Monophyly of Rhizaria and Multigene Phylogeny of Unicellular Bikonts

Burki, Fabien; Pawlowski, Jan Wojciech

How to cite

BURKI, Fabien, PAWLOWSKI, Jan Wojciech. Monophyly of Rhizaria and Multigene Phylogeny of Unicellular Bikonts. In: Molecular biology and evolution, 2006, vol. 23, n° 10, p. 1922–1930. doi: 10.1093/molbev/msl055

This publication URL: https://archive-ouverte.unige.ch/unige:171586
Publication DOI: 10.1093/molbev/msl055

Monophyly of Rhizaria and Multigene Phylogeny of Unicellular Bikonts

Fabien Burki and Jan Pawlowski

Department of Zoology and Animal Biology, University of Geneva, Geneva, Switzerland

Reconstructing a global phylogeny of eukaryotes is an ongoing challenge of molecular phylogenetics. The availability of genomic data from a broad range of eukaryotic phyla helped in resolving the eukaryotic tree into a topology with a rather small number of large assemblages, but the relationships between these "supergroups" are yet to be confirmed. Rhizaria is the most recently recognized "supergroup," but, in spite of this important position within the tree of life, their representatives are still missing in global phylogenies of eukaryotes. Here, we report the first large-scale analysis of eukaryote phylogeny including data for 2 rhizarian species, the foraminiferan *Reticulomyxa filosa* and the chlorarachniophyte *Bigelowiella natans*. Our results confirm the monophyly of Rhizaria (Foraminifera + Cercozoa), with very high bootstrap supports in all analyses. The overall topology of our trees is in agreement with the current view of eukaryote phylogeny with basal division into "unikonts" (Opisthokonts and Ameobozoa) and "bikonts" (Plantae, alveolates, stramenopiles, and excavates). As expected, Rhizaria branch among bikonts; however, their phylogenetic position is uncertain. Depending on the data set and the type of analysis, Rhizaria branch as sister group to either stramenopiles or excavates. Overall, the relationships between the major groups of unicellular bikonts are poorly resolved, despite the use of 85 proteins and the largest taxonomic sampling for this part of the tree available to date. This may be due to an acceleration of evolutionary rates in some bikont phyla or be related to their rapid diversification in the early evolution of eukaryotes.

Introduction

Resolving the structure of the phylogenetic tree of eukaryotes is of crucial importance for understanding the major evolutionary steps that could possibly explain the relationships between species. During the last 2 decades, the advances in molecular systematics led to establishing new monophyletic assemblages and helped in drawing the relations between the numerous lineages recognized on the basis of morphological and ultrastructural data. At first, based almost exclusively on the small-subunit rRNA (SSU rRNA) gene (Sogin et al. 1989; Sogin 1991; Kumar and Rzhetsky 1996; Pawlowski et al. 1996; Sogin and Silberman 1998), molecular phylogenies of eukaryotes were subsequently tested with protein-coding genes (Yamamoto et al. 1997; Moreira et al. 1999; Philippe et al. 2000). Despite their important role in the early days of molecular phylogenetics, single-gene phylogenies are now known to be highly sensitive to variation of evolutionary rates, which often led to false representation of early eukaryotic evolution (Stiller and Hall 1999; Morin 2000; Philippe 2000; Philippe and Germot 2000).

