Unexpected Dynamic Gene Family Evolution in Algal Actins

Min Wu, Josep M. Comeron,* Hwan Su Yoon,*† and Debashish Bhattacharya**

*Department of Biology and the Roy J. Carver Center for Comparative Genomics, University of Iowa; and †Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, ME

Actin is a conserved cytoskeletal protein that is well studied in model organisms although much less is known about actin molecular evolution in taxonomically diverse algae. Here, we analyzed 107 novel partial algal actin sequences and report some unexpected results. First, monophyletic actin gene families in multiple, phylogenetically distantly related algal taxa contain two distinct clades of sequences. One of these clades contains highly conserved sequences, whereas the second has multiple members with a significantly elevated substitution rate. This rate difference is associated with an excess of synonymous substitutions, strongly suggesting that both isoforms are active. These results paint a novel picture of actin gene evolution in algae showing it to be a remarkably dynamic system with duplication, homogenization, and potential functional diversification occurring independently in distantly related lineages.

Introduction

Actin is an abundant cytoskeletal protein that plays a central role in cell motility (Ghosh et al. 2004), cell division (Hill et al. 1996), determination of cell shape (Howard et al. 1993), and cytoplasmic streaming (Staiger et al. 1994). Despite the fundamental importance of actin, relatively little is known about its evolutionary history and functional diversification in nonmodel photosynthetic eukaryotes (algae). Algae are distributed throughout the tree of life but are particularly rich in the putative eukaryotic supergroups Plantae and Chromalveolata (for review, see Reyes-Prieto et al. 2007). In unicellular members of these supergroups, actin genes often occur as a single copy (Bhattacharya and Ehlting 1995), whereas in land plants, actin is present in complex gene families. Here, we investigated the evolutionary history of 107 novel, partial actin genomic fragments isolated from different Plantae and chromalveolates (see supplementary table S1 in the Supplementary Material at the MBE web site) to understand the evolutionary history of actin within and among the major photosynthetic protist lineages.

An overall picture of actin evolution was gained by analyzing our data within the context of 269 partial DNA sequences from the seven major algal phyla: red algae, green algae, glaucophytes, stramenopiles, dinoflagellates, haptophytes, and cryptophytes. A schematic representation of this tree is shown in supplementary figure S1, Supplementary Material online (the complete tree is available upon request from D.B.). This analysis shows as expected that actin genes support the monophyly of the different algal lineages (for exceptions, see supplementary fig. S1, Supplementary Material online). We then focused on the actin tree in each algal clade. An example is the nucleotide-based (fig. 1*A*) and protein-based (fig. 1*B*) trees of red algal actins that are broadly consistent with the recent classification of these taxa (Yoon et al. 2006) and previous analyses of bangiophyte actins (Hoef-Emden et al. 2005).

These trees and the actin DNA phylogenies for stramenopiles + dinoflagellates and cryptophytes + haptophytes (supplementary figs. S2 and S3, Supplementary Material online) are however more intriguing for the insights they provide into the evolutionary forces acting on this protein. Noteworthy in this regard is that *Flintiella sanguinaria*, *Glaucosphaera vacuolata*, and *Stylonema* sp. that are found in three different red algal classes together with the haptophytes *Pavlova gyrans* and *Pleurochrysis carterae* all share a similar pattern of actin gene duplication. Each species contains at least two distinct actin isoforms that form a monophyletic group (i.e., *Pavlova salina* is closely related to *P. gyrans*) comprised of one highly diverged actin lineage with a long branch that has accumulated a large number of DNA substitutions and another conserved isoform with a relatively short branch (fig. 1 and supplementary fig. S3, Supplementary Material online). These are well-studied (e.g., Hoef-Emden et al. 2002; Yoon et al. 2006) unialgal cultures; therefore, we have no reason to believe that this result reflects culture contamination with closely related taxa containing genes of widely differing divergence rates. Two observations argue against the possibility that these intriguing results are solely explained by polymerase chain reaction (PCR) or sequencing errors. First, under this scenario, we would expect a roughly equal distribution of actin gene gamily members (i.e., reflecting random sequencing errors) in each actin clade. This is often not the case (e.g., *F. sanguinaria Stylonema* sp., and *P. carterae* [see also *Storeatula* sp.]) with an excess of closely related actin gene sequences occurring in the diverged clade. Second, although rare, some of these actin genes contain spliceosomal introns (for complete alignment of novel actin genes, see supplementary fig. S4, Supplementary Material online). An example is shown in figure 1*C* for *Chrysochromulina polylepis* that encodes an intron after the first position in codon 123 (the conserved glutamine) in *Chlamydomonas reinhardtii* (XP_001699068) and other actins. We sequenced seven clones from this strain and found that each contained the intron. The sequence of this intron contains a microsatellite repeat that differs between the clones, sometimes with respect to a single CA repeat (i.e., DB5 vs. DB7) although one intron has many more changes (DB3). The DNA-based actin tree for *C. polylepis* shown in supplementary figure S3, Supplementary Material online, reveals a cluster of diverged actin genes similar to that described above that appears to arise at least partly from gene duplications.

