J Mol Evol (1999) 49:509–523

JOURNAL OF MOLECULAR EVOLUTION

The Rooting of the Universal Tree of Life Is Not Reliable

Hervé Philippe,[1] Patrick Forterre[2]

[1] Phylogénie et Evolution Moléculaires (UPRESA 8080 CNRS), Bâtiment 444, Université Paris-Sud, 91405 Orsay-Cedex, France
[2] Institut de Génétique et Microbiologie (UMR 8621 CNRS), Bâtiment 409, Université Paris-Sud, 91405 Orsay-Cedex, France

Abstract. Several composite universal trees connected by an ancestral gene duplication have been used to root the universal tree of life. In all cases, this root turned out to be in the eubacterial branch. However, the validity of results obtained from comparative sequence analysis has recently been questioned, in particular, in the case of ancient phylogenies. For example, it has been shown that several eukaryotic groups are misplaced in ribosomal RNA or elongation factor trees because of unequal rates of evolution and mutational saturation. Furthermore, the addition of new sequences to data sets has often turned apparently reasonable phylogenies into confused ones. We have thus revisited all composite protein trees that have been used to root the universal tree of life up to now (elongation factors, ATPases, tRNA synthetases, carbamoyl phosphate synthetases, signal recognition particle proteins) with updated data sets. In general, the two prokaryotic domains were not monophyletic with several aberrant groupings at different levels of the tree. Furthermore, the respective phylogenies contradicted each others, so that various ad hoc scenarios (paralogy or lateral gene transfer) must be proposed in order to obtain the traditional Archaebacteria–Eukaryota sisterhood. More importantly, all of the markers are heavily saturated with respect to amino acid substitutions. As phylogenies inferred from saturated data sets are extremely sensitive to differences in evolutionary rates, present phylogenies used to root the universal tree of life could be biased by the phenomenon of long branch attraction. Since the eubacterial branch was always the longest one, the eubacterial rooting could be explained by an attraction between this branch and the long branch of the outgroup. Finally, we suggested that an eukaryotic rooting could be a more fruitful working hypothesis, as it provides, for example, a simple explanation to the high genetic similarity of Archaebacteria and Eubacteria inferred from complete genome analysis.

Key words: Root of the tree of life — ATPase — Carbamoyl phosphate synthetase — Elongation factor — tRNA synthetase — Signal recognition particle — Mutational saturation — Long branch attraction

Introduction

According to ribosomal RNA (rRNA) sequence comparisons, all extant cellular organisms have been classified into one of three domains Eubacteria, Archaebacteria, and Eukaryota (Woese 1987). The classification of the living world in three groups is supported by numerous molecular phenotypic traits specific for each domain (for recent reviews, see Brown and Doolittle 1997; Forterre 1997; Olsen and Woese 1997). Comparison of molecular biology and central metabolism between Eubacteria, Archaebacteria and Eukaryota is expected to help in reconstituting the characteristics of the last common ancestor to all extant cellular life, or cenancestor, here called the Last Universal Cellular Ancestor (LUCA). This would require polarizing characters found in one or two domains to determine if they are primitive or derived features (plesiomorphies or synapomorphies sensu Hennig 1966). The rooting of the universal tree of life would

Correspondence to: H. Philippe, *e-mail:* herve.philippe@bc4.u-psud.fr

facilitate this task since homologous traits only shared by two domains, which are not sister-groups, should be ancestral. For example, considering that Archaebacteria and Eubacteria have a similar type of genome organization (chromosome size and number, operons, mode of cell division), rooting the universal tree in the eubacterial or the archaebacterial branch would suggest that these traits were already present in LUCA. In contrast, if the universal tree is rooted in the eukaryotic branch, these characters could either have been present in LUCA or have appeared in the branch common to the prokaryotes, i.e. correspond to an evolved state.

At the end of the eighties, two research teams tentatively rooted the universal tree of life in the eubacterial branch (Gogarten et al. 1989; Iwabe et al. 1989). This was inferred from the construction of universal trees for two pairs of paralogous proteins, which originated by gene duplication before LUCA. The proteins used were the elongation factors (EF), namely EF-1α(Tu) versus EF-2(G), and the catalytic versus regulatory subunits of eubacterial F-ATPases, and V or V-like-ATPases found in Eukaryota and Archaebacteria. The root always turned out to be located in the eubacterial branch. Later on, Brown and Doolittle (1995) used the same strategy to root a universal tree of lle-tRNA synthetases (lle-tRS) versus paralogous Val- and Leu-tRS, Lawson et al. (1996) a carbamoyl phosphate synthetase (CPS) tree using an internal gene duplication, Brown et al. (1997) a Tyr-tRS tree versus paralogous Trp-tRS, and Gribaldo and Cammarano (1998) a Signal Recognition Particle (SRP) 54kD protein using paralogous SRP receptor SR-α. In all cases, the root again separated Eubacteria from the two other domains. Moreover, a reanalysis of the elongation factor data set with more sequences and a refined alignment strengthened the eubacterial rooting (Baldauf et al. 1996).

The eubacterial rooting supports the current view that LUCA was a prokaryotic-like organism since characters shared by Archaebacteria and Eubacteria are considered primitive. Furthermore, it fits intuitively well with the finding that several features of the cellular information processing system are more similar between Eukaryota and Archaebacteria than between Archaebacteria and Eubacteria (Olsen and Woese 1997), and the common assumption that these features are more “evolved” than their eubacterial counterparts. Accordingly, this rooting was rapidly accepted and advertised in the community of evolutionary biologists and beyond, being now systematically used to draw universal trees in review papers, and even textbooks. The eubacterial rooting was also endorsed to support several evolutionary hypotheses, such as the origin of life at high temperature (Stetter 1992). Last but not least, Woese and coworkers recruited this rooting to support their new nomenclature for the three domains of life (removing the suffix bacteria from Archaebacteria) (Woese et al. 1990), since Archaebacteria are the sister group of Eukaryota, not of Eubacteria, when the universal tree is rooted in the eubacterial branch.

However, the validity of sequence comparison to infer ancient phylogenies has been questioned on various grounds. With more and more sequences available, it turned out that most protein phylogenies contradict each others as well as the rRNA tree (reviewed in Brown and Doolittle 1997; Doolittle and Brown 1994; Forterre 1997). In several cases, archaebacterial proteins were found more closely related to eubacterial ones than to eukaryotic ones, whilst in some cases eukaryotic proteins appeared close to eubacterial ones. This situation led to two major reactions. Some people suggested new scenarios of early cellular evolution based on their favorite proteins, or else diverse scenarios of fusion between primitive lineages to take into account contradictions between different phylogenies (Gupta and Golding 1993; Martin and Muller 1998; Moreira and Lopez-Garcia 1998; Rivera and Lake 1992; Sogin 1991; Zillig 1987). Other evolutionists argued that the proteins from the information processing system were intrinsically better than others because they are less prone to inter-domain transfer than metabolic proteins. Accordingly, since many proteins of the archaebacterial transcription, translation or replication apparatus resemble their eukaryotic homologues more than their eubacterial ones, they suggested that their phylogenies (even unrooted) testify for the eubacterial rooting of the tree of life (see for example Brown and Doolittle 1997).

However, it is possible that contradictions observed between universal phylogenies obtained with rRNA and various proteins do not require specific ad hoc hypotheses but simply reflect the weakness of the tree reconstruction methods that have been used to infer these phylogenies (Forterre 1997; Philippe and Laurent 1998). In particular, when the elongation factor and tRNA synthetase data sets were analyzed for the slowly evolving positions which should have been a priori the most informative, we did not found a significant signal for any rooting (Forterre 1997; Forterre et al. 1992). Up to now, these criticisms have not been sufficiently taken into account. Nevertheless this situation could change now, following recent developments in the study of early eukaryotic evolution which showed that some molecular phylogenies might be highly misleading. Indeed, in the case of eukaryotes, strikingly different trees can be obtained depending on the molecule analyzed (either rRNA, actin or tubulin). The order of emergence of the various groups at the base of the eukaryotic tree mainly depends on the rate of evolution of the protein used (the more rapidly evolving taxon emerging first) because the long branches of these groups are attracted by the long branch of the outgroup that roots the tree (Philippe and Adoutte 1998).

These considerations prompted us to revisit in detail

all the phylogenies that have been used up to now to root the universal tree of life, using updated data sets that especially include many novel archaebacterial and eubacterial sequences obtained from complete genome sequencing efforts. Here we report our studies on CPS, SRP, elongation factors, lle-tRS, Trp/Tyr-tRS and ATPase genes. We demonstrate that the phylogenies are highly confusing due to the combining effects of gene duplication, gene loss, lateral gene transfer and tree reconstruction artefact. Moreover the six genes appear to be highly mutationally saturated, suggesting that very few ancient phylogenetic signal remains. Finally we suggest that the eubacterial rooting is the result of a long branch attraction artefact and we discuss the hypothesis of a eukaryotic rooting.

