nearly always, it is the catalytic mechanism or cofactor-binding properties that are conserved or slightly modified and the substrate specificity that is changed (*25*). This suggests that it is much easier to evolve new binding sites than new catalytic mechanisms. Most of the members of these families are distributed across different pathways (*25–27*). There are only a tiny number of cases in which domains conserve their substrate binding properties and occur in the same pathway. An inspection of where homologs are found in the network of pathways shows that recruitment primarily occurred on the basis of catalytic mechanism or cofactor binding. This has led to a mosaic pattern of protein families with little or no coherence in the evolutionary relationships in different parts of the network.

To what extent are pathways conserved over a range of different organisms? The same pathway in different organisms can contain species-specific sets of isozymes (*28*, *29*). The comparison of enzymes in the same pathway in different organisms also shows that proteins responsible for the particular functions can belong to unrelated protein families. This phenomenon is called "nonorthologous displacement" (*30*). Variations come not just from changes in specific enzymes. In some organisms, sections of the standard pathway are not found and the gaps are bypassed through the use of alternative pathways (*28*). Together, these variations produce widespread plasticity in the pathways that are found in different organisms; much of this is described in the Clusters of Orthologous Groups (COGs) database (*4*, *31*).

For other sets of pathways, we expect duplications of the type described here, though possibly with more duplications within pathways that have arisen late in evolution, such as those of signal transduction and the immune system.

Causes and Consequences

The earliest evolution of the protein repertoire must have involved the ab initio invention of new proteins. At a very low level, this may still take place. But it is clear that the dominant mechanisms for expansion of the protein repertoire, in biology as we now know it, are gene duplication, divergence, and recombination. Why have these mechanisms replaced ab initio invention? Two plausible causes, which complement each other, can be put forward. First, once a set of domains whose functions are varied enough to support a basic form of life had been created, it was much faster to produce new proteins with different functions by duplication, divergence, and recombination. Second, once the error-correction procedures now present in DNA replication and protein synthesis were developed, they made the ab initio invention of proteins a process that is too difficult to be useful.

Consequently, even the simplest genomes of extant bacteria are the product of extensive gene duplication and recombination (*3*, *15*). An organism's complexity is not directly related to its number of genes; flies have fewer genes than nematodes, and humans have fewer than rice. However, complexity does seem to be related to expansions in particular families that underlie the more complex forms of life. This means that we will be able to trace much of the evolution of complexity by examining the duplications and recombinations of these families in different genomes (*32*, *33*).

