Universal tree of life

SIR — Phylogenetic comparisons of complete small-subunit ribosomal RNA (rRNA) sequences have changed well-established ideas about early events in the evolution of eukaryotes. Nonetheless, the incongruence of rRNA-based phylogenies with molecular trees derived from elongation factors or DNA-dependent RNA polymerases presents a challenge to molecular evolutionists. In rRNA phylogenies the earliest branches are nonphotosynthetic, amitochondriate taxa. They are separated from the more recently diverged kingdoms of plants, animals and fungi by a series of independent protist branches, including Entamoeba[1].

Hasegawa and Hashimoto in Scientific Correspondence[2] suggest that unusual G+C compositions erroneously place the diplomonad *Giardia lamblia* (74.7% G+C) and the microsporidian *Vairimorpha necatrix* (37.5% G+C) rather than *Entamoeba histolytica* (the earliest divergence in elongation factor and DNA-dependent RNA polymerase phylogenies) at the base of the eukaryotic tree. We share their concern about potential misleading effects of biased nucleotide compositions in rRNAs used to infer evolutionary histories. Because there is no accepted theoretical method for compensating for effects of biased nucleotide compositions in phylogenetic inferences, we have sequenced the small-subunit rDNA of the diplomonad *Hexamita inflata*[3] (50% G+C) and the microsporidian *Spraguea lophii* (49% G+C). Using distance, parsimony and maximum-likelihood methods, the overall picture of eukaryote small-subunit rRNA phylogeny remains unchanged. Diplomonads, trichomonads and microsporidians represent the earliest diverging lineages, but their relative branching order is influenced by the G+C compositions of prokaryote outgroups. By contrast, *Entamoeba* consistently diverge higher in the tree after the separation of euglenoids/kinetoplasts, acellular slime moulds and amoeboflagellates.

Reconstruction of phylogenetic history from molecular sequence data is a probability exercise based on a specific model of genetic change. When various models or different genes are used, statistical measurements can provide strong support for contradictory phylogenies. Deciding between discordant branching patterns frequently reduces to arguments about the "correct model" or "best molecular document" for inferring evolutionary history. In the final analysis, conflicting molecular data sets can be judged by considering the biology of the considered organisms. The fit between trees derived from the small-subunit rDNA data and morphological and ultrastructure data is unmatched by any other gene used to infer phylogenetic frameworks. When measured by these criteria, the reliability of rRNA-based phylogenies is remarkable and unparalleled in phylogenetic reconstructions of the universal tree of life.

Mitchell L. Sogin
Gregory Hinkle
Detlef D. Leipe
Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA

1. Sogin, M. L. *Curr. Opin. Genet. Dev.* **1**, 457–463 (1992).
2. Hasegawa, M. & Hashimoto, T. *Nature* **361**, 23 (1993).
3. Leipe, D. D., Gunderson, J. H., Nerad, T. A. & Sogin, M. L. *Molec. biochem. Parasit.* (in the press).

SIR — Hasegawa and Hashimoto correctly point out[1] that rRNA trees might be misleading in defining the evolutionary relationships between very distantly related organisms. They suggest instead that protein trees should be used for this purpose. Indeed, protein trees encompassing the three domains of life (archaebacteria, eubacteria and eukaryotes) are becoming more popular as the number of available protein sequences from archaebacteria increases. In particular, several authors have rooted the universal tree of life in the eubacterial branch, using two composite trees of duplicated protein families (elongation factors and membrane ATPases α and β subunits)[2,3].

Nevertheless, protein trees could be as misleading as rRNA trees for very distantly related organisms. Well-known mistakes in extrapolating species trees from protein trees have arisen from lateral gene transfers, unrecognized paralogy (duplication of genes before separation of the lineages under investigation) and unequal rates of evolution. We recently summarized[4] all the data obtained so far by comparing archaebacterial housekeeping proteins with their homologues from eubacteria and eukaryotes at the sequence level.

As we expected, our analysis revealed contradictions between protein trees and the rRNA tree, and between protein trees themselves. We also find that the composite ATPase and elongation factor trees used to root the tree of life are misleading. First, the finding of V-type ATPases in two eubacteria suggested that V- and F0/F1-type ATPases are paralogous. This was recently confirmed by the discovery of an F0/F1-type ATPase in an archaebacterium which already harbours a V-type enzyme[5]. As a consequence, the eubacterial rooting previously obtained from ATPase evolution is inconsistent as it was based on phylogenetic trees in which these two paralogous families were mixed in single trees. Second, cladistic analysis suggests that the elongation factors EF1α(Tu) and EF2(G) are too divergent to root with confidence the composite tree of their two families, again invalidating the inferred eubacterial rooting.

The difficulty of rooting the universal tree of life using protein trees is also emphasized by our recent analysis of glutamate dehydrogenase phylogeny[6]. Trying to root one subfamily of glutamate dehydrogenase harbouring representatives of the three domains of life, using the paralogous subfamily as an outgroup, we obtained different roots according to the method of tree construction used. Interestingly, the root was located either in the eukaryotic branch or in the archaebacterial domain, but never in the eubacterial branch.

Caution is therefore necessary in drawing definite conclusions from either rRNA or protein-tree analyses. In particular, it is by no means clear that the problem of rooting the tree of life is now solved.

Patrick Forterre
Nadia Benachenhou-Lafha
Bernard Labedan
Institut de Génétique et Microbiologie, Université Paris-Sud, Centre d'Orsay, 91405 Orsay Cedex, France

1. Hasegawa, M. & Hashimoto, T. *Nature* **361**, 23 (1993).
2. Gogarten, J. P. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 6661–6665 (1989).
3. Iwabe, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 9355–9359 (1989).
4. Forterre, P. *et al. Biosystems* **28**, 15–32 (1993).
5. Sumi, M. *et al. FEBS Lett.* **3**, 207–210 (1992).
6. Benachenhou-Lafha, N., Forterre, P. & Labedan, B. *J. molec. Evol.* **36**, 335–346 (1993).

Phytoplankton productivity?

SIR — Falkowski and Wilson[1] elegantly demonstrate that historical Secchi disk measurements show no evidence of a significant increase in North Pacific phytoplankton biomass in this century. Unfortunately, they also conclude that increased absorption of anthropogenic CO_2 by phytoplankton in the North Pacific is therefore equally unlikely. This conclusion assumes that phytoplankton productivity can be usefully indexed by biomass levels, which is inappropriate when applied to the North Pacific.

Secchi disk measurements in the North Pacific, while clearly a surprisingly sensitive measure of phytoplankton biomass, cannot be used to infer changes in phytoplankton productivity. Unlike the North Atlantic, phytoplankton biomass in the North Pacific is controlled by zooplankton grazing throughout the year[2], including the spring bloom. In-