Materials and Methods

All sequences homologous to the carbamoyl phosphate synthetase (CPS), ATPase, lle- and Val-tRNA synthetase (tRS), Trp- and Tyr-tRS, signal recognition protein (SRP) proteins, and elongation factors EF1α and EF2 available in data banks were identified by a BLAST search using the sequence from *Escherichia coli* as query sequence. The programs blast2retp and retp2ali (Philippe Lopez, personal communication) allowed us to retrieve all the sequences automatically and to write them into a MUST-compatible file. The alignment of these sequences was carried out visually with the help of the ED program of the MUST package version 1.0 (Philippe 1993). Some sequences were discarded because they were either partial or redundant or because they contained likely sequencing errors. The *Pyrobaculum aerophilum* sequences were kindly provided by Drs. Sorel Fitz-Gibbon and Jeffrey Miller. Preliminary sequence data were obtained from the Institute for Genomic Research website at http://www.tigr.org. The resulting alignments contained 339, 122, 344, 104, 118, and 103 sequences for ATPase, CPS, EF, lle/Val-tRS, SRP, and Trp/Tyr/tRS, respectively. Due to computer time limitation, only 124 and 75 sequences were selected for ATPase and EF, while keeping the greatest possible phylogenetic diversity. Positions that could not be unambiguously aligned were excluded from the analysis, yielding 201, 271, 158, 322, 184, and 83 usable positions for ATPase, CPS, EF, lle/Val-tRS, SRP, and Trp/Tyr-tRS, respectively. All alignments are available from HP upon request.

Phylogenetic trees were constructed with maximum-likelihood (ML), maximum-parsimony (MP), and distance-based methods, with the programs PROTML (Adachi and Hasegawa 1996) version 2.3, PAUP (Swofford 1993) version 3.1, and NJ in the MUST package (Philippe 1993) version 1.0, respectively. The distances were computed with the substitution model of Kimura (1983). MP trees were obtained by 10 random addition heuristic search replicates. Due to the high number of species used, the search for the ML tree was limited to a reduced sample of species and to local rearrangement method (option R) starting from the MP and the neighbor-joining (NJ) trees. The model of amino acid substitution used was JTT. Bootstrap proportions (BP) were calculated by analysis of 1000 replicates for NJ analysis (Saitou and Nel 1987). The results obtained by MP and ML methods are not shown because they are very similar to those of the NJ method.

The saturation level of the phylogenetic markers was estimated with the use of the method of Philippe et al. (1994). The inferred number of substitutions between each couple of species was estimated from the MP or the ML trees as the sum of the lengths of all the branch on the pathway linking these two species, using the program TREEPLOT (Philippe 1993). Using the program COMP_MAT, a plot was drawn to estimate the saturation level by displaying all the pairs of species with an abscissa value equal to the number of inferred substitutions and an ordinate equal to the number of observed differences. The mutational saturation was revealed by the presence of a plateau within which the number of substitutions increased, whereas the number of differences remained constant.

Results and Discussion

Confusing Phylogenies

Updated phylogenies are shown in Figs. 1–5 for the two CPS domains, Ile- and Val-tRS, Tyr- and Trp-tRS, the V- and F-ATPases, and the SRP and its receptor. The number of sequences has considerably increased during the last 2 years, thanks to the numerous genome projects that have been completed or are in progress (http://www.tigr.org/tdb/mdb/mdb.html). Furthermore, the availability of complete eubacterial and archaebacterial genomes allows us to identify putative gene loss, gene transfer, and gene duplication more safely. None of the six updated trees offers the classical Woese's picture, e.g., the monophyly of the three domains and the eubacterial rooting of each subtree using the other as an outgroup.

The more puzzling phylogeny was observed for the Tyr/Trp-tRS tree (Fig. 1), since the monophyly of both types of synthetase was not recovered, bacterial Trp- and Tyr-tRS being grouped together. This peculiar phylogeny was initially obtained by Ribas de Pouplana et al. (1996), using a limited data set that did not include archaebacterial sequences. These authors speculated that the divergence between Trp- and Tyr-tRS might have occurred only after the separation of prokaryotes and eukaryotes. Later on, the monophyly of both types of synthetase was nevertheless recovered by Brown et al. (1997) using an expanded data set and a different alignment. They concluded that the phylogeny previously obtained by Ribas de Pouplana and co-workers was due to long branch attraction (LBA) and that inclusion of archaeal sequences had allowed to infer the correct phylogeny by breaking the longest branches. However, in our present analysis, we get anew the topology first obtained by Ribas de Pouplana and coworkers (1996). It should be noted that the alignment of these tRNA synthetases is very difficult to perform. This is confirmed by the fact that a blast search using the Tyr-tRS of *E. coli* as a query sequence detects the eubacterial Tyr-tRS only, and neither the other Tyr-tRS nor the Trp-tRS. We were able to align only 83 positions unambiguously, which is significantly fewer than Brown et al. (1997) (147 or 184) and Ribas de Pouplana et al. (1996) (between 190 and 230). This difference in the alignment together with the use of various species sampling can explain the instability of the inferred phylogenies. Since the two eubacterial branches are much longer than all others in this phylogeny, we think that LBA is indeed the most likely hypothesis to explain the nonmonophyly of each type of synthetase. Examination of the local part of the Tyr and

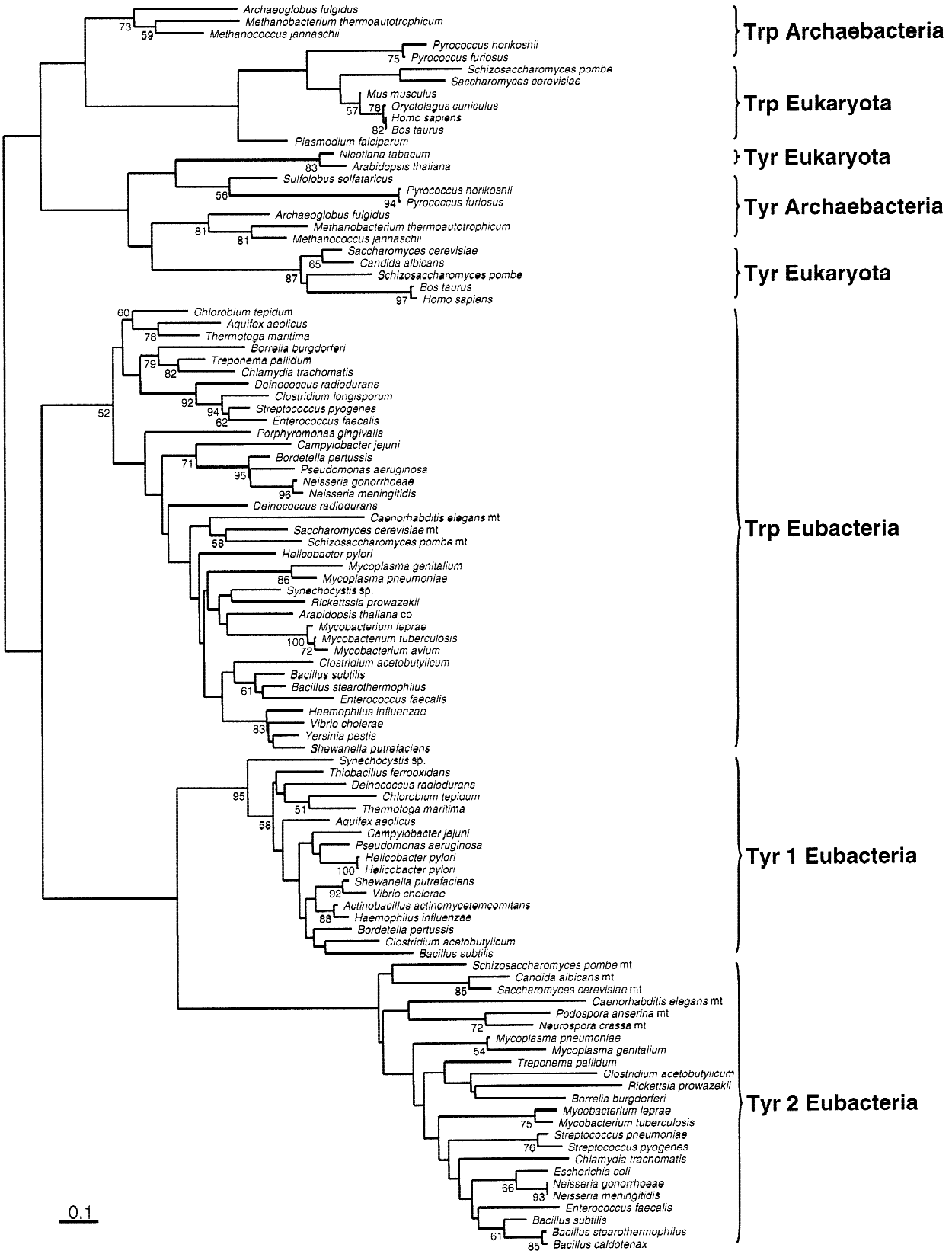

Fig. 1. Phylogenetic tree based on comparison of Trp- and Tyr-tRNA synthetase sequences; 83 unambiguously aligned positions were used. The tree was constructed with the NJ method employing the Kimura method of distance calculation. Bootstrap proportions are indicated when greater than 50%. A scale bar corresponding to 10 substitutions per 100 positions is given at bottom.