References and Notes

1. M. F. Perutz, J. C. Kendrew, H. C. Watson, *J. Mol. Biol.* **13**, 669 (1965).
2. M. Lynch, J. S. Conery, *Science* **290**, 1151 (2000).
3. J. Gough *et al.*, *J. Mol. Biol.* **313**, 903 (2001). Data used here includes updated results that can be found at http://supfam.org.
4. E. V. Koonin, M. Y. Galperin, *Sequence-Evolution-Function* (Kluwer Academic, Boston, MA, 2003).
5. A. G. Murzin *et al.*, *J. Mol. Biol.* **247**, 536 (1995). Names of domain families are taken from the SCOP database at http://scop.mrc-lmb.cam.ac.uk/scop/.
6. Y. I. Wolf *et al.*, *Genome Res.* **9**, 17 (1999).
7. A. Muller *et al.*, *Genome Res.* **12**, 1625 (2002).
8. V. A. Kuznetsov, *J. Biol. Syst.* **10**, 381 (2002).
9. M. A. Huynen, E. van Nimwegen, *Mol. Biol. Evol.* **15**, 583 (1998).
10. J. Qian *et al.*, *J. Mol. Biol.* **313**, 673 (2001).
11. E. V. Koonin, Y. I. Wolf, G. P. Karev, *Nature* **420**, 218 (2002).
12. C. A. Wilson, J. Kreychman, M. Gerstein, *J. Mol. Biol.* **297**, 233 (2000).
13. A. E. Todd, C. A. Orengo, J. M. Thornton, *J. Mol. Biol.* **307**, 1113 (2001).
14. M. G. Rossmann *et al.*, *Nature* **259**, 194 (1974).
15. S. A. Teichmann, J. Park, C. Chothia, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14658 (1998).
16. E. V. Koonin, L. Aravind, A. S. Kondrashov, *Cell* **101**, 573 (2000).
17. G. Apic, W. Huber, S. A. Teichmann, *J. Struct. Funct. Genomics*, in press.
18. G. Apic, J. Gough, S. A. Teichmann, *J. Mol. Biol.* **310**, 311 (2001).
19. S. Wuchty, *Mol. Biol. Evol.* **18**, 1694 (2001).
20. M. Bashton, C. Chothia, *J. Mol. Biol.* **315**, 927 (2002).
21. C. Vogel, C. Berzuini, S. A. Teichmann, unpublished data.
22. N. H. Horowitz in *Evolving Genes and Proteins*, V. Bryson, H. J. Vogel, Eds. (Academic Press, New York, 1965), pp. 15–23.
23. R. A. Jensen, *Annu. Rev. Microbiol.* **30**, 409 (1976).
24. M. Riley, M. H. Serres, *Annu. Rev. Microbiol.* **54**, 341 (2000).
25. S. A. Teichmann *et al.*, *J. Mol. Biol.* **311**, 693 (2001).
26. S. C. G. Rison, S. A. Teichmann, J. M. Thornton, *J. Mol. Biol.* **318**, 911 (2002).
27. R. Alves, R. A. Chaleil, M. J. Sternberg, *J. Mol. Biol.* **320**, 751 (2002).
28. T. Dandekar *et al.*, *Biochem. J.* **343**, 115 (1999).
29. O. Jardine *et al.*, *Genome Res.* **12**, 916 (2002).
30. E. V. Koonin, A. R. Mushegian, P. Bork, *Trends Genet.* **12**, 334 (1996).
31. R. L. Tatusov, E. V. Koonin, D. J. Lipman, *Science* **278**, 631 (1997).
32. S. A. Chervitz *et al.*, *Science* **282**, 2022 (1998).
33. C. Vogel, S. A. Teichmann, C. Chothia, unpublished data.

34. We thank M. Madera, E. Koonin, and A. Finkelstein for comments on the manuscript. J.G. has a Burroughs-Welcome Fellowship from the Program in Mathematics and Molecular Biology, and C.V. has a Boehringer Ingelheim Predoctoral Fellowship.

VIEWPOINT

The Deep Roots of Eukaryotes

S. L. Baldauf

Most cultivated and characterized eukaryotes can be confidently assigned to one of eight major groups. After a few false starts, we are beginning to resolve relationships among these major groups as well. However, recent developments are radically revising this picture again, particularly (i) the discovery of the likely antiquity and taxonomic diversity of ultrasmall eukaryotes, and (ii) a fundamental rethinking of the position of the root. Together these data suggest major gaps in our understanding simply of what eukaryotes are or, when it comes to the tree, even which end is up.

Introduction

Molecular phylogenetic trees have gradually assigned most of the cultivated and characterized eukaryotes to one of eight major groups (Fig. 1). Although these data have largely failed to resolve relationships among these major groups, with the benefit of hindsight it was perhaps somewhat naïve that we ever thought they would. While similarities among gene sequences may indicate the relatedness of the organisms that harbor them, this relationship is far from straightforward, particularly for ancient "deep" branches. Only a fraction of sites in any gene are free to mutate, and these have only so many states (nucleotides or amino acids) to toggle through before they start repeating themselves, and their true history becomes obscured.

With more data, improved methods, and just a better idea of what we're doing, an outline of the tree seems to be emerging. This

Department of Biology, University of York, Box 373, Heslington, York YO10 5YW, UK. E-mail:slb14@york.ac.uk

comes largely from (i) analyses of concatenated (combined) multigene data sets [e.g., (*1*, *2*)], (ii) recognition of the common patterns in all data (consensus) (*3*, *4*), and (iii) a better understanding of phylogenetic artifacts, particularly the tendency of highly divergent sequences to "attract" each other in phylogenetic trees ("long branch attraction") (*5*, *6*). Nonetheless, progress has been slow. Large data sets are still hard going for eukaryotes; there are little or no data on most taxa, including whole major groups, and genome sequences are few and far between, taking large collaborative efforts and years to complete. Two recent developments could dramatically change this. These are the discovery of a taxonomically diverse world of extremely small eukaryotes (*7*–*11*) and a possible radical rerooting of the entire tree (*12*).

