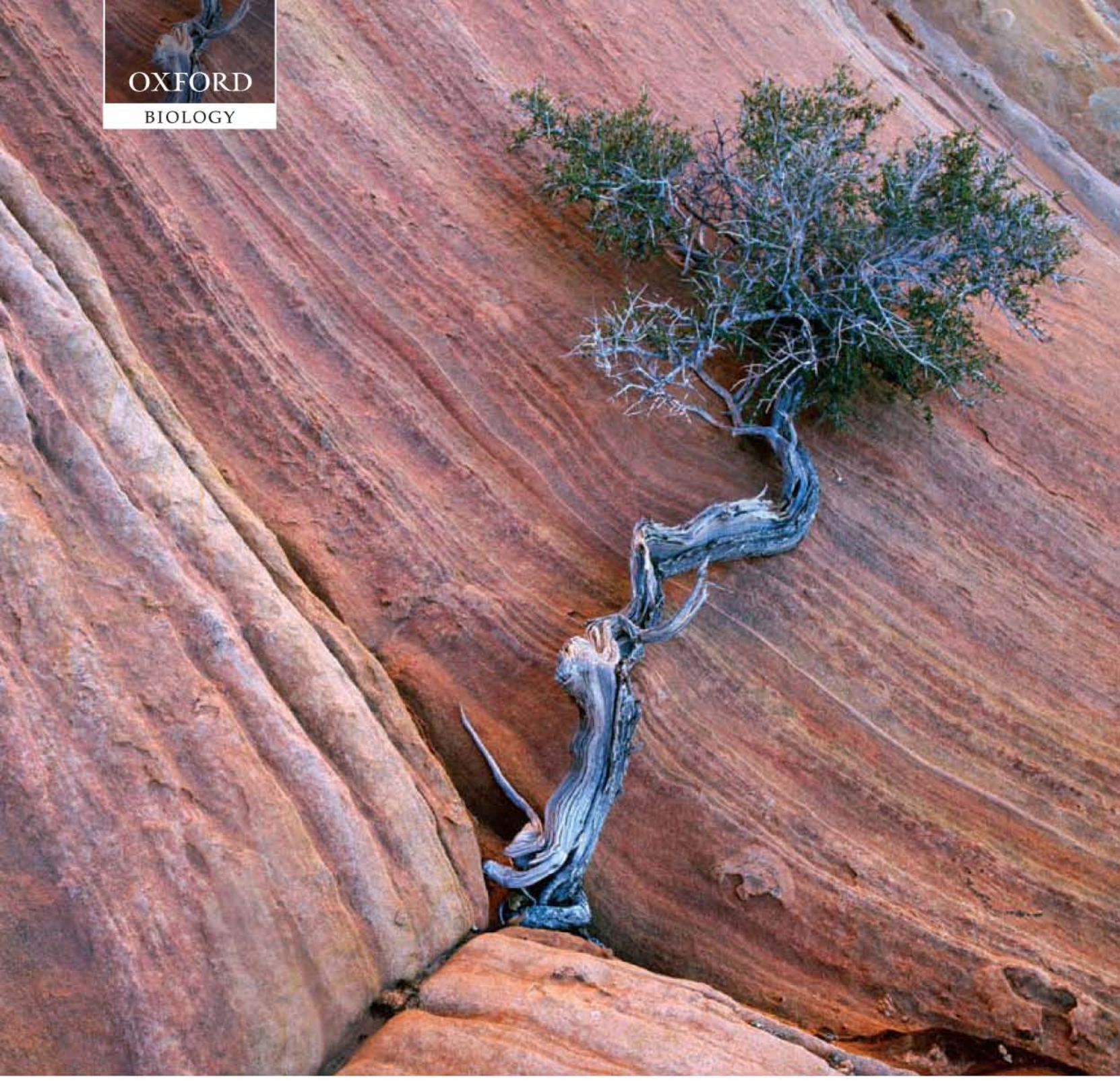

OXFORD
BIOLOGY

the TIMETREE *of* LIFE

edited by **S. BLAIR HEDGES** *and* **SUDHIR KUMAR**

foreword by James D. Watson

Eukaryotes (Eukaryota)