Over time, the accumulation of protein sequences from a large variety of eukaryotes has made it possible to test single-gene phylogenies using combined data (Baldauf et al. 2000). A new view of global phylogeny of eukaryotes emerged from a growing number of evidence based on several different kinds of mutually reinforcing data, such as 1) multiple gene phylogenies (Bapteste et al. 2002; Yoon et al. 2002; Philippe et al. 2004; Hampl et al. 2005; Harper et al. 2005; Philippe et al. 2005; Rodriguez-Ezpeleta et al. 2005; Simpson et al. 2006; Steenkamp et al. 2006), 2) individual phylogenies converging on the same relationships (Fast et al. 2002; Simpson, Roger, et al. 2002; Longet et al. 2004), 3) discrete characters (Baldauf and Palmer 1993; Keeling and Palmer 2001; Stechmann and Cavalier-Smith 2002; Archibald, Longet, et al. 2003), and 4) morphological and ultrastructural data (Simpson, Radek, et al. 2002). Overall, the vast majority of the known diversity of eukaryotes seems to be distributed among only 5 to 6 major divisions that are probably all monophyletic, referred to as the plants, excavates, chromalveolates, Rhizaria (all belonging to the assemblage of the so-called "bikonts") and the "unikonts," which comprise the Opisthokonts and Amoebozoa (Keeling et al. 2005). Identifying these natural supergroups raised the new challenge of understanding the relationships among them, which, for most of the eukaryotic tree, has yet to be confirmed.

Rhizaria (Cavalier-Smith 2002) is a recently emerged supergroup of eukaryotes enclosing organisms as diverse as filose testate amoebae, cercomonads, chlorarachniophytes, Foraminifera, plasmodiophorids, haplosporidians, gromiids, and radiolarians (Adl et al. 2005). The first hints for the evolutionary meaning of the group came from SSU rRNA–based phylogenies (Bhattacharya et al. 1995; Cavalier-Smith and Chao 1997). Rapidly, the phylum Cercozoa was created to accommodate this new assemblage (Cavalier-Smith 1998). Further molecular studies confirmed the heterogeneity of this phylum, with various protists being included in it (Burki et al. 2002; Cavalier-Smith and Chao 2003; Polet et al. 2004). Protein data indicated a relationship between Foraminifera and Cercozoa (Keeling 2001; Archibald, Longet, et al. 2003; Longet et al. 2003), and a combined analysis of SSU rRNA and actin confirmed their relation with Radiolaria (Nikolaev et al. 2004). Finally, a study of a single or double amino acid insertion in the protein polyubiquitin suggests that Radiolaria represent the most basal branch of Rhizaria, followed by Foraminifera and Cercozoa (Bass et al. 2005).

Despite their now well-accepted taxonomic status, the Rhizaria are still missing in most of the multigene phylogenies published to date (Bapteste et al. 2002; Philippe et al. 2004; Hampl et al. 2005; Rodriguez-Ezpeleta et al. 2005; Steenkamp et al. 2006). Until recently, the only available rhizarian genomic information was an expressed sequence tag (EST) data set for the chlorarachniophyte *Bigelowiella natans*, comprising about 3,500 sequences (Keeling and Palmer 2001). Some of these sequences have been used in studies with other purposes than exploring the phylogenetic

Key words: eukaryotes, unicellular bikonts, Rhizaria, multigene phylogeny.

E-mail: fabien.burki@zoo.unige.ch.

Mol. Biol. Evol. 23(10):1922–1930. 2006
doi:10.1093/molbev/msl055
Advance Access publication July 7, 2006

position of Rhizaria (de Koning et al. 2005; Harper et al. 2005) or are even absent from the final trees because of a suspected artifactual position (Simpson et al. 2006).

To include Rhizaria in multigene phylogenies of eukaryotes, we have recently conducted an EST project on the freshwater naked foraminiferan *Reticulomyxa filosa*, which led to approximately 1,600 high-quality sequences (Burki et al. forthcoming). Combining the available genomic information, we assembled in this study a data set of 85 orthologous proteins for 37 eukaryotic species, including the 2 rhizarian species *R. filosa* and *B. natans*, in order to 1) confirm the monophyly of Rhizaria when using a large number of protein-coding genes and 2) infer the phylogenetic position of this supergroup within eukaryotes.

