

The eocyte hypothesis and the origin of eukaryotic cells

John M. Archibald¹

The Canadian Institute for Advanced Research, Program in Integrated Microbial Biodiversity, Department of Biochemistry and Molecular Biology, Dalhousie University, Sir Charles Tupper Medical Building, 5850 College Street, Halifax, NS, Canada B3H 1X5

In the June 1984 issue of PNAS, James Lake and colleagues (1) published a provocative article in which they proposed that eukaryotes (animals, fungi, plants, and protists) evolved from a specific group of thermophilic prokaryotes, the “eocyte” archaeobacteria (1). Few questions capture the imagination of biologists like the origin of eukaryotic (nucleus-containing) cells such as our own, and as additional support accumulated (e.g., refs. 2–4) Lake’s eocyte hypothesis garnered considerable attention. The idea that eukaryotes could have arisen from within an already diversified archaeobacterial lineage was eventually overshadowed by Woese’s (5) “three-domains” view of life in which archaeobacteria (including eocytes) represent a natural (i.e., monophyletic) group to the exclusion of eukaryotes and eubacteria (5). In this issue of PNAS, Cox *et al.* (6) revisit the possibility of an archaeobacterial origin for eukaryotes by using expanded molecular sequence datasets and ultra-modern phylogenetic approaches. Their analyses are a model of rigor and rekindle interesting and important ideas about the prokaryotic antecedents of eukaryotic cells.

Evolving Views on the Tree of Life

Next to life itself, the origin of complex cells is one of the most fundamental, and intractable, problems in evolutionary biology. Progress in this area relies heavily on an understanding of the relationships between present-day organisms, yet despite tremendous advances over the last half-century scientists remain firmly divided on how to best classify cellular life. Many adhere to the textbook concept of 2 basic types of cells, prokaryotes and eukaryotes, as championed by Stanier and van Niel (7). Others posit that at its deepest level life is not a dichotomy but a trichotomy comprised of cells belonging to the domains Bacteria, Archaea, and Eukarya, each monophyletic and sufficiently distinct from one another to warrant equal status (5, 8). The conceptual and practical challenges associated with establishing a genealogy-based classification scheme for microbes have been fiercely debated for decades (see ref. 9 for recent review), and the literature is rich in philosophy and rhetoric.

The genomics revolution of the 1990s brought tremendous optimism to the

field of microbial systematics: if enough genomes from diverse organisms could be sequenced and compared, definitive answers to questions about evolutionary relationships within and between eubacteria, archaeobacteria, and eukaryotes would surely emerge. More specifically, it should be possible to discern how eukaryotes evolved from prokaryotes (if indeed that is what happened), and perhaps even who among modern-day prokaryotic lineages is our closest ancestor. Unfortunately, with the sequences of hundreds of eubacterial, archaeobacterial, and eukaryotic genomes has come the realization that the number of universally distributed genes suitable for global phylogenetic analysis is frustratingly

Cox *et al.* revisit the possibility of an archaeobacterial origin for eukaryotes.

small (10). Lateral (or horizontal) gene transfer has shown itself to be a pervasive force in the evolution of both prokaryotic and eukaryotic genomes, and even if a “core” set of genes can be identified (and there is much debate on this issue), how confident are we that the phylogenetic signal in these genes reflects the vertical history of cells? How meaningful are sequence alignment-independent, gene content-based approaches to resolving the “tree of life” (11)? To what extent is a “net of life” a more accurate and useful metaphor for describing the full spectrum of life on Earth (10, 12–14)?

These vexing questions aside, it is important to consider what analyses of universally distributed genes have, and have not, revealed about the relationships between prokaryotes and eukaryotes. The advent of rRNA gene sequencing led to a revolution in microbiology with the discovery of the third domain of life, the archaeobacteria (15), but rRNA phylogenies do not in and of themselves unambiguously support the monophyly of the group. Some analyses in fact place the crenarchaeotes [Lake’s eocytes (1)] as the sister group of eukaryotes to the exclusion of the

other main archaeobacterial line, the euryarchaeotes (e.g., refs. 2 and 4). Similarly, although analyses of anciently diverged paralogs such as genes encoding translation elongation factors often [but not always (16)] place the root of the tree of life on the branch leading to eubacteria, they are equivocal as to whether archaeobacteria are monophyletic or whether crenarchaeotes are adjacent to eukaryotes (e.g., refs. 17 and 18). A striking 11-aa insertion in the elongation factor 1 α (EF-1 α) protein identified by Rivera and Lake (3) supports the eocyte tree, as does the phylogeny of EF-1 α itself (e.g., refs. 3 and 17).

