

Commentary

Of Archae and Eo: What's in a name?

Russell F. Doolittle

Center for Molecular Genetics, University of California at San Diego, La Jolla, CA 92093-0634

On November 3, 1977, the *Los Angeles Times* carried a front page story reporting that a “new” life form had been discovered. The report noted that the “new” form arose earlier than the two previously known forms and was most like the common ancestor of all life on earth. What the article was describing was the finding by Carl Woese and his co-workers that a group of organisms, previously thought of as bacteria, actually formed a distinctly separate group when considered on the basis of their ribosomal RNA sequences (1). Woese called these organisms the archaeobacteria (archae = primitive), in line with the notion that they resembled the most ancient form of life on earth. It was apparent, however, that although the phylogenetic trees constructed with ribosomal RNA clearly showed three distinct clusters—corresponding to the eubacteria, the archaeobacteria, and the eukaryotes—the resemblance of any of them to a more primitive ancestor could only be supposed and not “rooted” in any firm sense. In fact, the three groups appeared to be equidistant from each other (Fig. 1A).

Earlier that same year, Woese and Fox (2) had coined the term “progenote” to describe a primitive hypothetical ancestor for prokaryotes and eukaryotes. Because there was microfossil evidence that suggested bacteria have remained virtually unchanged, in a morphological sense, over the course of 3.5 billion years, the progenote must have been very ancient. It was described as a primitive cell with a rudimentary translation apparatus, the loosely assembled genome of which may have been wholly RNA-based (3). Unraveling the true evolutionary relationships of the three “urkingdoms” (ur = primary) and their earlier ancestral stock has proved vexingly difficult, however, and widely different scenarios have been postulated since the initial report (Fig. 1). Now in this issue of the *Proceedings*, James Brown and W. Ford Doolittle present findings that go a long way toward resolving the arguments (4). A review of some of the conflicting observations that has beset this field may help explain why their study is so important.

The concept of three widely divergent “urkingdoms” was sustained as more ribosomal RNA sequences were reported. Initially, the new kingdom had comprised

only methanogens, but as exploration continued, it was found to embrace halophilic and extremely thermophilic bacteria as well. As such, the group seemed to be composed of organisms whose lifestyles seemed in harmony with the presumed harsh conditions of an early earth.

The notion of three ancient lineages was further confirmed by electron microscopy of ribosomes, each of the three kingdoms having recognizably different features (5). At least one characteristic prominence was common to the archaeobacteria and eukaryotes but absent from eubacteria; nonetheless, these authors, sensitive to the fashionable rules of strict parsimony, demurred from concluding that eukaryotes evolved from archaeobacteria, cautiously proposing instead that all three lineages sprang from a hypothetical ancestor called the “paleocyte” (“ancient + cell”). At about the same time, comparisons of 5S RNA sequences also suggested that the archaeobacteria were more similar to eukaryotes, leading to the suggestion that the group would be better named “metabacteria” (6), a reckoning more in line with Fig. 1D.

Not long thereafter, the electron microscopy group claimed even greater morphological resolution, the new details revealing that one subgroup of archaeobacteria, hyperthermophiles like *Sulfolobus*, had ribosomes that were distinctive from

other archaeobacteria and significantly more similar to ribosomes from eukaryotes (7). Once again, adhering to the rules of strict parsimony, they concluded that what this group and eukaryotes had in common were “primitive” traits and that the two groups were “older” than, or at least as old as, other archaeobacteria and eubacteria. They suggested the name “eocyte” (“dawn + cell”) for the subgroup, which they felt should be elevated to kingdom status.

Subsequently, several newly determined protein sequences seemed to be at odds with a primordial archaeobacteria interpretation, the archaeobacterial sequences more often than not appearing more similar to eukaryotic homologues than to those from the eubacteria. The bulk of the evidence seemed to be shifting to a phylogeny like that described in Fig. 1D. The matter became increasingly disputatious, with arguments and counterarguments about the methodology of sequence comparison and phylogenetic tree construction often becoming heated. The usual problems of changing rates and arbitrary rooting were argued incessantly.