Materials and Methods

Construction of the Alignment

Using our *R. filosa* ESTs as queries, we performed BlastX searches against the UniProt protein database on the Swiss Institute of Bioinformatics server to find sufficiently conserved genes in a broad taxonomic sampling of eukaryotes. A homemade perl script linking the Blast output and the seqret program from the EMBOSS package (http://emboss.ch.embnet.org/EMBOSSDOC/programs/html/seqret.html) allowed us to retrieve and store in different files (each corresponding to a different *R. filosa* gene) all sequences from the database with an *e*-value $< 10^{-40}$. This relatively stringent cutoff was defined in order to avoid the integration of paralogous genes. The homologous proteins were then aligned with ClustalW (Thompson et al. 1994) and kept for further analyses if they 1) showed a reasonable taxonomic distribution and 2) were conserved enough across all eukaryotes.

To increase the number of eukaryotes represented, we downloaded all available nucleotide sequences from GenBank through the taxonomy browser at National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) for the stramenopiles, ciliates, alveolates, *Entamoeba*, *Physcomitrella*, Rhodophyta, *Strongylocentrotus*, *Schistosoma*, *Giardia*, *Trichomonas*, *Alexandrium*, and *B. natans*. We searched for homology between this constructed data set and our *R. filosa* ESTs by performing local TBlastX (threshold $< 10^{-40}$) and added the resulting matching sequences to our alignments. At this point, only the genes found either in both *R. filosa* and *B. natans* or only in *R. filosa* had been retained. To increase the number of genes, we repeated the blasting and selecting procedures using this time the *B. natans* sequences as query. Overall, it resulted in a data set of homologous aligned genes containing for Rhizaria both *R. filosa* and *B. natans*, only *R. filosa*, or only *B. natans*, in addition to all other eukaryotic species. Alignments were eye checked and refined manually with BioEdit 7.0.5 (Hall 1999), and ambiguously aligned positions were removed with Gblocks (Castresana 2000).

Because of the limited data for certain groups and to maximize the number of genes by taxonomic assemblage, some higher taxa were represented by different closely related species: *Paramecium*, *Phytophthora*, *Cryptosporidium*, Rhodophyta, and *Theileria* (for details see Supplementary Table S1, Supplementary Material online). To decide on the final set of genes used in this study, we checked for orthology between all the retrieved sequences for each selected genes by first carrying out a Neighbor-Joining (NJ) analysis with the program PROTDIST 3.6 (Felsenstein 2004), allowing us to discard very distant paralogous genes. To refine our selection, we then constructed for each gene a maximum likelihood (ML) tree using PHYML (JTT + F + Γ_4) (Guindon and Gascuel 2003) so that we were able to keep genes only where clear orthology between species could be identified.

Phylogenetic Analyses

We concatenated all genes into alignments that were analyzed with both ML and Bayesian Inference (BI). ML analyses utilized the programs PHYML (Guindon and Gascuel 2003) and TREEFINDER (Jobb et al. 2004). Following the Akaike Information Criterion (AIC) (Posada and Buckley 2004) computed with ProtTest 1.2.6 (Abascal et al. 2005), the RtREV + F + Γ model allowing between-site rate variation was chosen (calculations were done with 8 gamma categories). Coming right after according to the AIC, the WAG model was also tested and gave the same topologies. To estimate the robustness of the phylogenetic inference, we used the bootstrap method (Felsenstein 1985) with 100 pseudoreplicates generated and analyzed with PHYML and TREEFINDER.

Bayesian analyses using the WAG + F + Γ_4 model were performed with the parallel version of MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003). Each inference, starting from a random tree and using 4 Metropolis-coupled Markov Chain Monte Carlo (MCMCMC), consisted of 1,000,000 generations with sampling every 100 generations. The average standard deviation of split frequencies was used to assess the convergence of the 2 runs. Bayesian posterior probabilities were calculated from the majority rule consensus of the tree sampled after the initial burn-in period as determined by checking the convergence of likelihood values across MCMCMC generations (corresponding to roughly 20,000–50,000 generations, depending on the analysis).