Debashish Bhattacharya[a,*], *Hwan Su Yoon*[b], *S. Blair Hedges*[c], *and Jeremiah D. Hackett*[d]

[a]Department of Biological Sciences and Roy J. Carver Center for Comparative Genomics, 446 Biology Building, University of Iowa, Iowa City, IA 52242, USA; [b]Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, ME 04575, USA; [c]Department of Biology, 208 Mueller Laboratory, Pennsylvania State University, University Park, PA 16802-5301, USA; [d]University of Arizona, Ecology and Evolutionary Biology Department, Biosciences West 336, Tucson, AZ 85721, USA

*To whom correspondence should be addressed (debashi-bhattacharya@uiowa.edu).

Abstract

Complex multicellular eukaryotes such as plants, animals, and fungi evolved independently from unicellular ancestors (protists). The relationships of these and other major groups of eukaryotes have proven difficult to determine because of a sparse fossil record and lack of a consensus among molecular phylogenies. Nonetheless, time estimates based on multiple nuclear genes indicate that eukaryotic photosynthetic organelles (plastids) arose from a symbiotic event with a cyanobacterium ~1600–1400 million years ago (Ma). The timetree suggests that most major groups of living eukaryotes also arose during the Mesoproterozoic (1600–1000 Ma) in an evolutionary radiation probably associated with new niches created by eukaryotic algae.

Resolving the tree of life for eukaryotes is an important challenge for biologists. This challenge has largely been taken on by molecular evolutionists because of the increasing availability and ability of genome data to resolve ancient relationships, combined with the lack of an extensive fossil record from this period in eukaryotic history. The earliest eukaryote fossil that can be assigned to a living lineage is the sexual red algal fossil *Bangiomorpha,* which is dated at 1200 Ma (*1, 2*), and only a handful of other taxonomically resolved eukaryotes are known between then and the Ediacaran Period (~635 Ma) (*3–5*). All multicellular clades trace their roots to protist—eukaryotes that are not plants, animals, or fungi (*6*)—ancestors, therefore solving the basal splits in the tree with regard to protists (Fig. 1) as well providing a timeline for these events are of paramount importance.

The first attempts at generating a pan-eukaryotic tree of life relied on comparisons of rDNA genes and although these were highly useful, it turned out that a single-gene framework could not resolve all ancient protist relationships. In addition, like most single-gene markers, rDNA trees produced some controversial results, for example a tree with deeply branching lineages below and with a cluster of recently radiated lineages above (*7, 8*). This tree shape was partially explained by artifactual "long-branch" attraction of some (but not all) highly diverged protist lineages. The next step was to focus on multigene datasets of conserved well-studied proteins (e.g., *9–12*) to increase the phylogenetic power. However, taxon representation in many of these analyses was sparse and marker choice was limited to a handful of proteins (e.g., actin, tubulins). In addition, attaining data from divergent phyla using degenerate PCR approaches proved to be very costly and time-consuming. This led to the current trend to generate eukaryotic trees using genome-wide (i.e., complete genome or expressed sequence tag, EST) data sets and large (e.g., >100) multiprotein alignments.