Then, in 1989, two groups appeared to offer insurmountable evidence that the archaeobacteria were really more akin to eukaryotes than eubacteria (Fig. 1D). Iwabe *et al.* (8) and, independently, Gogarten *et al.* (9) used the stratagem of

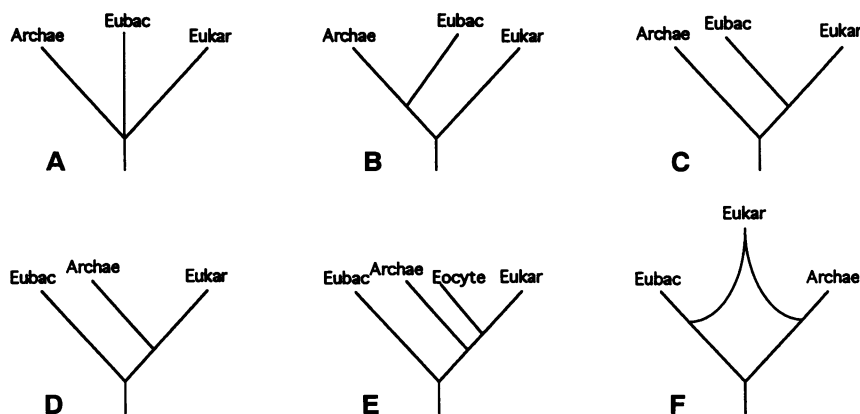


FIG. 1. Simple depictions of the “Tree of Life” showing various topologies and relationships that have been suggested. (A) “Trivergence” from a common ancestor (strictly speaking, trivergences never occur, of course). (B) Depiction showing eubacteria and archaeobacteria being more closely related than either is to eukaryotes. (C) Depiction with eubacteria being more closely related to eukaryotes. (D) Depiction showing that archaeobacteria and eukaryotes have had a common ancestor more recently than either has had with a eubacterium. (E) Depiction showing hyperthermophiles as a separate group (“eocytes”). (F) Depiction showing a eukaryote to be a chimera of a eubacterium and an archaeobacterium.

comparing duplicated gene products to root their trees. Thus, if all three groups of organisms have sets of proteins that arose by prior gene duplication, each of the two groups of duplicated sequences can be used to root the other's tree. When this was done for the elongation factors tu and g, on the one hand (8), and the catalytic and regulatory subunits of the F-1 and vacuolar ATPases (8, 9), on the other, all trees grouped the archaeobacteria with the eukaryotes. The implication was that the archaeobacteria, their name notwithstanding, are not necessarily ancient, which is to say, organisms like contemporary eubacteria may have existed before the advent of the archaeobacteria (Fig. 1D). Eventually, Woese *et al.* (10) came to agree that the most likely phylogeny resembled one shown in Fig. 1D, but nevertheless they clung to the prefixial "archae," suggesting the three urkingdoms be called Archaea, Bacteria, and Eukarya. The justification for keeping "archae" was based on the reasoning that they were indeed ancestral to eukaryotes.

Meanwhile, further studies of protein sequences were not turning out to be so clearcut. An analysis of glutamate dehydrogenase sequences suggested that the kinship problem could not be resolved—certain methods of tree construction favoring archaeobacteria being closer to eubacteria, others to eukaryotes (11). Indeed, in the case of glutamine synthetase, the archaeobacterial sequence is obviously more similar to its eubacterial counterparts than to eukaryotic ones, and the last-resort appeal to the horizontal gene transfer has had to be invoked (12, 13).

Furthermore, the ATPase story was becoming bogged down as additional homologues were found among the eubacteria, including a vacuolar-like one, casting doubt on whether or not the correct comparisons had been made in the initial studies (8, 9). Again, the possibility of a confounding horizontal gene transfer has been suggested (14). Even the elongation factor analysis was challenged on the basis that the sequences were not sufficiently long to allow statistically significant conclusions (15). Further complicating the results of that analysis was a report that the elongation factor 1 α from hyperthermophiles ("eocytes") has a skein of amino acids in common with the equivalent factor from eukaryotes but absent in elongation factors from other archaeobacteria and eubacteria (16). These data are most readily accommodated by a phylogenetic tree of the sort shown in Fig. 1E. Finally, it was found that heat shock proteins of the HSP70 class have a stretch of amino acids that is present only in heat shock proteins of Gram-negative bacteria and all eukaryotes but is absent in those of Gram-positive bacteria and a halophile (archaeobacterium) (17–19).