We stress here that we do not argue that all members of the diverged actin clades define bona fide actin loci (i.e., PCR and sequencing errors likely account for some

Key words: actin, chromalveolates, gene duplication, rate acceleration, red algae.

E-mail: debashi-bhattacharya@uiowa.edu.

Mol. Biol. Evol. 26(2):249–253. 2009
doi:10.1093/molbev/msn263
Advance Access publication November 13, 2008

© The Author 2008. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

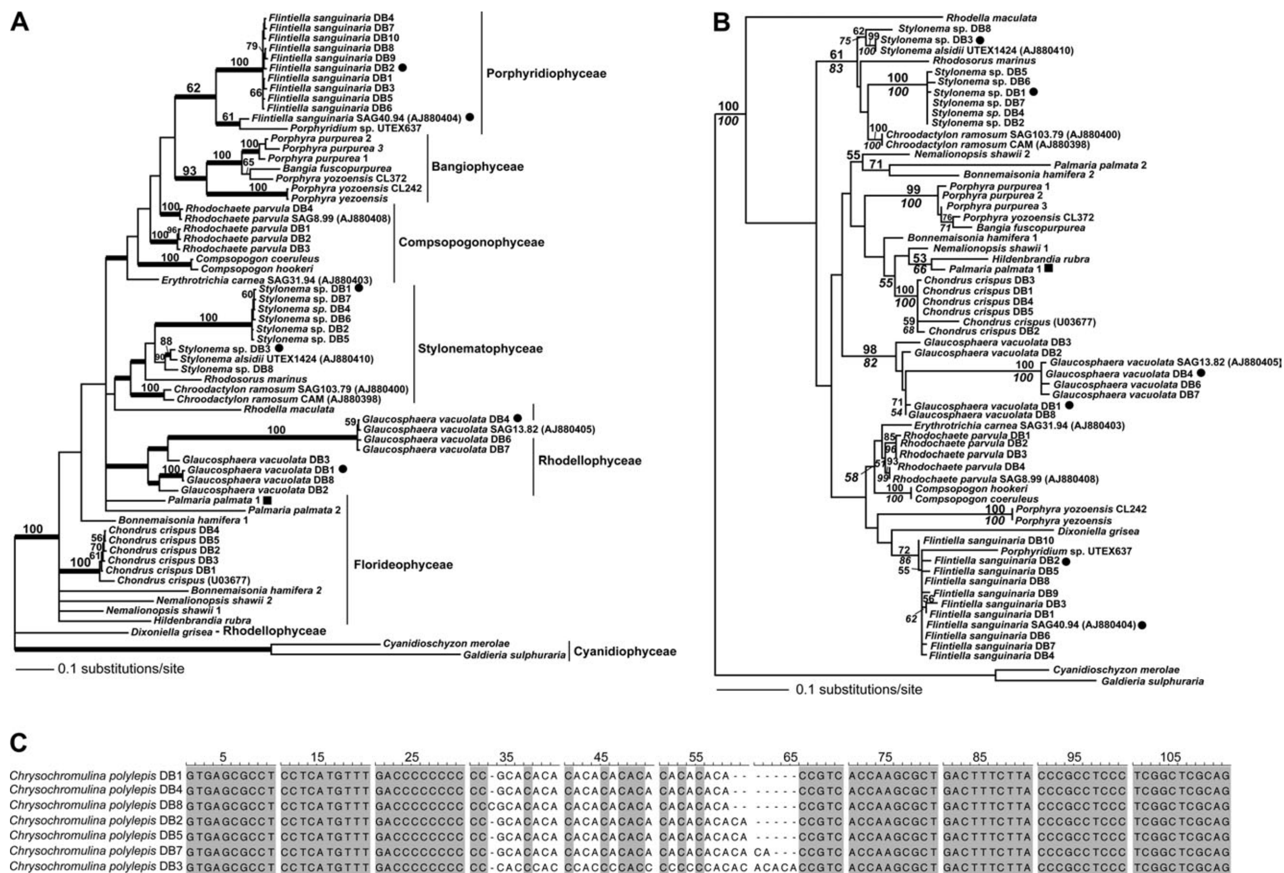

Fig. 1.—(*A*) Consensus phylogenetic tree of red algal actin DNA sequences based on Bayesian inference. The thick branches indicate >0.95 Bayesian posterior probability for the clade to the right. The numbers at the nodes are the result of a PHYML bootstrap analysis. (*B*) RAxML tree inferred from red algal actin protein sequences. The numbers above the nodes are the result of a RAxML bootstrap analysis, and those below the branches are from a PHYML bootstrap analysis. Only bootstrap values >50% are shown in both trees and branch lengths are proportional to the number of substitutions per site (see scale bars). (*C*) Alignment of a spliceosomal intron in seven *Chrysochromulina polylepis* genomic actin clones, with identical sites shown in the gray field.

sequence differences) but rather that the general pattern appears to be a valid result for these algae (see also below). The multiple independent cases across different algae also suggest this is a biologically relevant phenomenon. Finally, there is another case in which the same pattern of gene family evolution is found without taxon monophyly (i.e., *Storeatula* sp.), another for which we have less data but nonetheless contains conserved and diverged paralogs (*Goniomonas truncata*, supplementary fig. S3, Supplementary Material online), and others in which we apparently have identified one actin clade and not sampled the other using the PCR approach (e.g., *Diacronema* sp., *Chroomonas mesostigmatica*). We interpret these results as evidence for recent duplications of the actin genes followed by independent evolution in each lineage. The long-branched gene copies appear to have undergone concerted evolution resulting in sequence homogenization (e.g., *F. sanguinaria* and *G. vacuolata*) likely reflecting close genetic linkage of these loci in each of the algal genomes.