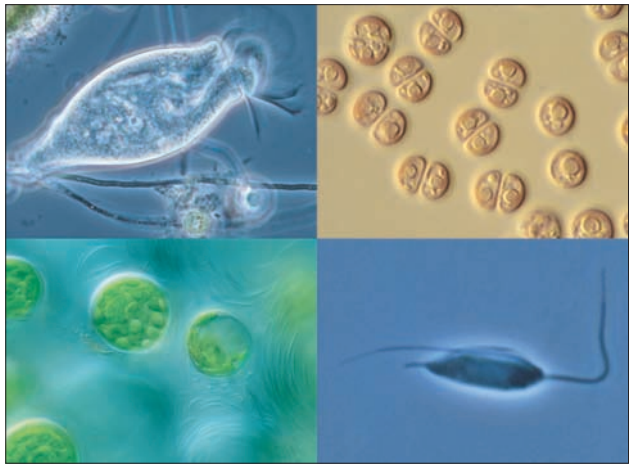

Fig. 1 *Epistylis*, a ciliate (upper left), *Rhodosorus*, a rhodophyte (upper right), *Leptosiropsis torulosa*, a chlorophyte (lower left), and *Jakoba ibera*, a jacobid excavate (lower right). Credits: D. J. Patterson and Aimlee Ladermann (upper left); B. Anderson and D. J. Patterson (upper right and lower left); and J. Cole and D. J. Patterson (lower right). Images provided by micro*scope (http://microscope.mbl.edu) under a creative commons license.

D. Bhattacharya, H. S. Yoon, S. B. Hedges, and J. D. Hackett. Eukaryotes (Eukaryota). Pp. 116–120 in *The Timetree of Life*, S. B. Hedges and S. Kumar, Eds. (Oxford University Press, 2009).

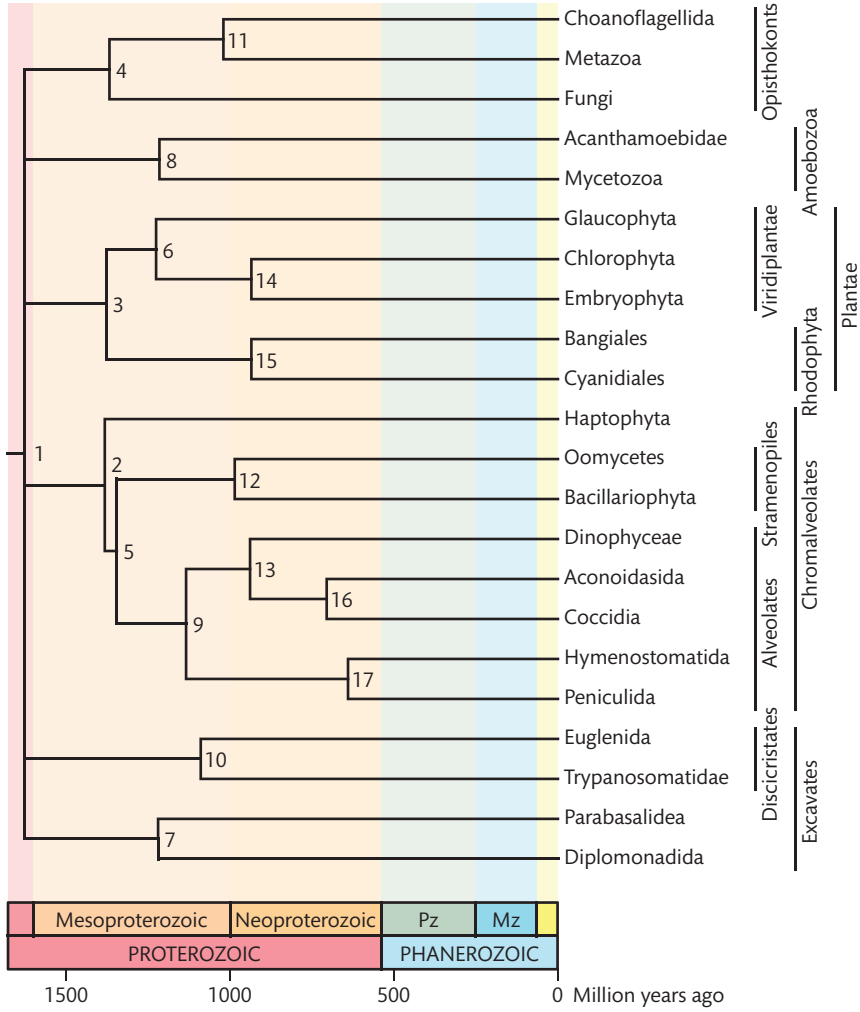

Fig. 2 A timetree of eukaryotes. Divergence times are shown in Table 1. *Abbreviations*: Mz (Mesozoic) and Pz (Paleozoic).