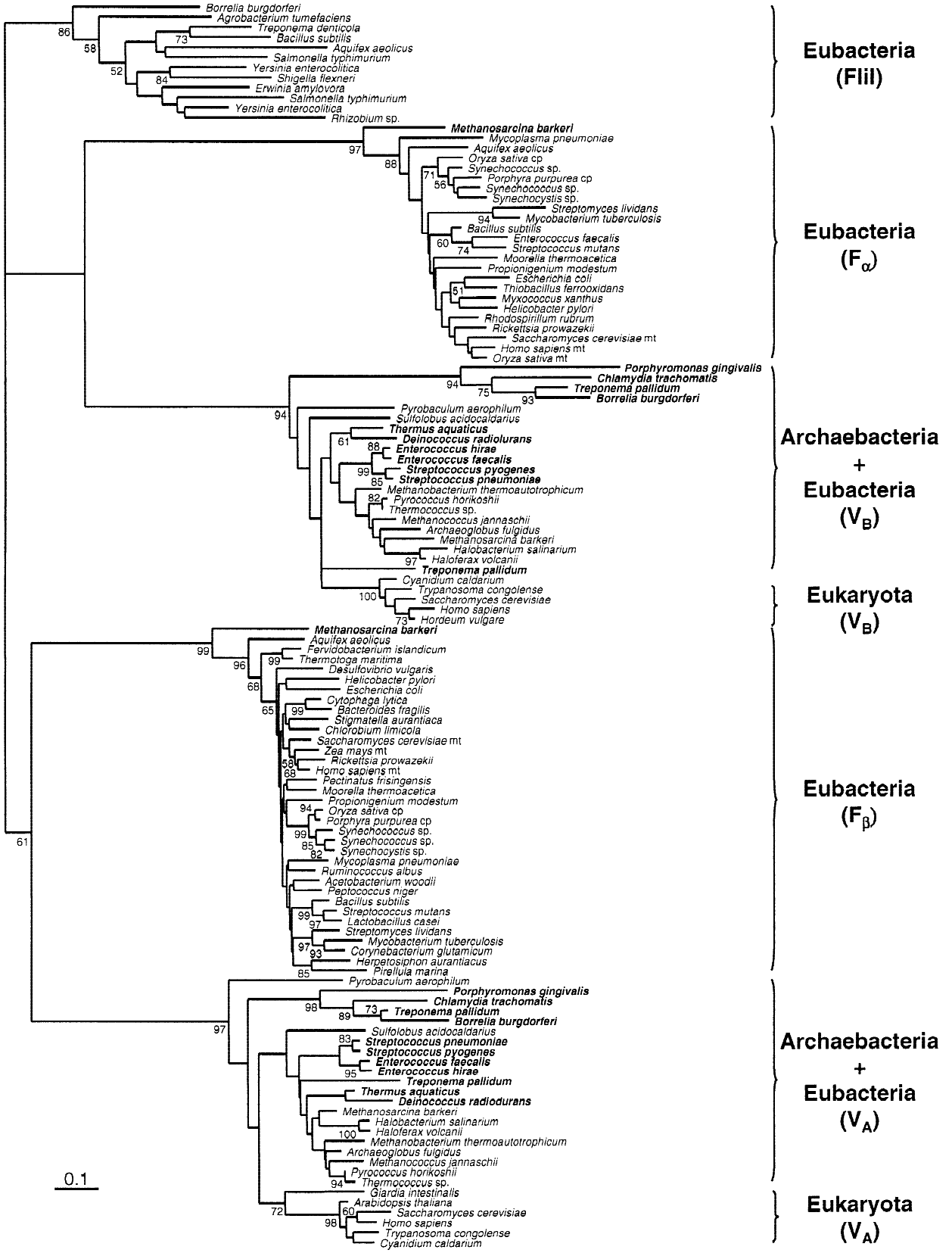

Fig. 2. Phylogenetic tree based on comparison of ATPase sequences; 201 unambiguously aligned positions were used. For method, see the legend to Fig. 1.

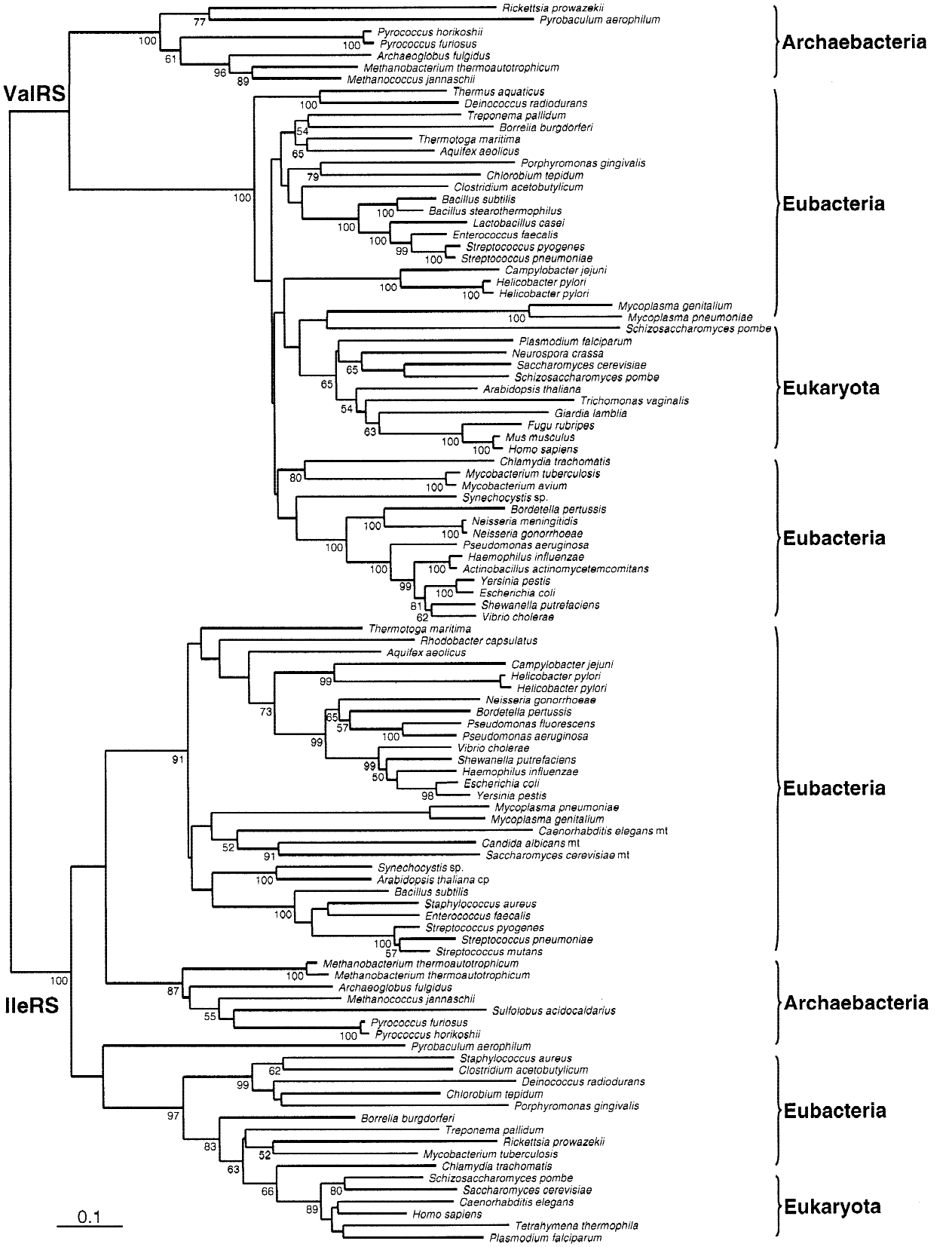

Fig. 3. Phylogenetic tree based on comparison of Ile- and Val-tRNA synthetase sequences; 322 unambiguously aligned positions were used. For the method, see the legend to Fig. 1.