We used Tajima's (1993) test in MEGA4 (Tamura et al. 2007) to test the molecular clock hypothesis for pairs of actin DNA sequences from each of the two distinct duplicated gene family clades in four algal species. For *F. sanguinaria*, *G. vacuolata* and *Stylonema* sp., we used the *Palmaria palmata* 1 actin gene as outgroup (marked with a filled square in fig. 1*A*). The pairs of tested genes for the red algae are marked with filled circles in figure 1*A*. This analysis showed that for two red algae, there are significant DNA rate differences between the gene family representatives (*G. vacuolata*, $P = 0.00001$; *Stylonema* sp., $P = 0.028$), whereas the *F. sanguinaria* pair does not differ significantly using this test ($P = 0.821$). The protein tree for red algae (fig. 1*B*) is consistent with these results showing however significant rate differences only for *G. vacuolata* ($P = 0.00021$) and not for *Stylonema* sp. and *F. sanguinaria*. For the fourth alga, the haptophyte *P. gyrans* (outgroup is *Isochrysis galbana*, ingroups are *P. gyrans* DB1 and *P. gyrans* DB2), there was a significant DNA divergence rate difference between the genes ($P = 0.011$). These results show that significant molecular forces have acted on some algal actin gene families that deviate from the expectations for a conserved molecular marker. A possible explanation for this result is that these algae contain two highly diverged alleles, and not paralogs, with variation in each potentially resulting from PCR–sequencing errors that we are interpreting as gene family members. We consider this interpretation unlikely for the reasons provided above and because there is no precedent to expect that

Table 1
Nucleotide Changes within and between Actin Isoforms in Four Candidate Species

Species	Fixed Syn (Fs)	Fixed Rep (Fn)	Polym. Syn (Ps)	Polym. Rep (Pn)	NI	*P*
Flintiella sanguinaria	102	0	5	12	—	<0.0001
Glaucosphaera vacuolata	135	48	3	5	4.69	0.039
Stylonema sp.	104	26	42	17	1.62	n.s.
Pavolva gyrans G1/G2	31	51	6	9	0.91	n.s.
P. gyrans G1/G3	23	28	31	55	1.46	n.s.