The study by Brown and Doolittle (4) set out to resolve some of these matters. Their chosen realm was the aminoacyl tRNA sequences, the 20 enzyme families of which are divided equally into two distinct evolutionary groups (20). Both groups are the results of a series of gene duplications that occurred well before the divergences leading to the three "urkingdoms" (21). Brown and Doolittle realized that, if sequences could be obtained from a reasonable number of representatives of all three "urkingdoms," the "rooting by paralogue" strategy could be used to resolve the arguments. They focused on the valine-isoleucine-leucine cluster of enzymes and began by using PCR to generate sequences for representative archaeobacteria (*Pyrococcus* and *Sulfolobus*) as well as a key early-diverging eukaryote (*Nosema locustae*) and some primitive eubacteria (*Aquifex* and *Thermotoga*). They also added to the valine aminoacyl tRNA synthetase inventory by determining the corresponding sequence from an amitochondrial protist (*Trichomonas vaginalis*). Next, they analyzed these sequences and a number of others already in the literature by a variety of phylogenetic tree construction methods. The trees are convincingly robust, and the result seems unassailable. The archaeobacterial sequences always cluster with the eukaryotes (Fig. 1D). Moreover, the clustering involves all three groups of archaeobacteria—halophiles, methanogens, and hyperthermophiles—in a monophyletic group.

But does this settle the matter? Not entirely. One stumbling block has to do with the process by which eukaryotic cells came into being. The orthodox view has been that the three major kingdoms evolved from each other incrementally, the last stage of eukaryotic evolution being the endosymbiotic acquisition of Gram-negative eubacteria destined to become mitochondria and plastids (22). The unorthodox position extends the concept of endosymbiotic acquisition to the nucleus and, directly or indirectly, to other architectural and functional features of the eukaryotic cell. Put more generally, this view regards the eukaryote as some kind of chimera of a eubacterium and an archaeobacterium.

No one disputes that archaeobacteria have some features in common with eubacteria and others that are more akin to the eukaryotes. For example, their circular genome and the nature of the transcriptional units on it are obviously similar in eubacteria and archaeobacteria. On the other hand, many other features found in archaeobacteria are distinctly eukaryotic (22–24). In this regard, it must be remembered that historically the archaeobacteria were denoted prokaryotic because, like other bacteria, they lack the nucleus that distinguishes the eukaryotic cell. As such, the genesis of the nucleus has often been

the focus of efforts to accommodate the seemingly irreconcilable observations. Some of these proposals trace back to suggestions made nearly a century ago to the effect that the nucleus is a relic of an endosymbiotic event (25). In fact, Woese and Fox (1) suggested that eukaryotic cells were the result of a progenote-like eukaryote engulfing endosymbionts destined to become organelles, with the host cell (which they deemed a "urkaryote") providing the RNA translational machinery. Lake *et al.* (5), in an effort to explain why archaeobacterial and eukaryotic ribosomes resemble each other morphologically, inverted that proposal, suggesting instead that an archaeobacterium was the "guest" (endosymbiont) and the contributor of ribosomes. Subsequently, Hartman (26) proposed that eukaryotic cells, with their uniquely configured cytoskeleton, were initially wholly RNA-based, either a eubacterium or an archaeobacterium being endosymbiosed to become the nucleus. Hartman called the engulfing cell a "kronocyte," after the figure in Greek mythology who swallowed its young. Sogin (27), anxious to reconcile discrepancies between rRNA and protein-based phylogenies, came down firmly on the side of the endosymbiont being an archaeobacterium. Zillig and co-workers (24, 28) and Gupta and Golding (17) downplayed the matter of the nucleus, but both presented arguments in favor of the eukaryotic cell being a fusion product of an archaeobacterium and a eubacterium (Fig. 1F).

The driving force behind these scenarios has gradually shifted in recent years. Initially, the prod was to reconcile the relationships predicated with rRNA sequences with protein sequence comparisons (26, 27). Now the major thrust is to provide an explanation for why some archaeobacterial sequences are like eukaryotic ones and others eubacterial.

There are certainly some valid reasons to take these suggestions seriously. Endosymbiosis *does* offer an attractive explanation of the double-membraned nature of the nucleus (29); but it doesn't explain at all why the guest ribosomes would displace those of the host cell. And certainly neither fusion nor endosymbiosis offers any insights about the origin of the eukaryotic cytoskeleton. It also seems unreasonable that the aminoacyl tRNA synthetase components of protein synthesis would come from a different root-stock than the RNA components with which they interact. In the end, the "chimera solutions" seem to introduce more problems than they solve.