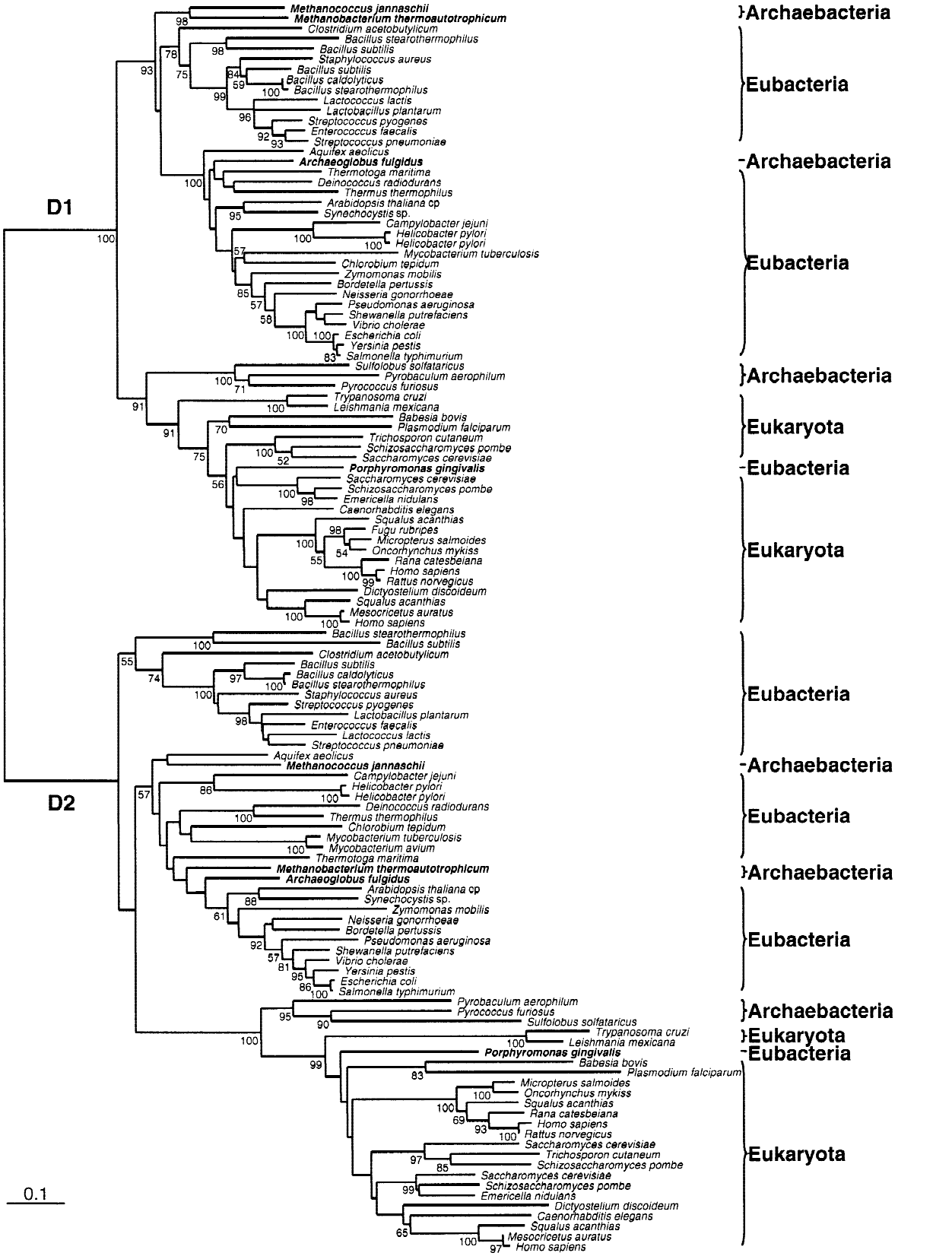

Fig. 4. Phylogenetic tree based on comparison of CPS sequences; 271 unambiguously aligned positions were used. For the method, see the legend to Fig. 1.

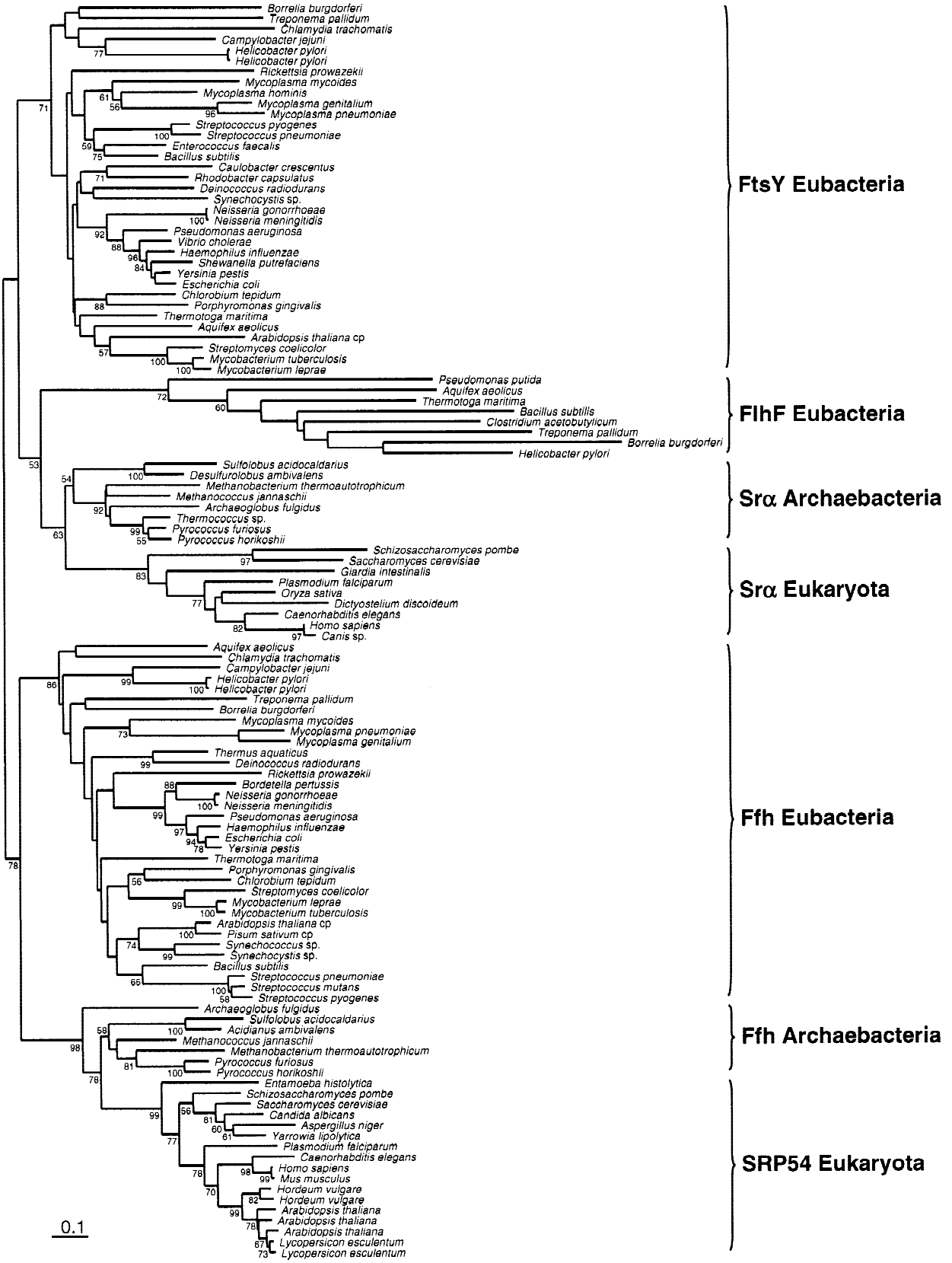

Fig. 5. Phylogenetic tree based on comparison of SRP sequences; 184 unambiguously aligned positions were used. For the method, see the legend to Fig. 1.

Trp-tRS tree revealed other oddities. In particular, *Pyrococcus* (Archaebacteria) sequences of Trp-tRS branch inside the eukaryotes, whereas plant Tyr-tRS branch inside Archaebacteria. Our updated phylogeny also identifies two groups of eubacterial Tyr-tRS (1 and 2). They have probably originated from a duplication in the eubacterial domain since they are both present in most bacterial kingdom, two species, *B. subtilis* and *Clostridium acetobutilicum*, containing the two genes. If this hypothesis is correct, one of the two genes should have been repeatedly lost during eubacterial evolution. All mitochondrial sequences are related to group 2, but they emerged at the basis of this group instead of branching with α-proteobacteria. Our conclusion is that Tyr- and Trp-tRS are very bad phylogenetic markers (probably in part because of the low number of residues which can be aligned) and cannot be used to root the tree of life confidently.

The ATPase tree (Fig. 2) resembles the Tyr and Trp-tRS tree in the difficulty to recover the presumed monophyly of the two proteins that were originally supposed to be paralogous, in that case catalytic versus regulatory subunits of V- and F-type ATPases. The ATPase tree exhibits five major groups with extremely long basal branches, such that the monophyly of each group is well supported but the relationships between these groups are very difficult to ascertain. To evaluate the robustness of the ATPase tree, we computed the likelihood of three quite different topologies, ((F_α, F_β), FliI),(V_A,V_B)), ((F_α, FliI), F_β),(V_A, V_B)), and ((F_β, FliI), F_α),(V_A, V_B)), applying Kishino/Hasegawa's test on a limited set of 90 species. The difference of log-likelihood (ΔL) with respect to the ML tree (similar to that in Fig. 2; log$L$ = −21252.9) turned out to be only −20.5 (2.3 SE). By comparison, ΔL for the tree constraining the monophyly of crenotes with the V_B gene was −23.0 (1.6 SE). Since this constraint was quite reasonable, a difference of 20 in the log likelihood cannot be rejected. As a result, the orthology between F_α and V_B, on one hand, and between F_β and V_A, on the other, is far from being strongly supported.