Eukaryotic Diversity: To (Nearly) Every Rule There Is an Exception

Eukaryotes (the domain of life to which we belong) are a highly distinct but also highly variable group: e.g., histones are dispensable (*13*); the genetic code varies (*14*, *15*); and multiple, sometimes highly differentiated nuclei have evolved multiple times (*16*–*18*), as has multicellularity (animals, fungi, plants, amoebozoans, heterokonts) (*1*–*4*, *17*, *18*). All extant eukaryotes appear to postdate the advent of mitochondria, but these organelles have been lost repeatedly or degenerated into small residual organelles of unknown function (parabasalids, entamoebids, fungi) (*19*, *20*). True chloroplasts were probably only "invented" once, unquestionably from cyanobacteria, but photosynthesis has spread widely by lateral transfer (secondary endosymbiosis), where one eukaryote engulfed another and kept its plastid (chlorarachniophytes, euglenoids, cryptophytes, chromophytes, dinoflagellates, apicomplexans) (*21*). There is one known tertiary plastid symbiosis (haptophytes) (*22*), and two cases where a stripped-down version of the original host nucleus is retained together with the plastid in the new host—a semiautonomous cell within a cell (chlorarachniophytes, cryptophytes) (*21*).

What We Thought We Knew But Didn't

Recent progress on resolving the relationships among these major groups has, however, consisted largely of backtracking. Early molecular trees gave an appealing picture of eukaryote evolution with a set of relatively simple-celled taxa inhabiting the deepest branches (*23*). These "Archezoa" lack mitochondria, golgi stacks, and meiosis, suggesting that early stages in the evolution of eukaryotic cellular complexity had survived. Meanwhile, almost everything else clustered together in a poorly resolved "crown radiation," suggesting that most of the remaining major groups arose in rapid succession. Both ideas have died slow, lingering deaths. We now know that the deep branching of at least some, if not all, of the "Archezoa" is an artifact of their highly divergent sequences (long branches) being "attracted" to a distant outgroup (see below) (*6*, *24*). All of these taxa also now appear to have once had mitochondria because mitochondrially derived genes occur in their nuclei (*19*). Meanwhile, the crown "radiation" probably simply reflects the inability of individual genes to resolve deep branches owing to a lack of sufficient clean phylogenetic information. We believe this because these deep branches are now being resolved by larger data sets consisting of many genes strung together (*1*, *2*).

The True Diversity of Life

This was all based on the assumption that we were working with a full deck, or at least as full as we could get, i.e., that we were sampling the vast majority of extant eukaryotes. It is now clear that this was not the case. New data now indicate a huge potential diversity of extremely small eukaryotes including new major subgroups scattered across the tree (Fig. 1) (*7*–*11*). These are nano- (2 to 20 μm) and pico- (<2 μm) eukaryotes, eukaryotes that overlap bacteria (~0.5 to 2 μm) in size. Most of these "taxa" were first detected by culture-independent surveys (ciPCR), a polymerase chain reaction (PCR)–based method of acquiring sequences from pooled environmental DNAs, bypassing the need to identify, much less isolate, new organisms (*25*). ciPCR studies have more than doubled the estimated number of prokaryotic phyla (*25*). However, eukaryotes were thought to be adequately surveyed by microscopy, and such global surveys showed the same relatively small set of taxa in the same habitats worldwide (*26*).

The existence of bacterial-sized eukaryotes has been known for some time. The smallest described eukaryote, *Ostreococcus tauri*, is <1 μm in diameter but still has a nucleus, 14 linear chromosomes, one chloroplast, and several mitochondria (*27*). Nonetheless, the potential diversity of small eukaryotes was only first recognized as a by-product of bacterial ciPCR. Surveys targeting eukaryotes were first reported only last year (*7*–*11*), but these already indicate a tremendous potential diversity of previously unknown ultrasmall taxa. These are scattered across the eukaryote tree and include major new subgroups (*11*), almost a parallel universe to the one we knew. To put this in perspective, there are ~40 major subgroups of previously known cultivated eukaryotes (Fig. 1), and ciPCR surveys so far indicate at least 10 substantial new subgroups and a possible 20 to 30 more.