This phylogenomic approach has resulted in some notable successes with respect to protists (*13–19*), although these studies have been hampered by the quality of data (i.e., partial, single-pass EST reads), significant missing data, and sparse taxon sampling. In spite of these issues, the tree has begun to take shape and formed the basis for a classification scheme that defines six putative "supergroups" of eukaryotes. These are the Opisthokonta (e.g., animals, fungi, choanoflagellates), Amoebozoa (e.g., lobose amoebae, slime molds), Archaeplastida or hereafter Plantae (red algae, green algae and land plants, and glaucophyte algae), Chromalveolata (e.g., apicomplexans, ciliates, giant kelps), Rhizaria (e.g., cercomonads, foraminifera), and Excavata (e.g., diplomonads, parabasalids). Although the validity of some supergroups (e.g., Chromalveolata, Excavata) is clearly in question (e.g., *20, 21*), the supergroup concept is increasingly used (e.g., *22, 23*) in the scientific literature and has permeated the field.

Here we use primarily a recently published maximum likelihood (PhyML) phylogenetic hypothesis based on a 17-protein alignment generated by Hackett *et al.* (*24*) and a molecular clock method (Bayesian inference) that relaxes the requirement for a strict molecular clock (*25*) to estimate the dates of key nodes in the eukaryotic tree. Using this approach, the ML tree topology was first used to calculate the branch lengths with the program estbranches, using the JTT protein evolution model (*26*), before Bayesian estimation of divergence times using the program multidivtime (*27*). The ML tree was rooted

Table 1. Divergence times (Ma) and their confidence/credibility intervals (CI) among eukaryotes.

Timetree		Estimates					
Node	Time	Ref. (*12*)		Ref. (*24*)(a)		Ref. (*24*)(b)	
		Time	CI	Time	CI	Time	CI
1	1594	1827	–	1335	–	1620	–
2	1382	–	–	–	–	–	–
3	1379	1428	1579–1277	1250	1490–1060	1460	1740–1220
4	1368	1513	1642–1384	1370	1660–1150	1220	1440–1050
5	1345	–	–	1170	1400–980	1520	1830–1260
6	1225	–	–	1140	1350–970	1310	1560–1100
7	1220	–	–	820	1020–650	1620	2030–1310
8	1215	–	–	1250	1520–1040	1180	1400–990
9	1135	–	–	990	1200–810	1280	1570–1040
10	1090	–	–	740	930–570	1440	1790–1160
11	1020	–	–	1070	1270–910	970	1120–840
12	985	–	–	860	1060–680	1110	1370–890
13	940	–	–	820	1010–640	1060	1320–830
14	936	968	1150–786	870	1010–770	970	1130–840
15	935	–	–	870	1060–710	1000	1220–810
16	705	–	–	610	780–460	800	1020–590
17	640	–	–	560	730–410	720	940–520

Note: Node times in the timetree represent the mean of time estimates from different studies and methods. For Node 2, an average of times from three studies was used (*30–32*) (see text). For ref. (*24*), the two columns represent alternative rootings: (a) opisthokont and (b) diplomonads + parabasalids. For Node 1, see text for method; these times are averages of multiple time estimates, each with confidence and credibility intervals as presented in the original references.

on either the branch leading to the opisthokonts (*28*) or the diplomonads + parabasalids branch due to uncertainty about the placement of the root (*29*). We placed eight time constraints on this analysis based on the fossil record (*24*). To accommodate the variation in split times due to the use of two different rooting schemes, we calculate the mean divergence time for the two alternate roots. When available, the relevant dates from Hedges *et al.* (*12*) were included in averages (Table 1).

