References

1 Nagafuchi, A. *et al.* (1987) *Nature* 329, 341–343
2 Edelman, G. M. *et al.* (1987) *Proc. Natl Acad. Sci. USA* 84, 8502–8506
3 Schneider, I. (1972) *J. Embryol. Exp. Morph.* 27, 353–365
4 Schneuwly, S., Klemenz, R. and Gehring, W. J. (1987) *Nature* 339, 61–64
5 Rubin, G. M. and Spradling, A. C. (1983) *Nucleic Acids Res.* 11, 6341–6351
6 Bunch, T. A., Grinblat, Y. and Goldstein, L. S. B. (1988) *Nucleic Acids Res.* 16, 1043–1061
7 Jockerst, R. S., Weeks, J. R., Zehring, W. A. and Greenleaf, A. L. (1989) *Mol. Gen. Genet.* 215, 266–275
8 Bourouis, M. and Jarry, B. (1983) *EMBO J.* 2, 1099–1104
9 Rio, D. C. and Rubin, G. M. (1985) *Mol. Cell. Biol.* 5, 1833–1838
10 Bastiani, M. J., Harrelson, A. L., Snow, P. M. and Goodman, C. S. (1987) *Cell* 48, 745–755
11 Patel, N. H., Snow, P. M. and Goodman, C. S. (1987) *Cell* 48, 975–988

12 Snow, P. M., Bieber, A. J. and Goodman, C. S. (1989) *Cell* 59, 313–323
13 Zinn, K., McAllister, L. and Goodman, C. S. (1988) *Cell* 53, 577–587
14 Grenningloh, G. *et al.* Cold Spring Harbor Symp. Quant. Biol. (in press)
15 Elkins, T. *et al.* (1990) *J. Cell Biol.* 110, 1825–1832
16 Steinberg, M. (1963) *Science* 141, 401–408
17 Bieber, A. J. *et al.* (1989) *Cell* 59, 447–460
18 Harrelson, A. L. and Goodman, C. S. (1988) *Science* 242, 700–708
19 Fujita, S. C. *et al.* (1982) *Proc. Natl Acad. Sci. USA* 79, 7929–7933
20 Van Vactor, D., Jr, Krantz, D. E., Reinke, R. and Zipursky, S. L. (1988) *Cell* 52, 281–290
21 Reinke, R., Krantz, D. E., Yen, D. and Zipursky, S. L. (1988) *Cell* 52, 291–301
22 Krantz, D. E. and Zipursky, L. S. (1990) *EMBO J.* 9, 1969–1977
23 Hashimoto, C., Hudson, K. L. and Anderson, K. V. (1988) *Cell* 52, 269–279
24 Keith, F. J. and Gay, N. J. (1990) *EMBO J.* 9, 4299–4306
25 Fehon, R. G. *et al.* (1990) *Cell* 61, 523–534
26 Wharton, K. A., Johansen, K. M., Xu, T. and Artavanis-Tsakonas, S. (1985) *Cell* 43, 567–581
27 Vässin, H., Bremer, K. A., Knust, E. and Campos-Ortega, J. (1987) *EMBO J.* 6, 3431–3440
28 Hortsch, M. *et al.* (1990) *Development* 110, 1327–1340
29 De la Escalera, S. *et al.* (1990) *EMBO J.* 9, 3593–3601
30 Barthalay, Y. *et al.* (1990) *EMBO J.* 9, 3603–3609
31 Hatta, K., Nose, A., Nagafuchi, A. and Takeichi, M. (1988) *J. Cell Biol.* 106, 873–881
32 Matsunaga, M., Hatta, K., Nagafuchi, A. and Takeichi, M. (1988) *Nature* 334, 62–64
33 Mege, R-M. *et al.* (1988) *Proc. Natl Acad. Sci. USA* 85, 7274–7278
34 Friedlander, D. R., Mege, R-M., Cunningham, B. A. and Edelman, G. M. (1989) *Proc. Natl Acad. Sci. USA* 86, 7043–7047

LETTERS

Tracing origins with molecular sequences: rooting the universal tree of life

In his recent *TIBS* article[1] Lake correctly pointed out the importance of molecular sequences in determining the phylogenetic relationships among diverse groups of organisms. The main emphasis of his article was on results that were found to be unreliable either because of insufficient data or because of possible artifacts stemming from unequal rates of evolution. I fear that many readers will get the impression that molecular data used as evolutionary markers are a very unreliable tool.

All algorithms available for phylogenetic analyses rely on assumptions that are sometimes difficult to verify but artifacts and unreliable results can usually be easily detected if the investigator takes care to compare the outcome of different algorithms and alignments. In many cases, distance matrix, maximum likelihood, parsimony and evolutionary parsimony analyses do give the same results (see, for example, Ref. 2). Thus, although the individual algorithms tend to overestimate the reliability of the results, and ignore a possible bias due to a particular alignment or algorithm, taken together they lend credibility to the obtained results.