In subsequent analyses, amino acid positions were successively removed from the complete alignment (CA) according to their substitution rates. Substitution rates at sites were computed with the program CODEML from the PAML package (Yang 1997), given the 15 possible trees uniting the bikonts when alveolates, stramenopiles, Rhizaria, and excavates are defined as a multifurcation, and the WAG model with all parameters to be estimated (12 gamma categories). Based on the substitution rates expressed in number of substitution per sites, we defined several categories of sites (i.e., going from the fastest evolving sites to slower evolving sites). Seven different alignments were generated, each having 1 category plus the faster categories of sites removed (see fig. 2 for the details).

PHYML and CODEML were executed on the Vital-IT computational facilities at the Swiss Institute of Bioinformatics (http://www.vital-it.ch). The parallel (MPI) MrBayes was run at the freely available University of Oslo Bioportal (http://www.bioportal.uio.no).

Testing Phylogenies

Phylogenetic hypotheses were tested using the approximately unbiased (AU) test (Shimodaira 2002). For each tested tree, site likelihoods were calculated using CODEML and the AU test was performed using CONSEL (Shimodaira and Hasegawa 2001) with default scaling and replicate values.

Results

Sequences and Alignments

Thirty-seven eukaryotic species representing a broad taxonomic sampling and for which a large amount of data are available were selected. From our initial data set, we retained 85 proteins (see Supplementary Table S2, Supplementary Material online) according to the following criteria: 1) at least 19 species out of the total of 37 ($> 50\%$) could be retrieved, 2) at least one out of the 2 rhizarian species were present, and 3) the orthology between all species was unambiguous on the base of ML trees. To minimize missing data in Rhizaria, sequences were shorted by removing all sites if present neither in *R. filosa* nor in *B. natans*, leading to a final concatenated alignment of 13,258 amino acid positions (CA). Overall, the average missing data across the alignment were 21% with a minimum of no missing data in *Homo sapiens* and *Drosophila melanogaster* (0%) and a maximum in *Alexandrium tamarense* (79.55%) (for a detailed list see Supplementary Table S1, Supplementary Material online).

We also considered for analyses a reduced alignment where genes not found in our *R. filosa* ESTs survey were taken off, leaving 9,947 amino acid positions (*R. filosa* no missing data alignment or NMDA). This has been done for 2 reasons. First, *R. filosa* is our organism of main interest, thus we wanted to have an alignment without any missing data for this species. Second, the *B. natans* EST data set contains a lot of sequences encoding plastid-targeted proteins with a chlorophyte green algal origin for the most part but also with streptophyte algae, red algae, or even bacteria origins (Archibald, Roger, et al. 2003). Although quite a few of these ESTs have already been annotated (Archibald, Roger, et al. 2003; Rogers et al. 2004), it was crucial to avoid the mixture of host genes with nonannotated endosymbiont or laterally transferred genes. Based on separate phylogenetic analyses for each selected gene, we were able to discard many questionable *B. natans* genes (i.e., *B. natans* genes that doubtfully branched very closely to plants), but one might still argue that some genes with only *B. natans* as rhizarian species in our CA have originated through secondary endosymbiosis or lateral gene transfer. Thus, considering the NMDA where for every *B. natans* sequence an orthologous rhizarian sequence from *R. filosa* was available lead to higher confidence in our results (see below).