Note.—NI represents the neutrality index (Rand and Kann 1996), computed as NI = (Pn/Ps)/(Fn/Fs). *P* indicates probability values of Fisher's exact test (two-tailed) of a 2 × 2 McDonald and Kreitman (1991) test; n.s. for $P > 0.05$. In *P. gyrans*, G1 contains three sequences: *P. gyrans* DB1, *P. gyrans* DB4, and *P. gyrans* DB6; G2 contains two sequences: *P. gyrans* DB2 and *P. gyrans* DB3a; G3 contains G2 and *P. gyrans* DB5 (supplementary fig. S3, Supplementary Material online). Syn and Rep are synonymous and replacement or amino acid changes, respectively.

alleles (clades of genes in this case) would simultaneously be highly divergent, accumulate changes at different rates, and segregate independently in populations of phylogenetically distantly related algae. Finally, there is no evidence of sexuality or existing data regarding ploidy in the bangiophyte red algal lineages studied in figure 1*A*. Therefore, it is also possible that these DNA samples arise from haploid cell lines, obviating issues about maintenance of diverged alleles.

The study of actin DNA sequences from the four target taxa shows that amino acid (replacement) differences are significantly less frequent than synonymous changes once corrected for their relative expectations based on the frequency of amino acid and synonymous sites (see table 1). This strongly suggests that both isoforms are active (i.e., non-pseudogenized). In the extreme case of *F. sanguinaria* (see table 1), there are no fixed amino acid changes between the two paralogs (i.e., see fig. 1*B*). Table 1 also shows the number of replacement and synonymous changes within actin isoforms. In *Stylonema* sp. and *P. gyrans*, the number of synonymous differences among isoforms is significantly greater than that expected (see above), and we can rule out the possibility that all differences among copies within each isoform are caused only by PCR-amplification errors. Therefore, we can confirm a minimum of two copies of actin in *F. sanguinaria* and *G. vacuolata* and multiple active copies within isoforms in *Stylonema* sp. and *P. gyrans*.

To study possible changes in selective trends after actin duplication, we compared the number of amino acid (replacement) and synonymous changes within and between actin isoforms (table 1; see Methods). When we focus on *Stylonema* sp. and *P. gyrans*, there is no evidence for a change in selective pressures at the protein level although we cannot exclude the possibility that the timing of the actin gene duplications was too distant in the past to allow detection of brief selective events immediately following the gene duplication event (see Lynch et al. 2006; Wagner 2008). This result, together with the observed differences in branch lengths (fig. 1*A*), suggests an increase in fixed synonymous changes that is also observed among isoforms. A rate difference that is restricted to synonymous changes of one isoform is not easily explained by neutral models. Therefore, we tentatively suggest that the cause for this increase in synonymous differences (between and within isoforms) might be associated with a change in expression rate. The translational selection hypothesis (Bulmer 1991; Akashi and Eyre-Walker 1998) proposes that the speed/accuracy of translation is influenced by the differential use of synonymous codons and that selection favors a set of preferred synonymous codons in highly expressed (i.e., translated) genes. Under this scenario, reduced actin gene expression levels would result in reduced selection leading to an increase in the synonymous substitution rate as observed in the divergent paralogs. Additional analyses of codon usage and transcript abundances in these species are however required to assess the validity of this proposal. In the other two species (*F. sanguinaria* and *G. vacuolata*) a direct comparison between differences between and within isoforms reveals a significant departure from expectations of steady selection (table 1), but, as indicated above, we cannot formally rule out the contribution of PCR errors to differences within isoforms.

In summary, we demonstrate that in different algae, recent actin duplications have resulted in at least two distinct paralog clades. These clades, in some cases, show significantly different evolutionary rates although there is no evidence suggesting that one duplicate has been pseudogenized. At the same time, the analysis of replacement and synonymous changes shows no evidence for relaxed (or positive) selection acting on amino acid sequence after duplication, with rate differences mostly associated with synonymous changes. These results paint a picture of actin evolution that stands in stark contrast to the widely accepted notion that unicellular algae maintain a single highly conserved actin gene (e.g., Bhattacharya and Weber 1997; Hoef-Emden et al. 2005), calling into question the use of actin as a phylogenetic marker among these taxa. It appears that gene duplications play a significant role in algal actin evolution that likely allows these taxa to explore the landscape of variation in actin function either through changes in gene expression or rare but potentially significant amino acid substitutions.