Data from these new "taxa" are already changing our ideas about eukaryotic evolution, diversity, and ecology, and even our basic idea of what a eukaryote is. The presence of deeply branching, ultrasmall taxa throughout the tree means that miniaturization is not a recent event and size was not necessarily an important factor in the early evolution of eukaryotes. In terms of deep branches, new taxa not only give us a more complete picture of who the major players are but also help clarify the relationships among them. (i) New taxa may bridge gaps between groups; this breaks up their long branches and decreases associated artifacts. (ii) Conservative members of currently known but highly derived groups may be identified and used in their place in evolutionary trees; this eliminates long branches altogether. Thus, phylogenetic analyses with new short-branched "pico-radiolarian" sequences recently demonstrated the monophyly of acantharians + polycystinids, a question dating back to Haeckel (*28*). A particularly enticing possibility is the discovery of taxa bridging the gap between eukaryotes and Archaea.

There are important caveats. Most of these new "taxa" are currently only PCR-amplified, small-subunit ribosomal RNA (SSU) sequences (phylotypes), defined as unique purely on the basis of their positions in SSU trees. However, we know that rapidly evolving sequences also appear as unique in SSU trees, e.g., Microsporidia group with amito-excavates in SSU trees, but are now recognized as rapidly evolving close relatives of fungi (*24*). Amoebozoans form at least three distinct groups with SSU sequences, but a single coherent group with nearly all other molecular data (*1*, *2*). Therefore, the identities of new "taxa" need to be confirmed by isolating the organisms involved, which will probably be easier than has been the case for prokaryotes [e.g., (*29*)]. Nonetheless, although some of these phylotypes are almost certainly only fast-evolving SSU sequences, many are clearly not, and some already constitute fairly substantial groups (*10*, *11*).

The Root of All Roots

The most important point in a phylogenetic tree is its root. The root is the oldest point in the tree and corresponds to the theoretical last common ancestor (LCA) of everything in the tree. The root gives directionality to evolution within the tree (relative order of branching events). It also identifies which groups are "true" groups. If the root of the tree lies within a "group," then it is not a group but a "grade." Resemblances among taxa at a grade are not unique defining features but rather primitive characteristics retained from their LCA, which is also the LCA of everything else in the tree.

"Rooting" a tree requires an external point of reference or "outgroup." For eukaryotes this

is either Archaea (nuclear housekeeping-gene trees) or Bacteria (mitochondrial gene trees). The antiquity of these relationships makes them almost by definition the mother of all long branches. Thus, although most rooted molecular trees place the amito-excavates nearest the root [e.g., (*2*, *3*, *30*, *31*); but see (*24*)], these taxa, also largely obligate parasites or symbionts, also tend to have very long branches. This makes their deep position look suspiciously like a long-branch attraction to the long branches of the distant outgroup (*6*).

An alternative approach to quantitatively calculating trees is to use macromolecular characters such as gene fusions, genomic rearrangements, or large insertions and deletions in conservative genes. These can be powerful phylogenetic markers because they are rare, complex, and largely irreversible and therefore unlikely to arise independently and be shared by unrelated taxa. A particularly dramatic example is the fusion of dihydrofolate reductase and thymidylate synthase (*11*), recently identified in nearly all eukaryotes except opisthokonts (animals, fungi, and their allies). Because gene fusions are rare, particularly among eukaryotes that lack operons, these genes probably fused only once, meaning that most eukaryotes share a unique common ancestor exclusive of opisthokonts.

Thus, these data place the root of the eukaryote tree between opisthokonts and nearly all the other major eukaryote taxa. Essentially, it turns the tree on its head, rooting it within the former "crown radiation." This is a radical reinterpretation and would mean that opisthokonts branched off very early from the main line of eukaryote descent. The LCA of all extant eukaryotes would then have been a far more complex organism than previously envisioned, and any similarities between, e.g., animals and plants would simply be universal eukaryote traits. It also suggests that opisthokonts may be older than previously thought, consistent with the diversity of single-celled protists now thought to be closely allied to animals and/or fungi (*32*).

There are many caveats. While compelling, this gene fusion is still only a single character and unsupported by any robust molecular trees, most of which still place the root close to or within amito-excavates. It is particularly disconcerting that these genes are missing altogether from amoebozoans and amito- excavates, which occupy pivotal positions in the two competing scenarios (Fig. 1). The antiquity of this event (1 to 2 billion years) allows alternative explanations such as reversal of the gene fission in an opisthokont ancestor (*33*) or replacement of the fused genes by lateral gene transfer from bacteria (*34*), where these genes are adjacent in an operon.