New Support for an Old Idea

Cox *et al.* (6) perform a critical examination of the evidence for and against the eocyte hypothesis and the 3-domains tree: they revisit the phylogeny of large subunit (LSU) and small subunit (SSU) rRNA and carry out a comprehensive analysis of a set of 51 genes/proteins that includes components of the core replication, transcription, and translation machinery, the so-called “informational” genes (19). Using sophisticated methodologies that accommodate among-site compositional heterogeneity in DNA and protein sequences (20) and lineage-specific compositional changes over time (21), Cox *et al.* show that a combined 40-taxon LSU–SSU rRNA dataset produces topologies consistent with the eocyte tree, not the 3-domains tree. Thirty-nine of the 51 core proteins analyzed were found to be heterogeneous in amino acid composition, and remarkably, only a single protein (the largest subunit of RNA polymerase I) generated a statistically robust topology in which archaeobacteria were monophyletic, i.e., crenarchaeotes and euryarchaeotes were each other’s closest relatives. The remaining trees provided no strong evidence for or against either hypothesis, although of the 35 trees in which eukaryotes emerge from within a paraphyletic archaeobacteria, 8 show the eocyte topology (6). Consistent with the

Author contributions: J.M.A. wrote the paper.

The author declares no conflict of interest.

See companion article on page 20356.

¹E-mail: jmarchib@dal.ca.

© 2008 by The National Academy of Sciences of the USA

results of Yutin *et al.* (22), it would appear that individual proteins analyzed in isolation do not paint a consistent picture as to whether the 3-domains tree should be favored over the eocyte tree or vice versa.

In search of increased resolution, Cox *et al.* (6) performed a battery of additional phylogenetic analyses on a concatenated set of 45 proteins culled from the original 51-protein dataset by eliminating alignments containing multiple eukaryotic paralogs. Using this supermatrix of 5,521 amino acid sites, traditional methods such as maximum parsimony resulted in the 3-domains topology, but a strongly-supported eocyte tree was obtained when maximum-likelihood and Bayesian analyses were performed, even under a “composition homogeneous” model. Use of a reduced amino acid alphabet to minimize the impact of saturation and/or compositional heterogeneity resulted in a robust crenarchaeote–eukaryote relationship. In sum, their analyses “. . . provide support for the eocyte tree, rather than the 3-domains tree” (6).

What are we to make of this conclusion? The gulf between prokaryotes and eukaryotes is obviously enormous, and

the thorn in the side of the eocyte hypothesis has always been how to rationalize crenarchaeote–eukaryote monophyly with the apparent unity of archaeobacteria from the perspective of molecular and cell biology. One of the most commonly raised objections is the fact that archaeobacteria possess glycerol-ether membrane lipids, unlike the glycerol-ester lipids found in most eubacteria and eukaryotes. If, as the eocyte hypothesis predicts, the eukaryotic nucleocytoplasm evolved from within archaeobacteria, eukaryotes would have had to replace their ancestral lipid biosynthesis pathway with a eubacterial-type system. At first glance this might seem problematic but it is not so difficult to imagine given that the “operational” genes of eukaryotes are primarily eubacterial in origin, not archaeobacterial (23).

Conversely, comparative genomic studies have revealed that most “informational” genes in archaeobacteria (e.g., those involved in DNA replication, transcription, and translation) are related to eukaryotic homologs (23, 24). Although it is often assumed that archaeobacteria are monophyletic, there is in fact still much to learn about the biology of cren-

archaeotes relative to euryarchaeotes. Indeed, the recent discovery of a novel cell division machinery in diverse crenarchaeal genera, unrelated to the FtsZ-based system found in euryarchaeotes and with proteins homologous to components of the eukaryote-specific endosomal protein sorting apparatus (25, 26), raises the possibility that a variety of molecular phylogeny-independent characters will eventually be brought to bear on the validity of the eocyte hypothesis.