Methods

Isolation and Analysis of Genomic Actin Sequences

The sources of the algal cultures used in this study are shown in supplementary table S1, Supplementary Material online. Total DNA and actin PCR-amplification products were prepared as described in Yoon et al. (2008). At least four clones were picked for each species, and inserts were sequenced using BigDye Terminator Cycle Sequencing Kit (PE-Applied Biosystems, Norwalk, CT). Sequences were run on an ABI-3100 at the Roy J. Carver Center for Comparative Genomics at the University of Iowa. Additional actin sequences were obtained from the following sources: expressed sequence tag (EST) and complete genome data

from the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/), from the Cyanidiales red alga *Galdieria sulphuraria* (Barbier et al. 2005; Michigan State University Galdieria Database http://genomics.msu.edu/galdieria), and from the chlorophyte green alga *C. reinhardtii* and the diatom *Thalassiosira pseudonana* (DOE Joint Genome Institute, http://www.jgi.doe.gov/).

To align the novel sequences, we used ClustalW (Thompson et al. 1994) with the default settings and manually refined the alignment (see supplementary fig. S4, Supplementary Material online). A nucleotide alignment containing actin sequences from photosynthetic protists was constructed as well as three separate nucleotide data sets: one only of red algae, one only with cryptomonads and haptophytes, and one only with stramenopiles and dinoflagellates. Phylogenies were inferred for each nucleotide data set using maximum likelihood (ML) and Bayesian methods. The ML analyses were done using PHYML V2.4.3 (Guindon and Gascuel 2003) with the appropriate evolutionary model and tree optimization. Bayesian inference was performed using MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001). Modeltest (http://darwin.uvigo.es/software/modeltest.html) was used to select the best-fit model (always general time-reversible [GTR] $+$ I $+\Gamma$) and parameters for the nucleotide data sets. To assess the stability of monophyletic groups in the ML trees, 100 replicates for bootstrap were run under PHYML, and a 50% majority rule consensus was erected to determine the support values for each node. Bayesian analyses were performed with the parallel version of MrBayes using the GTR $+$ I $+$ Γ model of sequence evolution with Metropolis-coupled Markov chain Monte Carlo from a random starting tree. The Bayesian analyses were run for 5,000,000 generations with trees sampled each 100 cycles. Four chains were run simultaneously of which three were heated and one was cold, with the initial 200,000 cycles (2,000 trees) being discarded as the "burn-in." A consensus tree was made with the remaining 48,000 phylogenies to determine the posterior probabilities at each node. These actin nucleotide data were then translated into protein sequences (278 aa) and analyzed using ML methods. We used RAxML (Stamatakis et al. 2008) to infer the protein tree and used both RAxML (100 replicates) and PHYML under the Whelan and Goldman [WAG] $+$ Γ substitution model (Whelan and Goldman 2001), and parameters estimated during the tree search (100 replicates) to infer bootstrap support for the nodes in the tree.

Analysis of Gene Function

Analysis of actin gene function was done with sequences from *F. sanguinaria*, *G. vacuolata*, *Stylonema* sp., and *P. gyrans*. The alignment for each species was prepared in FASTA format, and DnaSP 4.0 (Rozas et al. 2003) was used to calculate the number of synonymous and amino acid changes of each actin isoform in each species, together with the number of synonymous and amino acid (replacement) substitutions of fixed differences between actin isoforms in each species. We then applied the McDonald–Kreitman test (McDonald and Kreitman 1991) as implemented in DnaSP 4.0 to the four species to study possible departures from neutral expectations. In our case, we investigated whether the ratio of synonymous over replacement changes among alleles of an isoform is the same as that between isoforms. Fisher's exact test, two-tailed, was applied to assess the statistical significance of the McDonald–Kreitman 2×2 tests. The neutrality index (NI) was estimated following Rand and Kann (1996): NI indicates the extent to which the levels of polymorphic amino acid variation depart from those expected in the neutral model. NI > 1 indicates an excess of polymorphic amino acid variation, whereas NI < 1 indicates an excess of fixed amino acid variation, likely due to positive selection.