One of the five groups includes only eubacterial sequences corresponding to flagellar ATPases, suggesting a duplication and functional specialization in Eubacteria. The two groups of F-ATPases contain only bacterial sequences with a single exception, the archaeon *Methanosarcina barkeri*, suggesting a transfer from Eubacteria to Archaebacteria. However, in that case, one should imagine an LBA artifact to explain why *M. barkeri* emerged at the base of the two eubacterial subtrees. In contrast to F-ATPases, the two groups of V-ATPases contain sequences from the three domains. This type of ATPase was originally discovered in Archaebacteria and considered the bona fide archaebacterial enzyme (Gogarten et al. 1989). Later on, the discovery of this type of ATPase in two Eubacteria was interpreted as a lateral gene transfer from Archaebacteria to Eubacteria (Gogarten et al. 1996). Now V-ATPase appears to be present in several major branches of the eubacterial tree (Fig. 2). However, their phylogeny is very confused: Archaebacteria and Eubacteria turned out to be paraphyletic in both subtrees with eukaryotic sequences and most archaebacterial ones branching inside eubacterial sequences. Accordingly, besides the previous hypothesis of several gene transfers from Archaebacteria to Eubacteria, one can now argue as well for several transfers from Eubacteria to Archaebacteria.

Considering the general shape of the tree with both F- and V-ATPases, the long branches of the two eubacterial F-ATPases subtrees suggest that these proteins might have appeared by gene duplication and functional specialization in Eubacteria, as in the case of flagellar ATPases. It is also possible that F- and V-ATPases were already present in LUCA and that V-ATPase are orthologues in the three domains (Forterre et al. 1992). In any case, the ATPase data set appears unsuitable to root the universal tree of life since, as for the Tyr- and Trp-tRS, the evolutionary relationships between the various classes of enzymes are obscure.

Similar problems are now obvious with the Ile- and Val-tRS tree (Fig. 3). It has been shown previously that the Val-tRS phylogeny cannot be used to root the universal tree since the eukaryotic enzymes turned out to be of mitochondrial origin (Brown and Doolittle 1995; Hashimoto et al. 1998). This is confirmed by our analysis. However, the Ile-tRS tree was supposed to be safe for this rooting. This is no more true with our new data set. The updated tree revealed the existence of a very diverse group of eubacterial Ile-tRS, including sequences from many eubacterial lineages, which branch between the archaebacterial and eukaryotic enzymes (Brown et al. 1998). To save the Woesian structure of the Ile-RS tree, the existence of this group of sequence should be explained by the ancient transfer of an eukaryotic gene to Eubacteria and its rapid evolution in this new context. However, there is no objective reason to consider that one of the two eubacterial groups corresponds to the bona fide eubacterial gene. These two genes might be ancient paralogues that have been lost selectively during eubacterial evolution. Furthermore, the eukaryotic gene might have been recruited from one of these eubacterial species by the ratchet mechanism proposed by Doolittle (1998) or from mitochondria (*Rickettsia* possessing this "abnormal" gene). Many alternative scenarios can be proposed with no obvious possibility to make a rational choice. An interesting feature in Fig. 3 is that Archaebacteria are polyphyletic, the majority of them clustering with Eubacteria and only *Pyrobaculum* with eukaryotes. However, with MP and ML method, the Archaebacteria are paraphyletic and the sister group of eukaryotes. This switch between the eukaryotic and the eubacterial rooting depending on the tree reconstruction method could

be due to the limited resolving power of this gene. In any case, the only safe conclusion is that the Ile-tRS phylogeny cannot be used anymore to root the tree of life with confidence.

The universal tree inferred from the two CPS domains (D1 and D2) appears a priori less confused since the monophyly of each CPS domains is clearly recovered (Fig. 4). However, the root is no more in the eubacterial branches of the two subtrees, as it was the case in previous analysis that included only one archaebacterial sequence (Lawson et al. 1996). Now Archaebacteria turned out to be polyphyletic: several euryotes (*Archaeoglobus*, *Methanococcus*, and *Methanobacterium*) clustered with Eubacteria, and two crenotes (*Sulfolobus* and *Pyrobaculum*) and one euryote (*Pyrococcus furiosus*) clustered with eukaryotes. Moreover, one eubacterial sequence (*Porphyromonas*) emerged within eukaryotes and the complete genome sequencing of *Pyrococcus horikoshi* and *Pyrococcus abyssi* has revealed that this gene is absent in both species. Finally, gene duplication of CPS occurred in Eubacteria (two genes in *Bacillus*) and in eukaryotes (two genes in *Saccharomyces*). As a result, inferring the species phylogeny from the CPS gene phylogeny is a very difficult task because of the numerous gene duplications, gene loss and horizontal gene transfer. One can argue for an eubacterial rooting by assuming that the euryote sequences have been acquired by horizontal gene transfers or for a eukaryotic rooting by assuming that the crenote sequences have been acquired from eukaryotes.

In conclusion, all four of these genes (tRS, ATPase, and CPS) cannot be used confidently to root the tree of life because of the difficulty to choose between different evolutionary scenarios, knowing that gene duplication, gene loss, and lateral gene transfer have been frequent during prokaryotic evolution.

Classical Phylogenies

Of the six trees examined, the elongation factor (shown by Lopez et al. 1999) and the signal recognition particle/receptor trees are those that are more like the classical Woese's tree. However, they are again plagued with various problems. For SRP (Fig. 5), the three domains are monophyletic, with the root in the eubacterial branch, except for ffh, where Archaebacteria are paraphyletic with *Archaeoglobus* branching first. The situation is more complex in the SRα/Ftsy/FIhF subtree since it contains two eubacterial groups. The Ftsy group, which corresponds to functional analogues of SRα branches first in this subtree, whereas the FIhF, which corresponds to proteins involved in the biogenesis of the flagellum, is clustered with Archaebacteria and eukaryotes. The two bacterial groups likely originated from gene duplication in the bacterial domain and the long branch of the FIhF probably reflected rapid evolution due to functional change. The relations of orthology cannot be safely inferred, so that this part of the tree cannot be used to root the universal tree (Brinkmann and Philippe 1999). In the elongation factor tree (Lopez et al. 1999), eukaryotes and Eubacteria are monophyletic in the two subtrees, but Archaebacteria are paraphyletic in both cases. The EF-2 (EF-G) tree exhibits Lake's topology with crenotes grouped with eukaryotes. However, the euryotes are paraphyletic, the euryote *Halobacterium salinarium* branching between crenotes and other euryotes. Furthermore, the EF-1α tree exhibits a different topology, Archaebacteria being monophyletic, except the early emergence of the *Pyrococcus/Thermococcus* group.

In all six trees, the examination of intradomain phylogenetic patterns shows a mixture of correct groupings (e.g., *Thermococcus* with *Pyrococcus*, *Thermus* with *Deinococcus*) and wrong ones (e.g., the grouping in the Trp-tRNA synthetase tree of Rickettsia (an α-proteobacterium) with those of *Synechocystis* (a cyanobacterium)). There are too many of these aberrant grouping to be described in detail here, all the more so as that they are often not well supported. Although some of them can be explained by horizontal transfers, the remaining oddities should be explained by tree reconstruction artifacts (LBA, for example).

Saturation Analysis

The reanalysis of the six genes that previously supported the sisterhood between Archaebacteria and Eukaryota by using many new sequences led to a picture that was much more complex than that first reported. The failure to obtain the same phylogeny from different markers could thus be explained by the following: (1) the species tree was different from the gene tree, and (2) the tree reconstruction method was inappropriate. The first hypothesis was supported by the existence of several clearly enigmatic sequences, which implied at least several horizontal transfers but implied that it was almost-impossible to infer the good species tree. This point has been discussed in many papers (Doolittle 1998; Feng et al. 1997; Jain et al. 1999; Lawrence and Ochman 1998) and is not discussed in detail here. It should be noted, however, that even if horizontal transfers are relatively frequent and can severely disturb some gene phylogenies (Figs. 1–4), they do not mix up the genomes since the phylogeny based on gene content is similar to the phylogeny based on rRNA (Snel et al. 1999). The second hypothesis was of prime importance because (i) the phylogenetic relationships within the domains were incorrectly inferred, and (ii) the model of sequence evolution used was quite oversimplified (see Sullivan and Swofford 1997), as discussed in the accompanying paper (Lopez et al. 1999). One can conclude from all these analyses that the relationships among Eubacteria, Euryarchaeota, Crenarchaeota, and Eukaryota were still not solved, despite the use of six different genes.

In fact, such a question deals with very ancient events, at least 1 billion years ago and possibly more than 3 billion, and one should expect molecular phylogenetics to encounter many problems, since, for example, the complete mitochondrial genome was not able to recover the rodent monophyly (Philippe 1997), an event that occurred less than 100 million years ago. During more than 1 billion years, the evolutionary rate could have varied in the different lineages, generating erroneous phylogenies because of the LBA phenomenon. A more dramatic problem could be that numerous multiple substitutions occurred after the divergence of the three domains and masked the old phylogenetic signal. To test this hypothesis, the method of Philippe et al. (1994) was applied to the six data sets to evaluate the mutational saturation.