The phylogenetic position of haptophytes has been controversial, although several recent multigene studies have agreed in their placement as the closest relative of Chromalveolates + Rhizaria rather than nested within that group (*15, 16, 21*). Although haptophytes were not included in the two molecular clock studies used here (Table 1), their divergence from other major groups of eukaryotes has been timed in three other studies (*30–32*), resulting in a wide range of estimates (1900–1047 Ma) and an average of 1382 Ma. Therefore, we tentatively place them in the timetree (Fig. 2) in that position.

For estimating the root time (~1600 Ma) of the timetree (Fig. 2), we averaged the times of origin of the six included lineages in the polytomy from the two studies, Hackett *et al.* (*24*) (1335 and 1620 Ma, depending on the root) and Hedges *et al.* (*12*) (1827 Ma). Clearly, better knowledge of the phylogenetic position of the root will improve the time estimate of this node; the current estimate should be considered tentative. In this timetree, the split of fungi and animals was at ~1368 Ma, the split between stramenopiles and alveolates was at ~1345 Ma, and the origin of the photosynthetic organelle (plastid) in the Plantae ancestor was between ~1600 and ~1400 Ma (i.e., on the branch leading from the root to this supergroup). This time range for the primary origin of plastids (*33*) agrees with the conclusions of earlier molecular clock studies focused on the origin of plastids (*30*) as well as the separate studies used in constructing the current timetree (*12, 24*). Moreover, it corresponds to earliest undisputed eukaryotes (algae) in the fossil record; records from

before ~1600 Ma are debated as possibly being colonial prokaryotes (*1, 34–36*).

The striking pattern evident in the timetree (Fig. 2) is that nearly all of the divergences occurred in the Mesoproterozoic and earliest Neoproterozoic (~1600–900 Ma), in a relatively rapid evolutionary radiation. The likely explanation is the origin of plastids, thus creating eukaryotic algae, an increase in productivity, and an increase in ecological niches allowing diversification (*40*).

Are there other pan-eukaryotic molecular clock analyses that conflict with the results described earlier? The most comprehensive work in this regard is the multiprotein analysis by Douzery *et al.* (*37*), who suggested that the initial split among living eukaryotes was only 1085 Ma (1259–950 Ma), the split between animals and fungi 984 Ma (1127–872 Ma), and the important split between red algae and Viridiplantae 928 Ma (1061–825 Ma), with other dates equally young compared with previous estimates discussed earlier. Reanalyses of the Douzery *et al.* (*37*) data set were made by Roger and Hug (*38*) and Hug and Roger (*39*), who questioned the results and found that they were sensitive to the calibrations used. The specific calibrations and calibration methods used by Douzery *et al.* were also questioned elsewhere (*40*). In the study of Douzery *et al.* (*37*), each minimum calibration constraint was fixed as the younger boundary of the major geologic period containing the pertinent fossil rather than to the actual (older) geologic time constraints of the fossil itself, thus causing underestimates of resulting times. Douzery *et al.* also fixed maximum calibration constraints, arbitrarily, to the older boundary of the major geologic period containing the fossil rather than to an evolutionary event that might bear on the constraint. For example, the maximum calibration for the split of actinopterygian fish from mammals, 417 Ma, was essentially the same time as the oldest fossil on either branch, 416 Ma (*41*). However, there is little fossil information from this time period (Silurian) to establish that the divergence occurred precisely when the fossils appeared; more than likely it was much earlier, which would result in older Bayesian posterior time estimates. In addition, one of the maximum calibrations, the split between chelicerates and other arthropods (543 Ma), was fixed within the Cambrian, which is problematic because there is not an extensive fossil record before that period, showing morphological transitions, which could provide support for the use of a maximum calibration.