Future Prospects

The discovery, much less the characterization, of ultrasmall eukaryotes is barely in its infancy. Few habitats have been reported on, and so

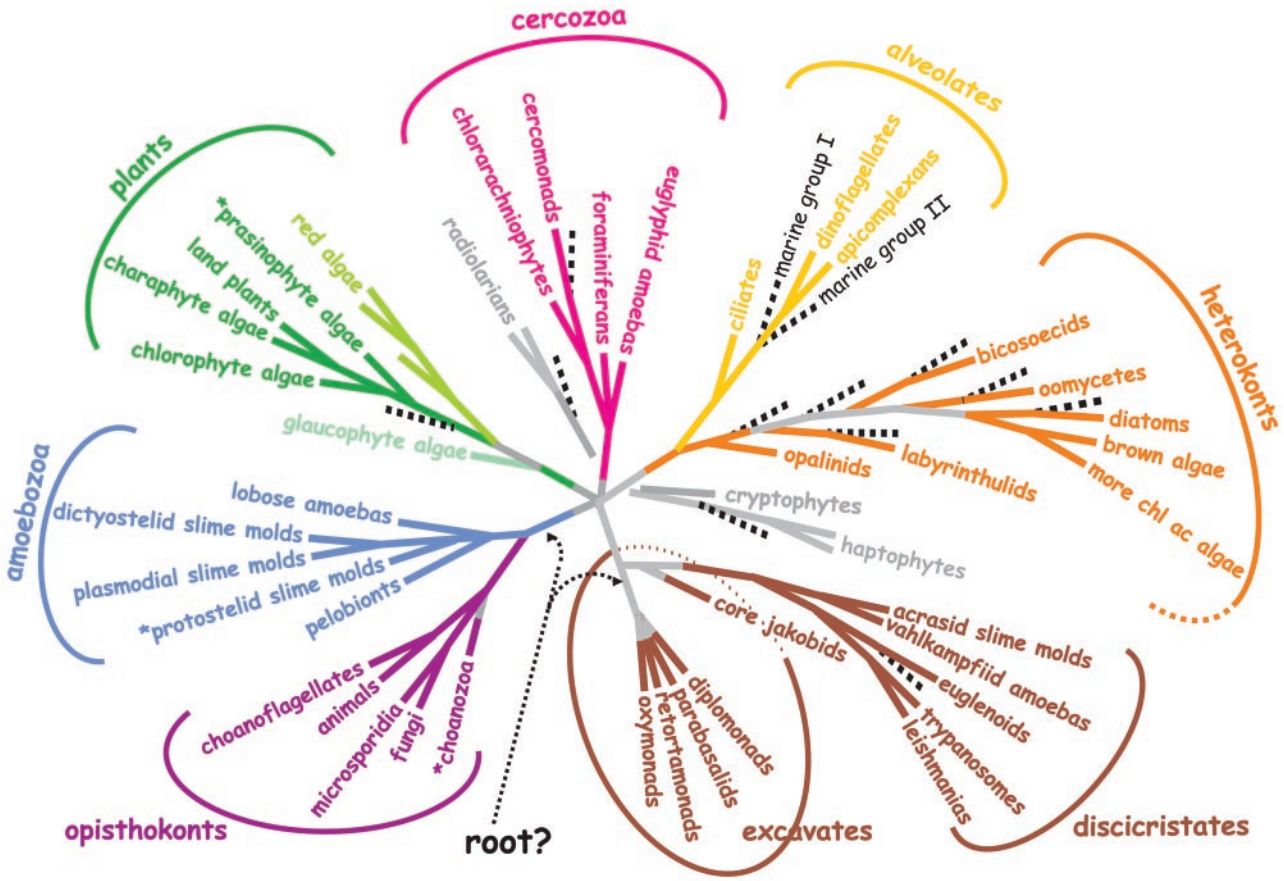

Fig. 1. A consensus phylogeny of eukaryotes. The vast majority of characterized eukaryotes, with the notable exception of major subgroups of amoebae, can now be assigned to one of eight major groups. Opisthokonts (basal flagellum) have a single basal flagellum on reproductive cells and flat mitochondrial cristae (most eukaryotes have tubular ones). Eukaryotic photosynthesis originated in Plants; theirs are the only plastids with just two outer membranes. Heterokonts (different flagellae) have a unique flagellum decorated with hollow tripartite hairs (stramenopiles) and, usually, a second plain one. Cercozoans are amoebae with filose pseudopodia, often living within tests (hard outer shells), some very elaborate (foraminiferans). Amoebozoa are mostly naked amoebae (lacking tests), often with lobose pseudopodia for at least part of their life cycle. Alveolates have systems of cortical alveoli directly beneath their plasma membranes. Discicristates have discoid mitochondrial cristae and, in some cases, a deep (excavated) ventral feeding groove. Amitochondrial excavates lack substantial molecular phylogenetic support, but most have an excavated ventral feeding groove, and all lack mitochondria. The tree shown is based on a consensus of molecular (*1*–*4*) and ultrastructural (*16*, *17*) data and includes a rough indication of new ciPCR "taxa" (broken black lines) (*7*–*11*). An asterisk preceding the taxon name indicates probable paraphyletic group.

many more subgroups, if not major groups, undoubtedly remain to be discovered. Simply a better understanding of the taxa already indicated should greatly facilitate resolution of the deep branches of the eukaryote tree or even define them clearly for the first time. An exciting possibility is the prospect of pico-eukaryote genomics. Owing to their size and complexity, there are few completely sequenced eukaryote genomes, mostly from opisthokonts. However, pico-eukaryotes also probably have simplified "pico" genomes; *Ostreococcus tauri*'s genome is ~8 Mb in size, less than twice that of the laboratory strain of *Escherichia coli* (*35*). Thus, these genomes should be highly amenable to sequencing, and we could relatively quickly accumulate a taxonomically broad enough set of eukaryote genomes to start making meaningful global comparisons.

Our understanding of eukaryote evolution, in terms of taxonomic diversity, genome structure, and ecology, is similar to that for prokaryotes 10 to 15 years ago. Genomics and ciPCR have together revolutionized nearly every aspect of our understanding of bacteria and archaea. It is fantastic to consider the probability that we are on the cusp of a similar revolution for eukaryotes.

