

Genomics update

Radical views of the Tree of Life

Howard Ochman*

*Department of Biochemistry and Molecular Biophysics,
University of Arizona, Tucson, Arizona 85718, USA.*

The tripartite division of cellular organisms into three domains – the *Bacteria*, the *Archaea* and the *Eukarya* – has become so ingrained that most of us have all but forgotten (or did not even know) that the ancestry and relationships among these ancient lineages are far from settled. The most widely accepted view, which emerged primarily from analyses of anciently duplicated genes (Gogarten *et al.*, 1989; Iwabe *et al.*, 1989), places *Bacteria* as the basal lineage and distinct from the common ancestor of archaea and eukaryotes. Accordingly, these studies frame the *Archaea*, despite their appellation, as archaic but not primordial, and the three domains are each viewed as separate lineages, with archaea and eukaryotes as sister taxa.

Early on, Lake (1988) developed an algorithm that positioned a clade of thermophilic, sulfur-metabolizing archaea (the *Crenarchaeota* or 'eocytes') as more closely related to eukaryotes than to other archaea, and the archaeal *Holobacteria* as more closely related to *Eubacteria*. This upended the view that *Bacteria*, *Archaea* and *Eukarya* were separate divisions; however, subsequent analyses of small- and large-subunit rRNA sequences in more conventional phylogenetic frameworks upheld the *Archaea* as a cohesive group (Gouy and Li, 1989). And although there has since been some tantalizing evidence that eukaryotes emerged from within the *Archaea*, particularly from analyses of elongation factor Tu sequences (Rivera and Lake, 1992; Baldauf *et al.*, 1996), the 'eocyte' hypothesis was all but forgotten . . . until last year.

The incentive to re-examine the deepest branches in the Tree of Life was no doubt impelled by the current number and diversity of genome sequences. The addition of genes and taxa has helped resolve some of the thorniest phylogenetic issues and, in this case, can help establish the degree of support for the 3-domain tree, the eocyte tree, or any of the other hypotheses that have

been posited over the years (Embley and Martin, 2006; Poole and Penny, 2007).

To examine the origin of the eukaryotes, Pisani *et al.* (2007) applied a supertree approach to each of three large overlapping data sets containing either 0, 8 or 21 eukaryotic genomes, in order to monitor the effects of taxon sampling on the resulting branching orders. Supertrees are phylogenies that are assembled by merging sets of more limited trees, each based on a gene that is not necessarily present in all taxa. Their results resolved bacteria and archaea as separate groups, but favour a chimeric origin of eukaryotes, as the majority of eukaryotic genes with prokaryotic homologues derive from multiple symbioses. These findings help explain how different genes originated and why particular genes show distinct phylogenetic histories, but they do not ascertain whether the cellular lineage leading to eukaryotes – i.e. the lineage that suffered these varied symbioses – preceded, or arose from within, the *Archaea*.

In a parallel approach, Yutin *et al.* (2008) sifted through nearly 1000 sets of putative orthologues present in all three domains and identified more than a hundred genes that provided sufficient signal to test the relationship of eukaryotes to the two major phyla within the *Archaea*. Although several individual genes showed closer alliance to the *Crenarchaeota*, their analyses, for the most part, placed eukaryotes outside the extant archaeal diversity, lending support to the 3-domain tree.

Numerous factors, aside from the presence of genes with different evolutionary histories, can complicate the accurate reconstruction of deep evolutionary relationships. Individual genes, or entire genomes, display different evolutionary rates due to selection or to the underlying pattern of mutations. For example, divergent organisms with AT-rich genomes often encode proteins that are enriched in amino acids specified by AT-rich codons, such as leucine, isoleucine and lysine, and such proteins are prone to converge instead of diverge with time, thereby distorting their actual relationships. This is rarely a problem when analysing closely related organisms, which tend to have similar rates and patterns of substitutions, but one might expect its effects to increase with organismal diversity and therefore, to be most pronounced when encompassing the diversity of all life forms.

*For correspondence. E-mail hochman@email.arizona.edu; Tel. 520-626-8355; Fax 520-621-3709.

2 Genomics update

In that most previous assessments failed to consider substitutional variation across organisms, Cox and colleagues (2008) applied a method that accounted for compositional heterogeneity when re-examining the phylogenetic relationships of the three domains based on the sequences of small- and large-subunit rRNAs as well as those of some 50 conserved proteins. Their phylogenetic analysis of concatenated protein sequences from 40 taxa placed eukaryotes as the sister group to the *Crenarchaeota*, favouring the eocyte hypothesis. And by allowing for base compositional heterogeneity, even rRNA sequences, which have long been the stronghold for the 3-domain tree, also supported the eocyte topology. Naturally, phylogenetic trees inferred from divergent sequences are only as good as their underlying models of evolution, and the present-day compositional bias within a genome is but one of many factors that have guided sequence evolution. However, support for the eocyte hypothesis forces us to re-think the evolution of features that uniquely link bacteria and eukaryotes as well as those confined to all contemporary *Archaea* (Cavalier-Smith, 2002).

As with the classification of *Reptiles* (which, as a group, comprises crocodiles, squamates and turtles but excludes birds) and of Great Apes (which is typically used to denote gorillas, chimps and orangutans to the exclusion of humans), the term '*Prokaryotes*' represents a paraphyletic grouping consisting of the *Bacteria* and the *Archaea*, the latter of which are certainly more closely related to, and possibly ancestral to, *Eukaryotes*. So perhaps we need to invoke a term that collectively groups the *Bacteria* and *Eukaryotes* – and might I suggest a radical name: the *AnArchaea*?

References

Baldauf, S.L., Palmer, J.D., and Doolittle, W.F. (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.
- Cavalier-Smith, T. (2002) The neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int J Syst Evol Microbiol* **52**: 7–76.
- Cox, C.J., Foster, P.G., Hirt, R.P., Harris, S.R., and Embley, T.M. (2008) The archaeobacterial origin of eukaryotes. *Proc Natl Acad Sci USA* **105**: 20356–20361.
- Embley, T.M., and Martin, W. (2006) Eukaryotic evolution, changes and challenges. *Nature* **440**: 623–630.
- Gogarten, J.P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E.J., Bowman, B.J., *et al.* (1989) Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proc Natl Acad Sci USA* **86**: 66661–66665.
- Gouy, M., and Li, W.-H. (1989) Phylogenetic analysis based on rRNA sequences supports the archaeobacterial rather than the eocyte tree. *Nature* **339**: 145–147.
- Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. (1989) Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc Natl Acad Sci USA* **86**: 9355–9359.
- Lake, J.A. (1988) Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* **331**: 184–186.
- Pisani, D., Cotton, J.A., and McInerney, J.O. (2007) Super-trees disentangle the chimerical origin of eukaryotic genomes. *Mol Biol Evol* **24**: 1752–1760.
- Poole, A.M., and Penny, D. (2007) Evaluating hypotheses for the origin of eukaryotes. *Bioessays* **29**: 74–84.
- Rivera, M.C., and Lake, J.A. (1992) Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257**: 74–76.
- Yutin, N., Makarova, K.S., Mekhedov, S.L., Wolf, Y.I., and Koonin, E.V. (2008) The deep archaeal roots of eukaryotes. *Mol Biol Evol* **25**: 1619–1630.