A separate reanalysis of the Douzery *et al.* (*37*) data set (S. B. Hedges, unpublished data) was conducted using the same methods and calibration taxa as the original authors but with corrections made to minimum calibrations only, based on the fossil record. This resulted in older time estimates of 1134 Ma (*Dictyostelium* root) and 1265 Ma (kinetoplastid root) for the initial split among living eukaryotes. However, when the problematic Cambrian arthropod maximum calibration was removed, and the actinopterygian–sarcopterygian maximum calibration was adjusted from 417 Ma (earliest fossil) to 495 Ma (more realistic), the initial eukaryote split was much older: 1857 Ma (*Dictyostelium* root) and 2216 Ma (kinetoplastid root). Therefore, this reanalysis of Douzery *et al.* (*37*) and those by Roger and Hug (*38*) and Hug and Roger (*39*) all agree that it was not the data set and relaxed clock method of Douzery *et al.* that resulted in relatively young times but rather the calibrations used. After correcting for calibration errors, significantly older times are produced, concordant now with the results of other studies.

Another analysis of note is that of Berney and Pawlowski (*42*), who used a broadly sampled data set of 240 small subunit rRNA genes. These authors were able to incorporate many more fossil dates (four maximum and 22 minimum constraints) in their analyses than genome-wide analyses due to the larger number of taxa in their tree that have a fossil record (e.g., coccolithophorids and diatoms). The increase in taxa and fossil constraints may have come with a loss of phylogenetic power due to the use of a single-gene framework. Other studies have documented the difficulties in inferring eukaryote-wide trees using single genes that sometimes show extreme rate variation and poor topological resolution. Moreover, according to de Vargas *et al.* (*31*), the resulting time estimates of Berney and Pawlowski (*42*) are underestimates attributed to the use of incorrect fossil dates for their clock calibration. The use of correct fossil calibrations would result in older time estimates (*31*).

Discovery that the relatively young time estimates for divergences among eukaryotes, found by Douzery *et al.* (*37*) and Berney and Pawlowski (*42*), are the result of miscalibrations goes a long way toward reconciling the differences between those studies and others bearing on the timescale of eukaryote evolution (e.g., *12, 24, 30*). In addition, the conflict regarding the time estimate (928 Ma) for the divergence of red and green algae in Douzery *et al.* (*37*), 300 million years younger than the earliest fossil at ~1200 Ma (*1*), is apparently resolved. Other time estimates for this divergence (Table 1, ~1380) are compatible with the fossil record. It also supports an earlier rise in organismal complexity, as measured in cell types

(*12*). In the future, studies that include a broader sample of taxa will permit more and better-constrained fossil calibrations that will presumably result in more reliable time estimates for the early evolutionary history of eukaryotes.

Acknowledgments

Support was provided by U.S. National Science Foundation (NSF) and National Aeronautics and Space Administration (NASA) to D.B. by NSF and the NASA Astrobiology Institute to S.B.H. and by the U.S. National Institute of Health through training grant support to J.D.H.