The principle of this method is to compare, for each pair of species, the number of observed differences and the number of substitutions inferred by the parsimony method, which is able to detect a fraction of the multiple substitutions occurring at a same position and thus gives an estimate of the real number of substitution. If the phylogenetic marker is saturated, the number of inferred substitutions will still increase, whereas the number of observed differences stays nearly constant, which generates a plateau in the graphical plot of the pairwise comparison. For the six genes analyzed, such a plateau was obviously present (Fig. 6). For example, the number of observed differences reached a maximum of 190 for ATPase (Fig. 6C), but the number of corresponding inferred substitutions varied from 200 up to 700. The plateau always contained all the pairwise comparisons of the paralogous sequences (displayed as an open circle). The saturation level within one orthologous gene was variable. The most saturated marker was the ATPase gene (Fig. 6C), because all the comparisons between a eubacterial sequence and its putative orthologous archaebacterial or eukaryotic sequence were located within the major plateau. The Ile-RS was almost as saturated, because the orthologous comparisons were mixed with the paralogous ones (Fig. 6B). Elongation factors were also highly saturated, as evidenced by a large plateau just below the plateau of the paralogous comparisons (Fig. 6D). The CPS was less saturated, because the orthologous comparisons displayed only a small tendency to form a plateau (Fig. 6A).

Since maximum parsimony is not the most efficient method to recover a phylogenetic tree and to detect multiple substitutions (Hasegawa et al. 1991), the same analysis was carried out with the use of the maximum likelihood method to infer the number of substitutions. A very similar pattern was observed (Fig. 7), and indeed the level of saturation appeared to be higher than with the parsimony estimation. This point was evidenced by the fact that the plateau of the paralogous comparisons began farther from the line with a slope equal to 1, which represented the theoretical case where no multiple substitutions occur. It was not surprising that the probabilistic approach allowed us to detect more multiple substitutions than the parsimony one, especially along the long unbroken branches. Moreover, it was likely that even the number of substitutions inferred by maximum likelihood was severely underestimated, at least for the very large distances.

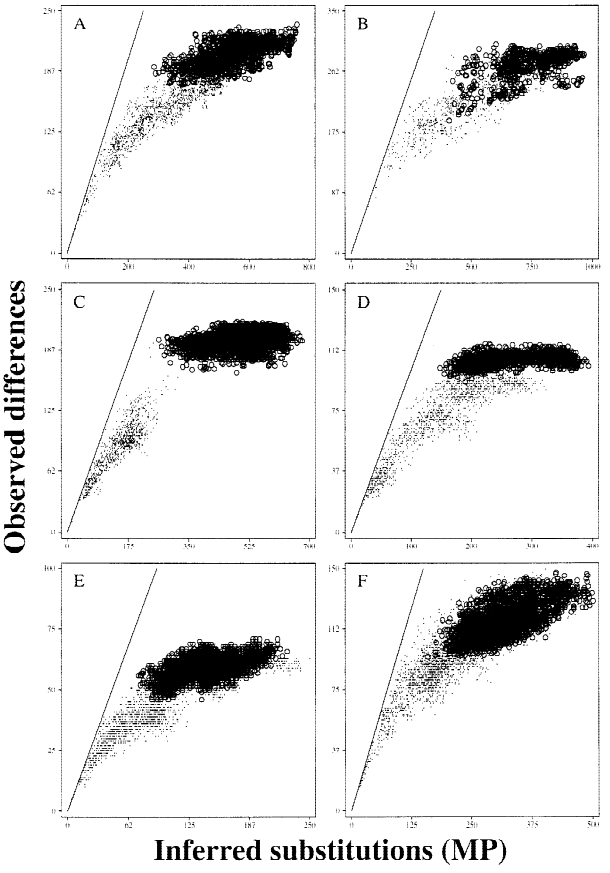

Fig. 6. Mutational saturation curves. **A,** CPS; **B,** Ile-tRS and Val-tRS; **C,** ATPases; **D,** EF-1α and EF-2; **E,** Trp-tRS and Tyr-tRS; **F,** SRP. *Y* axis: the observed number of differences between pairs of species sequences. *X* axis: the inferred number of substitutions between the same two sequences determined using the maximum-parsimony method. Each *dot* thus defines the observed versus the inferred number of substitutions for a given pair of sequences. It can be seen that in the six cases, the curve levels off after a given point, indicating that while the number of inferred mutations still increases (*X* axis), they are no longer detected as observed differences (leveling along the *Y* axis). Pairs of paralogous genes are represented by *open circles.* In the case of ATPase, the comparisons between Eubacteria and Archaebacteria/Eukaryota are also indicated by *open circles.* The *straight line* represents the ideal case, for which at most one substitution occurred by position.

As a result, the six genes used to root the universal tree of life were found to be highly saturated, probably much more than shown in Figs. 6 and 7. This raised a new interpretation of the inferred phylogeny. The reliability of the eubacterial rooting has been supported by the fact that the paralogous genes could have a constant evolutionary rate (Feng et al. 1997; Iwabe et al. 1989), thus avoiding the LBA artefact (Felsenstein 1978).

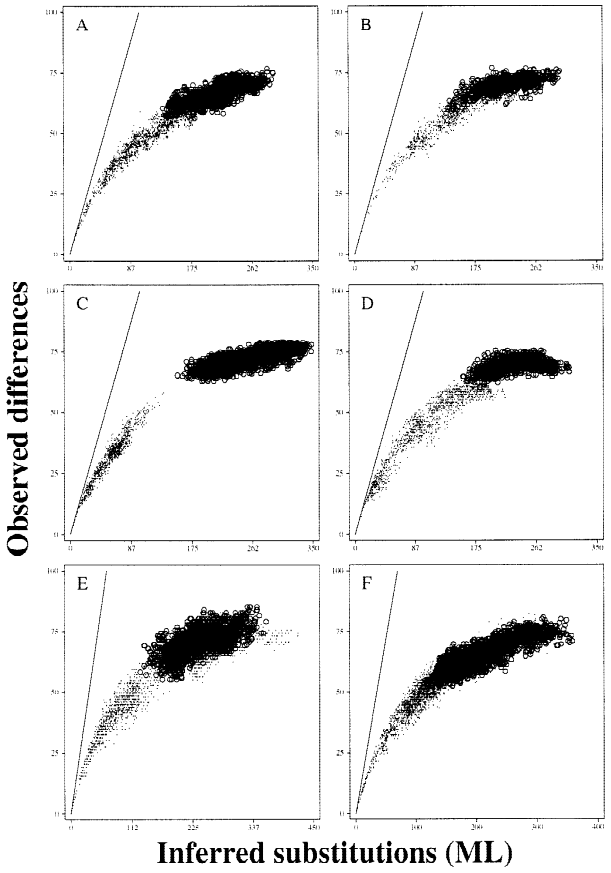

Fig. 7. Mutational saturation curves as in figure 5, except that the inferred number of substitutions between the same two sequences was determined using the maximum-likelihood method. The numbers of substitutions are represented as the frequency.

Brown and Doolittle (1997) also argue that orthologous proteins evolve at about the same rate in the three domains according to the relative rate test. But an important effect of mutational saturation is precisely an illusory molecular clockwise behavior of the phylogenetic marker, even if the evolutionary rate varies greatly between lineages (Philippe and Laurent 1998). To detect differences in evolutionary rate, the most commonly used method is indeed the relative rate test (Sarich and Wilson 1973). It consists in using an outgroup (O) and comparing the distance between it and two ingroup species, A and B. If the distance *d*(O,A) is significantly greater than *d*(O,B), then one can infer that species A has evolved faster than species B. On the other hand, if *d*(O,A) is equal to *d*(O,B), one can assume the constancy of the evolutionary rate. But mutational saturation produces exactly the result that the distance values reach a plateau (Figs. 6 and 7), irrespective of the real number of substitutions. In our case, saturation would be enough to make *d*(O,A) equal to *d*(O,B) even if species A does not evolve at the same rate as species B. The high level of saturation indicated that a relative rate test was inappropriate to detect any difference of evolutionary rates in the six paralogous genes.

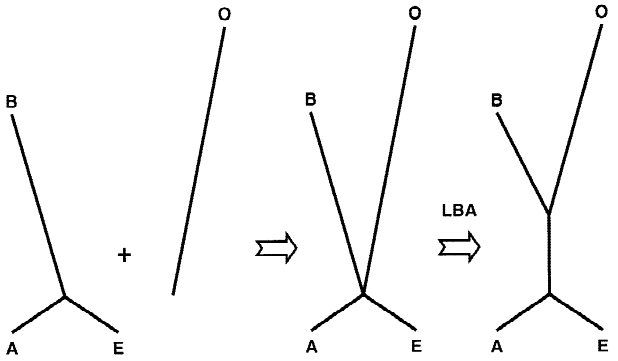

Fig. 8. The long branch attraction artifact and the rooting of the tree of life. **Left:** A hypothetical unrooted tree linking the three domains, for which the branch of Eubacteria (B) is much longer than those of Archaebacteria (A) and Eukaryota (E). The outgroup (O) represents a paralogous gene which obviously also has a very long branch. The resulting topology is very similar to the model used by Felsenstein (1978) to demonstrate the long branch attraction phenomenon. The root of the tree of life thus could artifactually be located in the longest branch.