The extent to which a small fraction of the genomes of living organisms can be used to trace the history of cellular lineages dating back >1 billion years will no doubt continue to be debated for years to come. Regardless, an important message to be taken from the analysis of Cox *et al.* (6) is just how sensitive the results of molecular phylogenies involving anciently diverged sequences can be to model misspecification and compositional heterogeneity. Beyond serving as a general impetus for reexamination of any number of tough phylogenetic problems, the results of Cox *et al.* serve to put the eocyte hypothesis firmly back “on the radar” of those struggling to understand the earliest events in the diversification of cellular lineages.

- Lake JA, Henderson E, Oakes M, Clark MW (1984) Eocytes: A new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc Natl Acad Sci USA* 81:3786–3790.
- Lake JA (1988) Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* 331:184–186.
- Rivera MC, Lake JA (1992) Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* 257:74–76.
- Tourasse NJ, Gouy M (1999) Accounting for evolutionary rate variation among sequence sites consistently changes universal phylogenies deduced from rRNA and protein-coding genes. *Mol Phylogenet Evol* 13:159–168.
- Woese CR, Kandler O, Wheelis ML (1990) Toward a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 87:4576–4579.
- Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM (2008) The archaeobacterial origin of eukaryotes. *Proc Natl Acad Sci USA* 105:20356–20361.
- Stanier RA, van Niel CB (1962) The concept of a bacterium. *Arch Microbiol* 42:17–35.
- Pace NR (2006) Time for a change. *Nature* 441:289.
- Sapp J (2006) Two faces of the prokaryote concept. *Int Microbiol* 9:163–172.
- Dagan T, Martin W (2006) The tree of one percent. *Genome Biol* 7:118.
- Wolf YI, Rogozin IB, Grishin NV, Koonin EV (2002) Genome trees and the tree of life. *Trends Genet* 18:472–479.
- Doolittle WF (1999) Phylogenetic classification and the universal tree. *Science* 284:2124–2129.
- Kurland CG, Canback B, Berg OG (2003) Horizontal gene transfer: A critical view. *Proc Natl Acad Sci USA* 100:9658–9662.
- McInerney JO, Cotton JA, Pisani D (2008) The prokaryotic tree of life: past, present... and future? *Trends Ecol Evol* 23:276–281.
- Woese CR (1987) Bacterial evolution. *Microbiol Rev* 51:221–271.
- Philippe H, Forterre P (1999) The rooting of the universal tree of life is not reliable. *J Mol Evol* 49:509–523.
- 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.
- 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.
- Jain R, Rivera MC, Lake JA (1999) Horizontal gene transfer among genomes: The complexity hypothesis. *Proc Natl Acad Sci USA* 96:3801–3806.
- Lartillot N, Philippe H (2004) A Bayesian mixture model for across-site heterogeneities in the amino acid replacement process. *Mol Biol Evol* 21:1095–1109.
- Foster PG (2004) Modeling compositional heterogeneity. *Syst Biol* 53:485–495.
- Yutin N, Makarova KS, Mekhedov SL, Wolf YI, Koonin EV (2008) The deep archaeal roots of eukaryotes. *Mol Biol Evol* 25:1619–1630.
- Esser C, *et al.* (2004) A genome phylogeny for mitochondria among α -proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol Biol Evol* 21:1643–1660.
- Rivera MC, Jain R, Moore JE, Lake JA (1998) Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci USA* 95:6239–6244.
- Lindas A-C, Karlsson EA, Lindgren MT, Ettema TJG, Bernander R (2008) A unique cell division machinery in the Archaea. *Proc Natl Acad Sci USA* 105:18942–18946.
- Samson RY, Obita T, Freund SM, Williams RL, Bell SD (2008) A role for the ESCRT system in cell division in Archaea. *Science*, 10.1126/science.1165322.