Phylogenetic Position of Rhizaria

The analyses of the CA and the NMDA give trees of generally similar structure (fig. 1 and Supplementary Material online), congruent with global eukaryotic phylogenies inferred in previous EST-based studies (Philippe et al. 2004). In all analyses, 3 major assemblages of species can be distinguished. The first assemblage comprises animals, fungi, and Amoebozoa, that is, the "unikonts" of Stechmann and Cavalier-Smith (2003). The second assemblage is composed of green plants and rhodophytes, which form a strongly supported grouping of the primary photosynthetic eukaryotes (Rodriguez-Ezpeleta et al. 2005). The third assemblage includes all other unicellular "bikonts" (stramenopiles, alveolates, rhizarians, and excavates). These 3 major assemblages are strongly supported in the analysis of the CA, but, with the exception of the MrBayes analysis, the support is globally weaker in the case of NMDA (Supplementary Figs. S1, S3, and S5, Supplementary Material online). Although most of the supergroups of eukaryotes, including Rhizaria, are recovered in all analyses, their relationships are not well resolved. In particular, the assemblage of unicellular bikonts appears as an unresolved radiation of 4 supergroups (fig. 1).

The phylogenetic position of Rhizaria varied depending on both the type of alignment and the method of analysis. In the ML (PHYML) analysis of the CA (Supplementary Fig. S2, Supplementary Material online), Rhizaria branch as sister group to stramenopiles, whereas in the Bayesian analysis (Supplementary Fig. S4, Supplementary Material online), they branch as sister group to excavates. This last topology was also found in the ML analysis using TREEFINDER, but in this case, the ciliates branched between Rhizaria and excavates (not shown). Both ML and Bayesian methods show Rhizaria branching as sister group to excavates in analysis of the NMDA, but the bootstrap support for this and other groupings was rather weak (Supplementary Figs. S3 and S5, Supplementary Material online).

To better examine the position of Rhizaria, we successively removed some fast-evolving lineages, which could potentially introduce systematic bias in our analyses, especially with analyses of large-scale data sets (Brinkmann et al. 2005; Jeffroy et al. 2006). In particular, to avoid a long-branch attraction (LBA) artifact (Felsenstein 1978), we reanalyzed our data in absence of excavates or ciliates, which appeared particularly unstable in our analyses. These modifications of species composition had different impacts on the rhizarian position, depending on both the alignment studied and the method used. After removing both *Giardia* and *Trichomonas*, or all excavates at the same time, the topology of the CA tree remained unchanged (see Supplementary Fig. S2, Supplementary Material online), whereas the NMDA topology was drastically changed as the relationship between Rhizaria and stramenopiles was recovered (data not shown). When ciliates were removed, Rhizaria branched as sister group to excavates in ML analyses of both complete and NMDA. Finally, because *R. filosa* has a slightly longer branch than *B. natans* (see fig. 1), we tested whether *B. natans* alone prefers the excavate or the stramenopile position by reconstructing a TREEFINDER tree (not shown). Interestingly, it branched as sister to stramenopiles preventing us to rule out the possibility that the relationship between Rhizaria and excavates is due to the rapid evolutionary rates of foraminifers.

This observed instability could indicate the presence in the data of 2 opposite signals of similar strength (a