Because of the mutational saturation, it was thus highly probable that the substitutions that occurred in the deep branches of the tree were completely masked by the innumerable substitutions that occurred later. The phylogenetic signal for ancient events could thus have been completely lost, suggesting a priori that all the phylogenies used to root the tree of life were prone to tree reconstruction artifact.

Eubacterial Rooting as the Result of Long Branch Attraction

Let us assume that, for a given gene, Eubacteria evolved faster than Eukaryota and Archaebacteria. When this gene is rooted through the addition of a paralogous gene, the phylogeny will contain two long branches (Eubacteria and outgroup) and two short branches (Eukaryota and Archaebacteria) as shown in Fig. 8. Such a topology is very similar to the one that has been used by Felsenstein (1978) to demonstrate the LBA artifact. It is likely that the two long branches will be grouped together because of this artifact, locating the root of the tree of life in the eubacterial branch.

In Fig. 9, the unrooted topologies of the ATPase, TyrtRS, SRP54, EF-1α, Ile-tRS, and 16S rRNA were displayed. The branch lengths were equal to the average distances from extant species to the trifurcating point estimated on an ML tree. A salient feature in this figure is that the branch lengths of the three domains were quite different according to the gene studied. For example, for the ATPase, the eubacterial branch was 6 times longer than the archaebacterial and eukaryotic ones, and for the rRNA, the eukaryotic branch was 1.4 times longer than the eubacterial one and 1.9 times longer than the archaebacterial one. This meant that the evolutionary rate var-

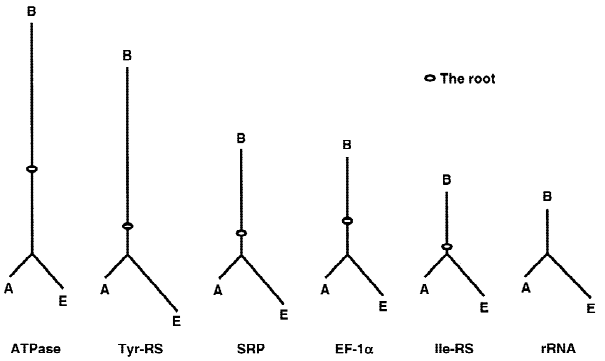

Fig. 9. Six unrooted trees linking the three domains. The branch lengths of each domain are represented as the average of the branch lengths on an ML tree between the trifurcation point and the species in the domain. When a paralogous gene is added to the analysis, the root is always located in the longest branch, as indicated by the *ellipse*. The long branch attraction artifact explains this phenomenon well (see Fig. 8).

ied greatly between genes and between domains and that probably none of these six genes was a good molecular clock. As discussed above, their apparently clock-like behavior is spurious and is due to a high level of mutational saturation. More interestingly, when these unrooted protein trees were rooted through the use of a paralogous gene, the root always fell within the longest branch, producing a result similar to a midpoint rooting approach. This strongly suggested that this rooting was artifactual, due to the LBA phenomenon. Further evidence for this hypothesis is provided by the ML analysis performed on a limited sample of species. The difference in likelihood between the eubacterial rooting and the eukaryotic rooting (i.e., the monophyly of prokaryotes) was proportional to the length of the eubacterial branch: about 15 SE for the ATPase, about 3 SE for the EF-1α, and only 1.3 SE for the Ile-RS. The scenario in Fig. 8 explaining the eubacterial rooting by the LBA artifact was thus probable.

This hypothesis also suggested an explanation for the following paradox. The six paralogous genes used in this study were highly saturated. This explained why the intradomain phylogeny, even their monophyly, was not correctly recovered and was supported by low bootstrap values (see Figs. 1–5). But it raised a serious question: Why were the deeper nodes such as the monophyly of the duplicated genes (in the majority of the cases) and the relationships between domains recovered and generally supported by high bootstrap values? The saturation should have erased all the signal for the relationship between the three domains, since the more ancient the phylogenetic signal is, the more the saturation should obscure it. If the tree reconstruction method inferred some groupings statistically supported by high bootstrap values despite the absence of phylogenetic signal, this must reflect some inconsistency of the reconstruction method, such as LBA. The best answer to the previous paradox, i.e., resolution of deep nodes despite the high level of mutational saturation, was therefore that the robust resolution of the rooting of the tree of life observed for these six genes was due to the LBA artefact. However, the explanation is probably more complex, because of the variation of invariant sites discussed in the companion paper (Lopez et al. 1999).

Exploration of a Possible Eukaryotic Rooting

We have shown in this paper that the rooting of the tree of life in the eubacterial branch has been based on unreliable phylogenies. To locate correctly the root of the universal tree, we must take up the challenge of inferring good phylogenies from highly saturated data. One way is to use the ML method with a very adequate model of sequence evolution (Sullivan and Swofford 1997). However, as discussed in the companion paper (Lopez et al. 1999) and in many others (Cao et al. 1998; Goldman et al. 1998; Halpern and Bruno 1998; Lockhart et al. 1998; Naylor and Brown 1998; Voelker and Edwards 1998; Yang et al. 1998), it is clear that current models poorly fit the data and thus we have no guarantee of finding the true phylogeny. Another way is to study the slowly evolving positions, which are much less saturated. We have applied this approach to the two genes that are apparently not affected by horizontal gene transfer or a paralogy problem, the elongation factors (Lopez et al. 1999), and the SRP (Brinkmann and Philippe 1999) and found that the eukaryotic rooting is the best-supported hypothesis.

This eukaryotic rooting would best explain the presence of many more eubacterial-like genes than eukaryotic-like ones in completely sequenced Archaebacterial genomes (Koonin et al. 1997), which cannot be easily explained in the frame of the current scenario. This will also best explain the presence of many unique processes in eukaryotes that involve the participation of structural RNAs or ribozymes reminiscent of the RNA world (Jeffares et al. 1998; Poole et al. 1998).

The existence of many eukaryotic features in the major archaebacterial cellular information processing systems (replication, transcription, and translation) is generally explained by the sisterhood of Archaebacteria and Eukaryota, implying that they are evolved states. But it can be also explained by an acceleration of the rate of evolution of these systems in the eubacterial lineage (and also by differential gene loss), which is equivalent to interpret them as primitive. The acceleration phenomenon can now be documented in the case of microsporidia, in which proteins involved in the translation machinery (rRNA, EF-1α, EF-2, and Ile-RS) all evolved at an accelerated evolutionary rate, leading to the artifactual early emergence of these peculiar fungi in the eukaryotic tree, because of the LBA artifact (Germot et al. 1997; Hirt et al. 1999).

A major difference between the Eubacteria and the

Archaebacteria in the transcription and translation systems is the smaller number of proteins in the Eubacteria. The differences in these systems are too complex to be simply explained by loss or gain of genes, but there is a clear trend toward simplicity in Eubacteria that could be interpreted by (1) the loss of these proteins in the Eubacteria or (2) their gain in the Archaebacteria. The second hypothesis fits well with the idea that Archaebacteria are en route toward the complex system of eukaryotes, but the selective constraints in favor of this hypothesis are very unclear, since apparently the archaebacterial system is not more efficient than the eubacterial one. In contrast, the first hypothesis is reasonable since it is advantageous to perform the same work with fewer proteins.

The loss of some proteins in the eubacterial translation apparatus could have produced an acceleration of the evolutionary rate of remaining proteins. Indeed, as first noted by Dickerson (1971), the physical contacts between a protein and several partners (proteins or nucleic acids) induce constraints for its evolution. If one or several of these contacts disappear, the corresponding constraint is thus removed and the sequence will evolve faster. This could well explain why some archaebacterial and eukaryotic proteins involved in identical protein–protein interactions with ribosomes, such as elongation factors, have conserved more similarities between them (ancestral characters) than each one with the eubacterial proteins (see Brown and Doolittle 1997).

A strong selective pressure that could have favored the loss of many proteins in Eubacteria is the coupling of transcription and translation. For that, the mRNA must be sufficiently accessible to the ribosome and its accessory proteins and must not be masked by the transcriptional apparatus. As a result, simply because of sterical hindrance, the coupling of transcription and translation in Eubacteria could have been favored by the loss of many proteins. The coupling of transcription and translation is generally assumed in Archaebacteria because of the absence of nucleus, but without any experimental evidence. If this coupling indeed exists, the similarity between Archaebacteria and Eukaryota proteins of the translation apparatus would be all the more striking because their functions would be quite different. If it were not, it would be a support for the proposed acceleration of the evolutionary rate of the transcriptional and translational proteins in Eubacteria and thus for an artefactual rooting in this branch.

