke M (1994) Import of a mitochondrial presequence into *P. denitrificans*. Insight into the evolution of protein transport. FEBS Lett 337:9–13

Rzhetsky A, Nei M (1994) Unbiased estimates of the number of nucleotide substitutions when substitution rate varies among different sites. J Mol Evol 38:295–299

Saitou N, Imanishi T (1989) Relative efficiencies of the Fitch-Margoliash, Maximum-Parsimony, Maximum-Likelihood, Minimum-Evolution, and Neighbour-joining methods of phylogenetic tree construction in obtaining the correct tree. Mol Biol Evol 6: 514–525

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

Sogin ML (1989) Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs. Am Zool 29:487–499

Van de Peer Y, De Wachter R (1993) TREECON: a software package for the construction and drawing of evolutionary trees. Comput Applic Biosci 9:177–182

Van de Peer Y, De Wachter R (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput Appl Biosci 10:569–570

Van de Peer Y, Neefs J-M, De Rijk P, De Vos P, De Wachter R (1994a) About the order of divergence of the major bacterial taxa during evolution. System Appl Microbiol 17:32–38

Van de Peer Y, Van den Broeck I, De Rijk P, De Wachter R (1994b) Database on the structure of small ribosomal subunit RNA. Nucleic Acids Res 22:3488–3494

Van de Peer Y, Neefs J-M, De Rijk P, De Wachter R (1993) Reconstructing evolution from eukaryotic small ribosomal subunit RNA sequences: calibration of the molecular clock. J Mol Evol 37:221–232

Van de Peer Y, Neefs J-M, De Wachter R (1990) Small ribosomal subunit RNA sequences, evolutionary relationships among different life forms, and mitochondrial origins. J Mol Evol 30:463–476

Wilmotte A, Van de Peer Y, Goris A, Chapelle S, De Baere R, Nelissen B, Neefs J-M, Hennebert GL, De Wachter R (1993) Evolutionary relationships among higher fungi inferred from small ribosomal subunit RNA sequence analysis. System Appl Microbiol 16:436–444

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271

Yang D, Oyaizu Y, Oyaizu H, Olsen GJ, Woese CR (1985) Mitochondrial origins. Proc Natl Acad Sci USA 82:4443–4447