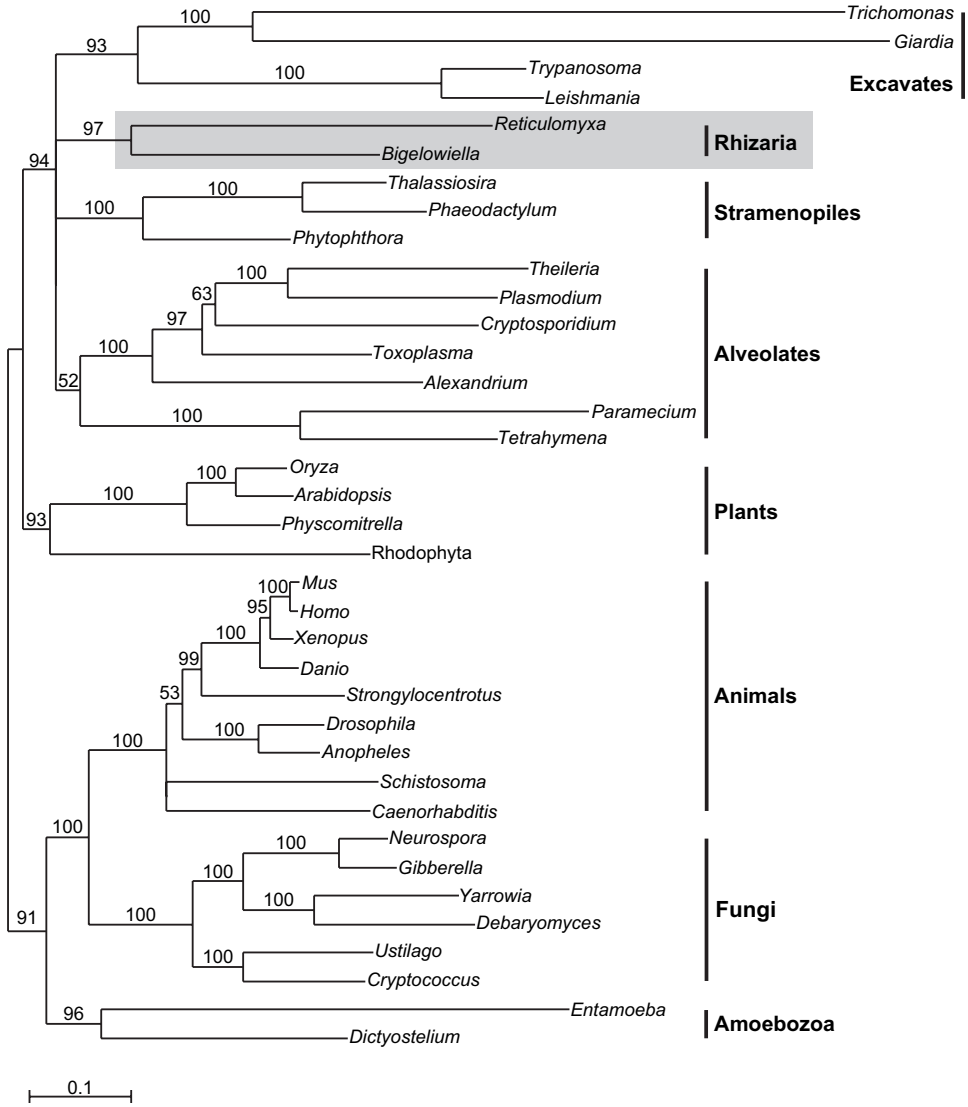
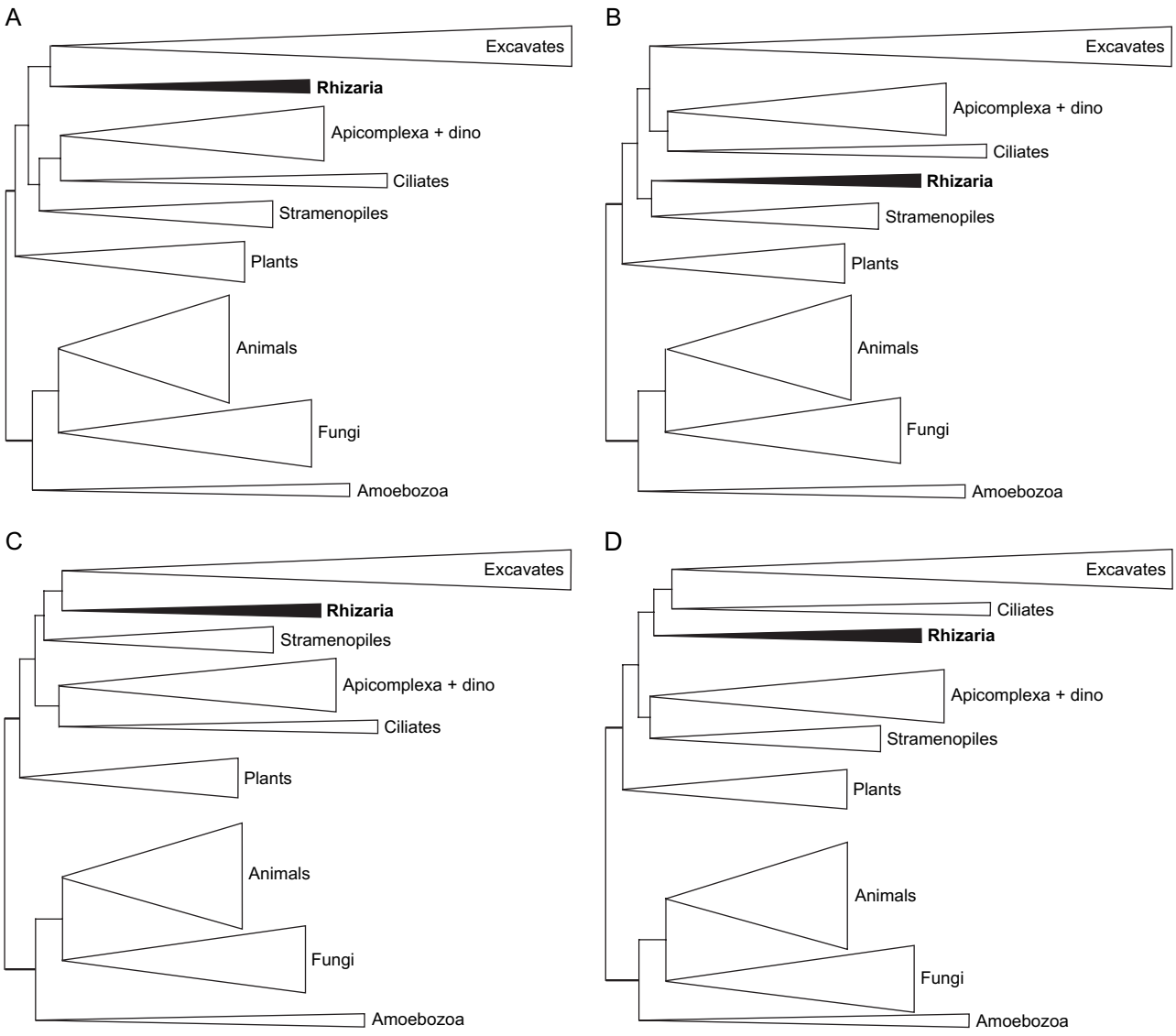