Another aspect of eubacterial evolution that could have reduced the number of proteins involved in various cellular processes and led to a general simplification of their molecular mechanisms is nonorthologous gene displacement (Forterre and Philippe 1999). Comparative genomics have shown that this phenomenon has occurred frequently during the divergence of the three domains and this could explain divergent rates of evolution such as those responsible for the long bacterial branches of many universal trees.

Acknowledgments. We would like to thank Henner Brinkmann, Jacqueline Laurent, Philippe Lopez, David Moreira, and Miklos Müller for helpful comments and readings of the manuscript. We would also like to thank Drs. J. Miller and S. Fitz-Gibbon, as well as TIGR for unpublished sequences and La Foundation des Treilles for giving a starting point to our collaboration. We acknowledge Karine Budin for her help in the elaboration of the figures. Genome sequencing was accomplished with support from DOE, MGRI, NIAID, and NIDR.

References

Adachi J, Hasegawa M (1996) MOLPHY version 2.3: Programs for molecular phylogenetics based on maximum likelihood. Comput Sci Monogr 28:1–150

Baldauf SL, Palmer JD, Doolittle WF (1996) The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. Proc Natl Acad Sci USA 93:7749–7754

Brinkmann H, Philippe H (1999) Archaea sister-group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. Mol Biol Evol 16:817–825

Brown JR, Doolittle WF (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc Natl Acad Sci USA 92:2441–2445

Brown JR, Doolittle WF (1997) Archaea and the prokaryote-to-eukaryote transition. Microbiol Mol Biol Rev 61:456–502

Brown JR, Robb FT, Weiss R, Doolittle WF (1997) Evidence for the early divergence of tryptophanyl- and tyrosyl-tRNA synthetases. J Mol Evol 45:9–16

Brown JR, Zhang J, Hodgson JE (1998) A bacterial antibiotic resistance gene with eukaryotic origins. Curr Biol 8:R365–R367

Cao Y, Waddell PJ, Okada N, Hasegawa M (1998) The complete mitochondrial DNA sequence of the shark *Mustelus manazo:* Evaluating rooting contradictions to living bony vertebrates. Mol Biol Evol 15:1637–1646

Dickerson RE (1971) The structures of cytochrome c and the rates of molecular evolution. J Mol Evol 1:26–45

Doolittle WF (1998) You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet 14:307–311

Doolittle WF, Brown JR (1994) Tempo, mode, the progenote, and the universal root. Proc Natl Acad Sci USA 91:6721–6728

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27:401–410

Feng DF, Cho G, Doolittle RF (1997) Determining divergence times with a protein clock: Update and reevaluation. Proc Natl Acad Sci USA 94:13028–13033

Forterre P (1997) Protein versus rRNA: rooting the universal tree of life? ASM News 63:89–92

Forterre P, Philippe H (1999) Where is the root of the universal tree of life? BioEssays (in press)

Forterre P, Benachenhou-Lahfa N, Confalonieri F, Duguet M, Elie C, Labedan B (1992) The nature of the last universal ancestor and the root of the tree of life, still open questions. BioSystems 28:15–32

Germot A, Philippe H, Le Guyader H (1997) Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae.* Mol Biochem Parasitol 87:159–168

Gogarten JP, Hilario E, Olendzenski L (1996) Gene duplications and horizontal gene transfer during early evolution. In: Roberts DM, Sharp P, Alderson G, Collins M (eds) Evolution of microbial life. Cambridge University Press, Cambridge, pp 109–126

Gogarten JP, Kibak H, Dittrich P, et al. (1989) Evolution of the vacu-

olar H^+-ATPase: Implications for the origin of eukaryotes. Proc Natl Acad Sci USA 86:6661–6665

Goldman N, Thorne JL, Jones DT (1998) Assessing the impact of secondary structure and solvent accessibility on protein evolution. Genetics 149:445–458

Gribaldo S, Cammarano P (1998) The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. J Mol Evol 47:508–516

Gupta RS, Golding GB (1993) Evolution of HSP70 gene and its implications regarding relationships between archaebacteria, eubacteria, and eukaryotes. J Mol Evol 37:573–582

Halpern AL, Bruno WJ (1998) Evolutionary distances for protein-coding sequences: Modeling site-specific residue frequencies. Mol Biol Evol 15:910–917

Hasegawa M, Kishino H, Saitou N (1991) On the maximum likelihood method in molecular phylogenetics. J Mol Evol 32:443–445

Hashimoto T, Sanchez LB, Shirakura T, Muller M, Hasegawa M (1998) Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny. Proc Natl Acad Sci USA 95:6860–6865

Hennig W (1966) Phylogenetic systematics. University of Illinois Press, Urbana

Hirt RP, Logsdon JM Jr, Healy B, Dorey MW, Doolittle WF, Embley TM (1999) Microsporidia are related to fungi: Evidence from the largest subunit of RNA polymerase II and other proteins. Proc Natl Acad Sci USA 96:580–585

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

Jain R, Rivera MC, Lake JA (1999) Horizontal gene transfer among genomes: The complexity hypothesis. Proc Natl Acad Sci USA 96:3801–3806

Jeffares DC, Poole AM, Penny D (1998) Relics from the RNA world. J Mol Evol 46:18–36

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Koonin EV, Mushegian AR, Galperin MY, Walker DR (1997) Comparison of archaeal and bacterial genomes: Computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. Mol Microbiol 25:619–637

Lawrence JG, Ochman H (1998) Molecular archaeology of the *Escherichia coli* genome. Proc Natl Acad Sci USA 95:9413–9417

Lawson FS, Charlebois RL, Dillon JA (1996) Phylogenetic analysis of carbamoylphosphate synthetase genes: Complex evolutionary history includes an internal duplication within a gene which can root the tree of life. Mol Biol Evol 13:970–977

Lockhart PJ, Steel MA, Barbrook AC, Huson D, Charleston MA, Howe CJ (1998) A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. Mol Biol Evol 15: 1183–1188

Lopez P, Forterre P, Philippe H (1999) The root of the tree of life in the light of the covarion model. J Mol Evol (in press)

Martin W, Muller M (1998) The hydrogen hypothesis for the first eukaryote. Nature 392:37–41

Moreira D, Lopez-Garcia P (1998) Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: The synthrophic hypothesis. J Mol Evol 47:517–530

Naylor GJP, Brown WM (1998) Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. Syst Biol 47:61–76

Olsen GJ, Woese CR (1997) Archaeal genomics: An overview. Cell 89:991–994

Philippe H (1993) MUST, a computer package of Management Utilities for Sequences and Trees. Nucleic Acids Res 21:5264–5272

Philippe H (1997) Rodent monophyly: Pitfalls of molecular phylogenies. J Mol Evol 45:712–715

Philippe H, Adoutte A (1998) The molecular phylogeny of Eukaryota: Solid facts and uncertainties. In: Coombs G, Vickerman K, Sleigh M, Warren A (eds) Evolutionary relationships among Protozoa. Chapman & Hall, London, pp 25–56

Philippe H, Laurent J (1998) How good are deep phylogenetic trees? Curr Opin Genet Dev 8:616–623

Philippe H, Sörhannus U, Baroin A, Perasso R, Gasse F, Adoutte A (1994) Comparison of molecular and paleontological data in diatoms suggests a major gap in the fossil record. J Evol Biol 7:247–265

Poole AM, Jeffares DC, Penny D (1998) The path from the RNA world. J Mol Evol 46:1–17

Ribas de Pouplana L, Frugier M, Quinn CL, Schimmel P (1996) Evidence that two present-day components needed for the genetic code appeared after nucleated cells separated from eubacteria. Proc Natl Acad Sci USA 93:166–170

Rivera MC, Lake JA (1992) Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. Science 257:74–76

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

Sarich VM, Wilson AC (1973) Generation time and genomic evolution in primates. Science 179:1144–1147

Snel B, Bork P, Huynen MA (1999) Genome phylogeny based on gene content. Nature Genet 21:108–110

Sogin ML (1991) Early evolution and the origin of eukaryotes. Curr Opin Genet Dev 1:457–463

Stetter K (1992) Life at the upper temperature border. In: Trân Thanh Van J, Trân Thanh Van K, Mounolou J, Schneider J, McKay C (eds) Frontiers of life. Editions Frontières, Gif-sur-Yvette, France

Sullivan J, Swofford DL (1997) Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J Mammal Evol 4:77–86

Swofford DL (1993) PAUP: Phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign

Voelker G, Edwards SV (1998) Can weighting improve bushy trees? Models of cytochrome b evolution and the molecular systematics of pipits and wagtails (Aves: Motacillidae). Syst Biol 47:589–603

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

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 87:4576–4579

Yang Z, Nielsen R, Hasegawa M (1998) Models of amino acid substitution and applications to mitochondrial protein evolution. Mol Biol Evol 15:1600–1611

Zillig W (1987) Eukaryotic traits in Archaebacteria. Could the eukaryotic cytoplasm have arisen from archaebacterial origin? Ann NY Acad Sci 503:78–82