A good case to illustrate this point is the rooting of the universal tree of life by means of genes that had already

undergone a gene duplication in the last common ancestor. By use of the DNA sequences encoding the catalytic (i.e. ATP hydrolysing) and non-catalytic (i.e. ATP binding, but not hydrolysing) subunits of F-, V- (vacuolar) and archaebacterial proton pumping ATPases it was shown that these ATPases are homologous to each other[2,3]. In addition, analysis showed that the gene duplication that gave rise to the catalytic and non-catalytic subunits occurred before the lines that lead to the three Urkingdoms or domains separated from each other[2]. Thus one can use the non-catalytic

subunits as an outgroup to root a tree which uses the catalytic subunits as markers for the organismal evolution (see Fig 1). Use of the non-catalytic subunit as an organismal marker and the catalytic subunit as an outgroup gives the same result.

Use of ATPase subunits as evolutionary markers suggests that the archaebacteria branch off from the line that leads from the last common ancestor to the eukaryotes. This result has also been obtained for *Sulfolobus*[1], *Methanococcus*[4], *Methanosarcina* (H. Kibak, J. P. Gogarten and L. Taiz, unpublished; sequence from
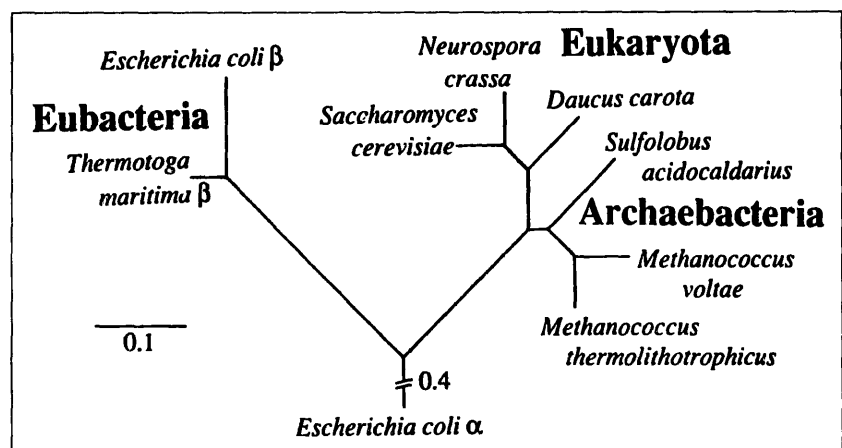

Figure 1
Phylogenetic tree showing the relations between the three Urkingdoms (domains). The topology and the branch lengths were calculated using Felsenstein's maximum likelihood method[14]. Branches are scaled in terms of the probability for change of the first base of the codon. Parameters for the algorithm, sequences and their alignment were as described in Ref. 4. All branch lengths were calculated to be positive at the 1% significance level. Using evolutionary parsimony the archaebacterial tree was significantly supported with p<2%.

Ref. 5), and *Halobacterium*[6]. The placement of the root between eubacteria on the one hand and archaebacteria and eukaryotes on the other is also supported by a 90 amino acid-long insertion (termed non-homologous region; Ref. 3) that is present in the archaebacterial and eukaryotic V-ATPase catalytic subunits, but is absent in the catalytic subunit of the eubacterial F-ATPase and in all non-catalytic subunits (i.e. in the outgroup).

These results were obtained with many different methods (including evolutionary parsimony) and alignments, thus artifacts due to unequal rates or biased alignments can be excluded. Analyses of other duplicated genes (tRNA[Met]s, dehydrogenases and elongation factors) suggest the same location of the root[7], thus making lateral gene transfer an unlikely explanation for the result obtained (one would have to invoke the lateral transfer of a substantial part of the genome). Although these results confirm the old notion of archaebacteria as proto-eukaryotes[8,9], they were far from being expected. Widely believed scenarios contradictory to these findings were that (1) the three Urkingdoms[10] had separated from each other at a very early stage of evolution (see, for example, Ref. 9) and 'equidistant from one another' (Ref. 10; however, the authors point out that their distance measure is not necessarily proportional to time); and (2) during the early evolution many events of lateral gene transfer were expected. In contrast to these expectations the findings outlined above suggest that the last common ancestor of all existing life was already highly developed. This ancestral organism was not a primitive progenote, defined as a organism that existed before the relationship between pheno- and genotype evolved[11].

The placement of the root in the universal tree of life between (eubacteria) and (archaebacteria and eukaryotes) is related to, but should not be confused with, the question eocyte versus archaebacterial tree. Whereas in the former case the data seem to be convincing (see, for example, Ref. 12), the latter case is still not settled. In other words, all archaebacteria branch off from the line leading from the common ancestor to the eukaryotes; however, it is not clear if all archaebacteria branch off in a single node (i.e. they are monophyletic, this corresponds to the archaebacterial tree) or if they represent separate side branches (the archaebacteria are paraphyletic, if the group containing *Sulfolobus* constitutes the most recent branch, this corresponds to the eocyte tree). Many recent publications favor the archaebacterial over the eocyte grouping (Refs 4, 13; Ref. 1 and references therein). However, due to the bias introduced by alignments

and algorithms, the obtained significance levels and probabilities for the different trees are not yet sufficient to settle this case unambiguously.