Supplementary Material

Supplementary figures S1–S4 and supplementary table S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/). The 107 novel actin sequences generated in this study are available in fasta format as figure S4 in the Supplementary Material.

Acknowledgments

This work was supported by grants from the National Science Foundation awarded to D.B. (EF 04-31117, EF 06-25440) and to J.M.C. (DEB-03-44209). We are grateful to two anonymous reviewers for their constructive criticisms.

Literature Cited

Akashi H, Eyre-Walker A. 1998. Translational selection and molecular evolution. Curr Opin Genet Dev. 8:688–693.

Barbier G, Oesterhelt C, Larson MD, Halgren RG, Wilkerson C, Garavito RM, Benning C, Weber APM. 2005. Genome analysis. Comparative genomics of two closely related unicellular thermo-acidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria* and significant differences in carbohydrate metabolism of both algae. Plant Physiol. 137:460–474.

Bhattacharya D, Ehlting J. 1995. Actin coding regions— gene family evolution and use as a phylogenetic marker. Archiv für Protistenkunde. 145:155–164.

Bhattacharya D, Weber K. 1997. The actin gene of the glaucocystophyte *Cyanophora paradoxa*: analysis of the coding region and introns, and an actin phylogeny of eukaryotes. Curr Genet. 31: 439–446.

Bulmer M. 1991. The selection–mutation-drift theory of synonymous codon usage. Genetics. 129:897–907.

Ghosh M, Song XY, Mouneimne G, Sidani M, Lawrence DS, Condeelis JS. 2004. Cofilin promotes actin polymerization and defines the direction of cell motility. Science. 304: 743–746.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52:696–704.

Hill KL, Catlett NL, Weisman LS. 1996. Actin and myosin function in directed vacuole movement during cell division in *Saccharomyces cerevisiae*. J Cell Biol. 135:1535–1549.

Hoef-Emden K, Marin B, Melkonian M. 2002. Nuclear and nucleomorph SSU rDNA phylogeny in the Cryptophyta and the evolution of cryptophyte diversity. J Mol Evol. 55:161–179.
Hoef-Emden K, Shrestha RP, Lapidot M, Weinstein Y, Melkonian M, Arad SM. 2005. Actin phylogeny and intron distribution in bangiophyte red algae (Rhodoplantae). J Mol Evol. 61:360–371.
Howard PK, Sefton BM, Firtel RA. 1993. Tyrosine phosphorylation of actin in *Dictyostelium* associated with cell-shape changes. Science. 259:241–244.
Huelsenbeck JP, Ronquist F. 2001. MRBAYES: bayesian inference of phylogenetic trees. Bioinformatics. 17:754–755.
Lynch VJ, Roth JJ, Wagner GP. 2006. Adaptive evolution of Hox-gene homeodomains after cluster duplications. BMC Evol Biol. 6:86.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature. 351:652–654.
Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. Mol Biol Evol. 13:735–748.
Reyes-Prieto A, Weber AP, Bhattacharya D. 2007. The origin and establishment of the plastid in algae and plants. Annu Rev Genet. 41:147–168.
Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.
Staiger CJ, Yuan M, Valenta R, Shaw PJ, Warn RM, Lloyd CW. 1994. Microinjected profilin affects cytoplasmic streaming in plant-cells by rapidly depolymerizing actin microfilaments. Curr Biol. 4:215–219.
Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web-servers. Syst Biol. 75:758–771.
Tajima F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics. 135:599–607.
Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596–1599.
Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673–4680.
Wagner A. 2008. Gene duplications, robustness and evolutionary innovations. Bioessays. 30:367–373.
Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol. 18:691–699.
Yoon HS, Grant J, Tekle YI, Wu M, Chaon BC, Cole JC, Logsdon JM, Patterson DJ, Bhattacharya D, Katz LA. 2008. Broadly sampled multigene trees of eukaryotes. BMC Evol Biol. 8:14.
Yoon HS, Muller KM, Sheath RG, Ott FD, Bhattacharya D. 2006. Defining the major lineages of red algae (Rhodophyta). J Phycol. 42:482–492.

Martin Embley, Associate Editor

Accepted November 7, 2008