In the meantime, the most vexing problems center around particular protein sequences and the matter of monophyly vs. polyphyly. In the case of heat shock protein HSP70, the molecular kinship of the Gram-positive bacteria with archaeobacteria (17, 19) would not surprise those who

long ago postulated that the archaeobacteria were derived from Gram-positive bacteria (22). But such an interpretation flies in the face of there being three distinct lineages. Similarly, the observation that elongation factor 1 α has a sequence feature in common with eukaryotic factors may indeed reflect the nature of the ancient archaeobacterium that gave rise to eukaryotes (16), but it is clearly at odds with the phylogenies of Brown and Doolittle (4) and also the expanded results of Miyata *et al.* (30), who first introduced the idea of using duplicated genes for rooting these trees (8), both sets of data clearly showing the archaeobacteria as a monophyletic group including hyperthermophiles. One encouraging note is that the aminoacyl tRNA synthetase data have hardly been exhausted; many more sets of sequences will soon be available to further test all these propositions. In the end, the true relationship of all these organisms will doubtless emerge.

1. Woese, C. R. & Fox, G. E. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5088–5090.
2. Woese, C. R. & Fox, G. E. (1977) *J. Mol. Evol.* **10**, 1–6.
3. Woese, C. R. (1983) in *Evolution from Molecules to Men*, ed. Bendall, D. S. (Cambridge Univ. Press, Cambridge, U.K.) Chap. 11, pp. 209–233.
4. Brown, J. R. & Doolittle, W. F. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 2441–2445.
5. Lake, J. A., Henderson, E., Clark, M. W. & Matheson, A. T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5948–5952.
6. Hori, H., Itoh, T. & Osawa, S. (1982) *Zentralbl. Bakteriol. Mikrobiol. Hyg. Abt. 1 Orig. C* **3**, 18–30.
7. Lake, J. A., Henderson, E., Oakes, M. & Clark, M. W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3786–3790.
8. Iwabe, N., Kuma, K.-I., Hasegawa, M., Osawa, S. & Miyata, T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9355–9359.
9. Gogarten, J. P., Kibak, H., Ditttrich, P., Taiz, L., Bowman, E. J., Bowman, B. J., Manolson, M. F., Poole, R. J., Date, T., Oshima, T., Konishi, J., Denda, K. & Yoshida, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6661–6665.
10. Woese, C. R., Kandler, O. & Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4576–4579.
11. Benachou-Lahfa, N., Forterre, P. & Labedan, B. (1993) *J. Mol. Evol.* **36**, 335–346.
12. Smith, M. W., Feng, D. F. & Doolittle, R. F. (1993) *Trends Biochem. Sci.* **17**, 489–493.
13. Brown, J. R., Masuchi, Y., Robb, F. T. & Doolittle, W. F. (1994) *J. Mol. Evol.* **38**, 566–576.
14. Hilario, E. & Gogarten, J. P. (1993) *BioSystems* **31**, 111–119.
15. Forterre, P., Benachou-Lahfa, N., Confalonieri, F., Duguet, M., Elie, C. & Labedan, B. (1993) *BioSystems* **28**, 15–32.
16. Rivera, M. C. & Lake, J. A. (1992) *Science* **257**, 74–76.
17. Gupta, R. S. & Golding, G. B. (1993) *J. Mol. Evol.* **37**, 573–582.
18. Boorstein, W. R., Ziegelhuffer, T. & Craig, E. A. (1994) *J. Mol. Evol.* **38**, 1–17.
19. Gupta, R. S., Aitken, K., Falah, M. & Singh, B. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 2895–2899.
20. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1989) *Nature (London)* **347**, 203–206.
21. Nagel, G. & Doolittle, R. F. (1990) *Proc. Natl. Acad. Sci. USA* **88**, 8121–8125.
22. Cavalier-Smith, T. (1987) *Ann. N.Y. Acad. Sci.* **503**, 17–54.
23. Cavalier-Smith, T. (1990) *Rev. Micropalaeontol.* **33**, 145–154.
24. Zillig, W. (1991) *Curr. Opin. Genet. Dev.* **1**, 544–551.
25. Mereschkowsky, C. (1905) *Biol. Zbl.* **25**, 595–604.
26. Hartman, H. (1984) *Speculations Sci. Technol.* **7**, 77–81.
27. Sogin, M. L. (1991) *Curr. Opin. Genet. Dev.* **1**, 457–463.
28. Pühler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.-P., Lottspeich, F., Garrett, R. A. & Zillig, W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4569–4573.
29. Lake, J. A. & Rivera, M. C. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 2880–2881.
30. Miyata, T., Iwabe, N., Kuma, K.-I., Kawanishi, Y.-I., Hasegawa, M., Mukohata, Y., Ihara, K. & Osawa, S. (1991) in *Evolution of Life*, eds. S. Osawa & T. Honjo (Springer, Tokyo), pp. 337–351.