References

1 Lake, J. A. (1991) *Trends Biochem. Sci.* 16, 46–50

2 Gogarten, J. P. *et al.* (1989) *Proc. Natl Acad. Sci. USA* 86, 6661–6665

3 Zimniak L. *et al.* (1988) *J. Biol. Chem.* 263, 9102–9112

4 Gogarten, J. P. *et al.* (1989) *Z. Naturforsch.* 44c, 641–650

5 Inatomi, K-I., Eya, S., Maeda, M. and Futai, M. (1989) *J. Biol. Chem.* 264, 10954–10959

6 Mukohata, Y. *et al.* (1990) *Proc. Japan Acad.* 66, 63–67

7 Iwabe, N. *et al.* (1989) *Proc. Natl Acad. Sci. USA* 86, 9355–9359

8 Searcy, D. G., Stein, D. B. and Searcy, K. B. (1981) *Ann. N. Y. Acad. Sci.* 361, 312–324

9 Woese C. R., Magrum, L. J. and Fox, G. E. (1978) *J. Mol. Evol.* 11, 245–252

10 Woese, C. R. and Fox, G. E. (1977) *Proc. Natl Acad. Sci. USA* 74, 5088–5090

11 Woese, C. R. and Fox, G. E. (1977) *J. Mol. Evol.* 10, 1–6

12 Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) *Proc. Natl Acad. Sci. USA* 87, 4576–4579

13 Iwabe, N. *et al.* (1991) *J. Mol. Evol.* 32, 70–78

14 Felsenstein, J. (1981) *J. Mol. Evol.* 17, 368–376

TIMOTHY P. LINKKILA AND JOHANN PETER GOGARTEN

Department of Molecular and Cell Biology, University of Connecticut, 75 North Eagleville Road, Storrs, CT 06269, USA.

A monophyletic holophyletic archaeal domain versus the 'eocyte tree'

In a recent article in *TIBS*, James Lake repeats arguments, already refuted elsewhere, in favour of the so-called eocyte phylogenic tree in which the eocytes (Crenarchaeota[1]) appear linked to eukaryotes and the halobacteria to the eubacteria[2]. Lake tries to disprove an overwhelming body of evidence for the monophyletic holophyletic nature of the archaebacteria (Archaea[1]) including the phylogenetic tree of large RNA polymerase components[3-11] by four main assertions.

(1) Lake claims that trees containing branches of differing length are distorted by unequal rate effects without considering the extent of the branch length differences. As Gouy and Li[10] before, we have used computer simulation experiments to estimate the limits beyond which such distortions occur. Only extremely frequent mutations that left less than 5% residual identity between the mutated protein sequences and gave branch length ratios of about ten led to incorrect branching orders in trees calculated according to Fitch and Margoliash[12]. In the tree constructed using RNA polymerase large components, the corrected[13] maximal branch length ratio is around two and the residual identity is more than 25%. Essentially the same results were obtained using DNA sequences and the maximum likelihood and DNA bootstrap algorithms of Felsenstein's Phylogeny Inference Package[14]. These results make unequal rate artifacts in the branching order highly improbable. All algorithms, including the evolutionary parsimony

method[15] of Lake, gave the archaebacterial tree with the eukaryotic RNA polymerase pol 1 lineage sharing a separate bifurcation with the eubacteria.

(2) Lake claims proteins to be unsuitable for molecular phylogeny because they would evolve about twice as fast as rRNAs. Like rRNAs, RNA polymerase sequences contain highly conserved regions. The alignment of sequences of 20 different amino acids is considerably less ambiguous than that of sequences of four different nucleotides. The total number of nucleotide positions in an alignment of the two largest components of RNA polymerases is about 4800, not considering the third positions in the codons. More than 3000 positions (about twice the total number in 16S rRNA) are without gaps and unambiguously aligned. In contrast, Gouy and Li[10] have shown that the number of positions available in 16S rRNA does not allow the construction of a significant evolutionary parsimony tree.

(3) Another argument used by Lake against the validity of our RNA polymerase component tree is that it shows 'the eubacteria as a subgroup of the eukaryotes' and is thus 'biologically untenable'. After checking the significance of the tree topology in several independent ways, we have offered two alternative explanations[9] in line with the data: a reduction hypothesis and, more probable in our opinion, a fusion hypothesis in which the eukaryotes are regarded as a bi- or oligophyletic chimera rather than a monophyletic lineage.

(4) Lake claims that biochemical evidence supports the eocyte tree. With one exception (the organization of rRNA operons) his evidence for a specific relatedness of halobacteria and eubacteria is invalid. For example, the existence of a box A as an anti-termination signal in the spacer of the