References

1. S. L. Baldauf, A. J. Roger, I. Wenk-Siefert, W. F. Doolittle, *Science* **290**, 972 (2000).
2. E. Bapteste *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1414 (2002).
3. T. M. Embley, R. P. Hirt, *Curr. Opin. Genet. Dev.* **8**, 624 (1998).
4. A. G. Simpson, A. J. Roger, *Curr. Biol.* **12**, R691 (2002).
5. J. Felsenstein, *Syst. Zool.* **27**, 401 (1978).
6. S. Gribaldo, H. Philippe, *Theor. Popul. Biol.* **61**, 391 (2002).
7. S. Y. Moon-van der Staay, R. De Wachter, D. Vaulot, *Nature* **409**, 607 (2001).
8. P. Lopez-Garcia, F. Rodriguez-Valera, C. Pedros-Alio, D. Moreira, *Nature* **409**, 603 (2001).
9. L. A. Amaral Zettler *et al.*, *Nature* **417**, 137 (2002).
10. S. C. Dawson, N. R. Pace, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8324 (2002).
11. D. Moreira, P. Lopez-Garcia, *Trends Microbiol.* **10**, 31 (2002).
12. A. Stechmann, T. Cavalier-Smith, *Science* **297**, 89 (2002).
13. P. J. Rizzo, E. R. Cox, *Science* **198**, 1258 (1977).
14. L. A. Klobutcher, P. J. Farabaugh, *Cell* **111**, 763 (2002).
15. P. J. Keeling, W. F. Doolittle, *EMBO J.* **15**, 2285 (1996).
16. D. M. Prescott, *Nature Rev. Genet.* **1**, 191 (2000).
17. J. J. Lee, G. F. Leedale, P. Gradbury, Eds., *The Illustrated Guide to the Protozoa* (Society of Protozoologists, Lawrence, KS, ed. 2, 2000).
18. K. Hausmann, N. Hulsmann, *Protozoology* (Thieme, New York, 1996).
19. S. D. Dyall, P. J. Johnson, *Curr. Opin. Microbiol.* **3**, 404 (2000).
20. B. A. Williams, R. P. Hirt, J. M. Lucocq, T. M. Embley, *Nature* **418**, 865 (2002).
21. J. M. Archibald, P. J. Keeling, *Trends Genet.* **18**, 577 (2002).
22. H. S. Yoon, J. D. Hackett, D. Bhattacharya, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11724 (2002).
23. M. L. Sogin, G. Hinkle, D. D. Leipe, *Nature* **362**, 795 (1993).
24. R. P. Hirt *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 580 (1999).
25. E. F. DeLong, N. R. Pace, *Syst. Biol.* **50**, 470 (2001).
26. D. J. Patterson, W. J. Lee, in *The Flagellates*, J. Green, B. S. C. Leadbeater, Eds. (Taylor and Francis, London, 2000), pp. 269–287.
27. C. Courties *et al.*, *J. Phycol.* **34**, 844 (1998).
28. P. Lopez-Garcia, F. Rodriguez-Valera, D. Moreira, *Mol. Biol. Evol.* **19**, 118 (2002).
29. R. Massana, L. Guillou, B. Diez, C. Pedros-Alio, *Appl. Environ. Microbiol.* **68**, 4554 (2002).
30. E. Hilario, J. P. Gogarten, *J. Mol. Evol.* **46**, 703 (1998).
31. S. L. Baldauf, J. D. Palmer, W. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7749 (1996).
32. L. A. Amaral-Zettler, T. A. Nerad, C. J. O'Kelly, M. L. Sogin, *J. Eukaryot. Microbiol.* **48**, 293 (2001).
33. I. Yanai, Y. I. Wolf, E. V. Koonin, *Genome Biol.* **3**, 0024 (2002).
34. J. O. Andersson *et al.*, *Curr. Biol.* **13**, 94 (2003).
35. E. Derelle *et al.*, *J. Phycol.* **38**, 1150 (2002).