References

1. N. J. Butterfield, *Paleobiology* **26**, 386 (2000).
2. N. J. Butterfield, *Precambrian Res.* **111**, 235 (2001).
3. N. J. Butterfield, *Paleobiology* **30**, 231 (2004).
4. A. H. Knoll, *Life on a Young Planet: The First Three Billion Years of Evolution on Earth* (Princeton University Press, Princeton, NJ, 2003).
5. S. M. Porter, *Paleontol. Soc. Pap.* **10**, 35 (2004).
6. D. J. Patterson, *Am. Nat.* **154**, S96 (1999).
7. J. C. Edman *et al.*, *Nature* **334**, 519 (1988).
8. M. L. Sogin, J. D. Silberman, *Int. J. Parasitol.* **28**, 11 (1998).
9. S. L. Baldauf, A. J. Roger, I. Wenk-Siefert, W. F. Doolittle, *Science* **290**, 972 (2000).
10. E. Bapteste, H. Philippe, *Mol. Biol. Evol.* **19**, 972 (2002).
11. H. S. Yoon, J. D. Hackett, G. Pinto, D. Bhattacharya, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15507 (2002).
12. S. B. Hedges, J. E. Blair, M. L. Venturi, J. L. Shoe, *BMC Evol. Biol.* **4**, 2 (2004).
13. N. Rodriguez-Ezpeleta *et al.*, *Curr. Biol.* **15**, 1325 (2005).
14. J. E. Blair, P. Shah, S. B. Hedges, *BMC Bioinf.* **6**, 53 (2005).
15. J. D. Hackett *et al.*, *Mol. Biol. Evol.* **24**, 1702 (2007).
16. F. Burki *et al.*, *PLoS ONE* **2**, e270 (2007).
17. N. J. Patron, Y. Inagaki, P. J. Keeling, *Curr. Biol.* **17**, 887 (2007).
18. N. Rodriguez-Ezpeleta *et al.*, *Syst. Biol.* **56**, 389 (2007).
19. N. Rodriguez-Ezpeleta *et al.*, *Curr. Biol.* **17**, 1420 (2007).
20. L. W. Parfrey *et al.*, *PLoS Genet.* **2**, 2062 (2006).
21. H. S. Yoon *et al.*, *BMC Evol. Biol.* **8**, 14 (2008).
22. S. L. Baldauf, *Science* **300**, 1703 (2003).
23. P. J. Keeling *et al.*, *Trends. Ecol. Evol.* **20**, 670 (2005).
24. J. D. Hackett, H. S. Yoon, N. J. Butterfield, M. J. Sanderson, D. Bhattacharya, in *Evolution of Primary Producers in the Sea*, P. G. Falkowski, A. H. Knoll, Eds. (Elsevier, Burlington, 2007), pp. 109–131.
25. J. L. Thorne, H. Kishino, I. S. Painter, *Mol. Biol. Evol.* **15**, 1647 (1998).
26. D. T. Jones, W. R. Taylor, J. M. Thornton, *Comput. Appl. Biosci.* **8**, 275 (1992).
27. H. Kishino, J. L. Thorne, W. J. Bruno, *Mol. Biol. Evol.* **18**, 352 (2001).
28. A. Stechmann, T. Cavalier-Smith, *Science* **297**, 89 (2002).
29. N. Arisue, M. Hasegawa, T. Hashimoto, *Mol. Biol. Evol.* **22**, 409 (2005).
30. H. S. Yoon, J. D. Hackett, C. Ciniglia, G. Pinto, D. Bhattacharya, *Mol. Biol. Evol.* **21**, 809 (2004).
31. C. de Vargas, M.-P. Aubry, I. Probert, J. Young, in *Evolution of Primary Producers in the Sea*, P. G. Falkowski, A. H. Knoll, Eds. (Elsevier Academic Press, Burlington, Massachusetts, 2007), pp. 251–285.
32. L. K. Medlin, A. G. Saez, J. R. Young, *Mar. Micropaleontol.* **67**, 69 (2008).
33. D. Bhattacharya, H. S. Yoon, J. D. Hackett, *BioEssays* **26**, 50 (2004).
34. J. Samuelsson, N. J. Butterfield, *Precambrian Res.* **107**, 235 (2001).
35. V. N. Sergeev, L. M. Gerasimenko, G. A. Zavarzin, *Microbiology* **71**, 623 (2002).
36. A. H. Knoll, E. J. Javaux, D. Hewitt, P. Cohen, *Philos. Trans. Roy. Soc. Lond. B* **361**, 1023 (2006).
37. E. J. P. Douzery, E. A. Snell, E. Bapteste, F. Delsuc, H. Philippe, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15386 (2004).
38. A. J. Roger, L. A. Hug, *Philos. Trans. Roy. Soc. Lond. B* **361**, 1039 (2006).
39. L. A. Hug, A. J. Roger, *Mol. Biol. Evol.* **24**, 1889 (2007).
40. S. B. Hedges, F. U. Battistuzzi, J. E. Blair, in *Neoproterozoic Geobiology and Paleobiology*, S. Xiao, A. J. Kaufman, Eds. (Springer, New York, 2006), pp. 199–229.
41. M. J. Benton, P. C. J. Donoghue, *Mol. Biol. Evol.* **24**, 26 (2007).
42. C. Berney, J. Pawlowski, *Proc. Roy. Soc. Lond. B* **273**, 1867 (2006).