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

methods like phylogenetic footprinting (*8*).

In studies of the vast diversity of microbes that have never been cultured in the laboratory, DNA can be isolated directly from the environment. To know what organisms that DNA came from, one needs a phylogenetic anchor sequence [e.g., ribosomal RNA (rRNA)] to link the DNA to the tree of life. This approach led to the identification of a novel form of phototrophy in the open ocean (*9*) and has revolutionized environmental microbiology.

In addition, the tree of life can be used to select species that "bracket" major evolutionary transitions. For example, comparing genomes on either side of the prokaryote- eukaryote transition has identified features conserved among eukaryotes but apparently absent from prokaryotes (*10*). The importance of this approach is one reason to sequence the genomes of monotremes (the earliest branching group in mammalian evolution that still has living species, including the echidna and the duck-billed platypus). A detailed knowledge not only of the structure of the tree but also of the origin of particular features is important for such studies, so as to avoid artificially grouping together features that evolved independently multiple times such as multicellularity (*11*).

The concept of a single tree implies that evolution of species follows a branching pattern in which genes and genomes are transmitted vertically from parents to offspring. However, this is not the only mode of evolution—genes can also be passed from one evolutionary branch to another in a process known as lateral gene transfer (LGT). Although the occurrence of LGT was known for many years, analysis of whole genomes has reinvigorated its investigation (*12*). The value of complete genome sequences is great in these studies since it allows one to screen for genes that may have unusual origins, without any preconceived notions of what those genes might be. This has allowed, for example, the identification of "pathogenicity islands" in bacteria (*13*–*15*). Analysis of complete genomes has contributed to better understanding of one of the most common forms of LGT—that of genes from organellar genomes (mitochondria and chloroplasts, which used to be free-living bacteria and still have their own genomes) to the nuclear genomes of eukaryotes. Analysis of complete genomes has been used to identify genes that are likely derived from chloroplasts in plant (*16*) and *Plasmodium* genomes (*17*).

Unfortunately, many of the claims of LGT turn out to be incorrect [e.g., the claim that the human genome was "infected" with hundreds of bacterial genes (*18*) has since been refuted (*19*–*21*)]. In practice, identifying cases of LGT is quite difficult, and distinguishing it from other phenomena that cause genes to look anomalous, such as unusual rates of evolution, strong selection, or gene loss, has been difficult (*22*). When done carefully, whole-genome studies suggest that LGT has been rare over the course of evolution and that it has not completely distorted the structure of the tree. Therefore, we should view claims of LGT with appropriate skepticism and ask whether alternative possibilities have been tested.

The value of evolutionary analysis in genomics goes beyond simply using the relationships among species. For example, the prediction of gene function is greatly improved by phylogenetic analysis of gene families (*23*, *24*). Evolutionary reconstructions have revealed that in bacteria the most common major rearrangements are inversions that are symmetric about the origin of replication (*25*, *26*), that the entire genome of the plant *Arabidopsis thaliana* has apparently been duplicated (*27*), and that many segmented duplications may have occurred in human history (*28*). Evolutionary analysis also allows the determination of the age of duplication events, which in turn can greatly aid in functional studies (i.e., recent duplications suggest the expansion of an activity in a species, old duplications likely reflect divergent functions) (*29*, *30*). In all such cases, evolutionary analysis helped by allowing homologous genes in different species to be divided into groups of orthologs and paralogs, which in turn allows the indentification of duplication events.

In conclusion, there is an ever-growing list of examples in which cross-talk between these two disciplines has enabled scientists to design better experiments and generate new insights. Just as development biology has embraced evolution and become known as EvoDevo, genomics and evolution should become one.

References

1. J. A. Eisen, P. C. Hanawalt, *Mutat. Res.* **435**, 171 (1999).
2. J. A. Eisen, D. Kaiser, R. M. Myers, *Nature Med.* **3**, 1076 (1997).
3. V. Daubin, M. Gouy, G. Perrière, *Genome Res.* **12**, 1080 (2002).
4. M. D. Katinka *et al.*, *Nature* **414**, 450 (2001).
5. T. D. Read *et al.*, *Science* **296**, 2028 (2002).
6. R. D. Fleischmann *et al.*, *J. Bacteriol.* **184**, 5479 (2002).
7. P. Hugenholtz, *Genome Biol.* **3**, reviews0003.1 (2002).
8. D. L. Gumucio *et al.*, *Mol. Cell. Biol.* **12**, 4919 (1992).
9. O. Beja *et al.*, *Science* **289**, 1902 (2000).
10. V. Wood *et al.*, *Nature* **415**, 871 (2002).
11. J. A. Eisen, *Nature* **415**, 845 (2002).
12. W. F. Doolittle, *Trends Cell Biol.* **9**, M5 (1999).
13. J. Parkhill *et al.*, *Nature* **413**, 523 (2001).
14. H. Tettelin *et al.*, *Science* **293**, 498 (2001).
15. H. Tettelin *et al.*, *Science* **287**, 1809 (2000).
16. T. Rujan, W. Martin, *Trends Genet.* **17**, 113 (2001).
17. M. J. Gardner *et al.*, *Nature* **419**, 498 (2002).
18. E. S. Lander *et al.*, *Nature* **409**, 860 (2001).
19. S. L. Salzberg, J. A. Eisen, *Science* **293**, 1048 (2001).
20. M. J. Stanhope *et al.*, *Nature* **411**, 940 (2001).
21. J. Roelofs, P. J. Van Haastert, *Nature* **411**, 1013 (2001).
22. J. A. Eisen, *Curr. Opin. Genet. Dev.* **10**, 606 (2000).
23. J. A. Eisen, *Nucleic Acids Res.* **26**, 4291 (1998).
24. J. A. Eisen, *Genome Res.* **8**, 163 (1998).
25. J. A. Eisen, J. F. Heidelberg, O. White, S. L. Salzberg, *Genome Biol.* **1**, research0011.1 (2000).
26. M. Suyama, P. Bork, *Trends Genet.* **17**, 10 (2001).
27. The *Arabidopsis* Genome Initiative, *Nature* **408**, 796 (2000).
28. J. A. Bailey *et al.*, *Science* **297**, 1003 (2002).
29. I. K. Jordan, K. S. Makarova, J. L. Spouge, Y. I. Wolf, E. V. Koonin, *Genome Res.* **11**, 555 (2001).
30. J. F. Heidelberg *et al.*, *Nature* **406**, 477 (2000).

VIEWPOINT

Preserving the Tree of Life

Georgina M. Mace,[1] John L. Gittleman,[2] Andy Purvis[3]

Phylogenies provide new ways to measure biodiversity, to assess conservation priorities, and to quantify the evolutionary history in any set of species. Methodological problems and a lack of knowledge about most species have so far hampered their use. In the future, as techniques improve and more data become accessible, we will have an expanded set of conservation options, including ways to prioritize outcomes from evolutionary and ecological processes.

If a species is at risk of extinction, its close relatives have a higher than average chance of being at risk too (*1*). As well as predicting extinction risk, phylogeny provides a powerful metaphor for biodiversity—the Tree of Life. If the lengths of all the branches are summed, phylogeny can go beyond metaphor to yield a natural measure of biodiversity ("evolutionary history" or "phylogenetic diversity" (*2, 3*) (Fig. 1).

Counting Extinctions

Documentation of the extinction crisis has tended to be about lists of species—those regarded as extinct, as committed to extinction, or as threatened with extinction. The

[1]Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK. [2]Department of Biology, Gilmer Hall, University of Virginia, Charlottesville, VA 22904, USA. [3]Department of Biological Sciences, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, UK.