VIEWPOINT

Phylogenomics: Intersection of Evolution and Genomics

Jonathan A. Eisen* and Claire M. Fraser

Much has been gained from genomic and evolutionary studies of species. Combining the perspectives of these different approaches suggests that an integrated phylogenomic approach will be beneficial.

Although it is generally accepted that genome sequences are excellent tools for studying evolution, it is perhaps less well accepted that evolutionary analysis is a powerful tool in studies of genome sequences. In particular, evolutionary analysis helps to place comparative genomic studies in perspective. Researchers can begin to understand how and even why some of the similarities and differences in genomes came to be, for example, the presence and absence of genes, the DNA substitution patterns seen in noncoding regions, and global patterns of synteny (conserved gene order) across species. These analyses, in turn, can be used to understand metabolism, pathogenicity, physiology, and behavior. An important component of such studies is the fact that certain evolutionary analyses are only possible with (or are greatly improved by) analysis of complete genome sequences. Gene loss cannot be unequivocally inferred for a species if one does not have the complete genome. The converse is also true—certain genomic studies are greatly improved by using evolutionary analysis. The feedback loops between genome analysis and evolutionary studies are so pervasive that we believe it is necessary to integrate the two approaches into a single composite, called phylogenomics (*1*, *2*).

In building the tree of life, analysis of whole genomes has begun to supplement, and in some cases to improve upon, studies previously done with one or a few genes. For example, recent studies of complete bacterial genomes have suggested that the hyperthermophilic species are not deeply branching; if this is true, it casts doubt on the idea that the first forms of life were thermophiles (*3*). Analysis of the genome of the eukaryotic parasite *Encephalitozoon cuniculi* supports suggestions that the group Microsporidia are not deep branching protists but are in fact members of the fungal kingdom (*4*). Genome analysis can even help resolve relationships within species, such as by providing new genetic markers for population genetics studies in the bacteria causing anthrax or tuberculosis (*5*, *6*). In all these studies, it is the additional data provided by a complete genome sequence that allows one to separate the phylogenetic signal from the noise. This is not to say the tree of life is now resolved—we only have sampled a smattering of genomes, and many groups are not yet touched (*7*).

Just as genomics can help resolve the branching patterns in the tree of life, an accurate picture of the tree is critical for genome studies. An accurate tree allows one to select species so as to best represent phylogenetic diversity or to select organisms that are optimally positioned for answering particular questions. For example, *Drosophila pseudoobscura* was selected in large part for genome sequencing because it is at an evolutionary distance from *D. melanogaster*, such that potential regulatory regions will be somewhat conserved and can be identified by

The Institute for Genomic Research, Rockville, MD 20850, USA.

*To whom correspondence should be addressed. E-mail: jeisen@tigr.org