FIG. 1.—Consensus ML phylogenetic tree as obtained with TREEFINDER after the analysis of the complete data set (CA). Hundred bootstrap replicates were done (bootstrap support are represented by the numbers at nodes), and the unresolved nodes correspond to relationships recovered in less than 50 replicates.

phylogenetic and a nonphylogenetic signal) that prevent phylogenetic methods from finding the true evolutionary tree (N. Rodriguez-Ezpeleta and H. Philippe, personal communication). One way to eliminate the nonphylogenetic signal and extract the true evolutionary information is the removal of potentially saturated fast-evolving sites. To do this, we divided the fastest evolving amino acid positions in the CA in different categories according to their evolutionary rates and inferred ML trees based on alignments successively shortened by removing a class of sites. Figure 2 (*A*, *B*, *C*, and *D*) shows the 4 different topologies we obtained and their occurrence (fig. 2*E*). As one can notice, the relationships were very dependent on both the alignment and the method. PHYML gave a mixture of topologies B and C, whereas TREEFINDER mostly found topology D but also found topology B when the 5 fastest categories were removed. Based on these comparisons, one cannot obviously decide in favor of a particular topology as no clear pattern appears.

Additionally, to assess a confidence level for the comparison of the topologies, we performed the AU test, which is considered as the least-biased and most rigorous test available to date (Shimodaira 2002). Precisely, the only 4 different topologies obtained during this study (i.e., topologies in fig. 2) were tested, given CA, NMDA, and the 7 alignments resulting from the removal of class of sites. The rows 2 to 5 of table 1 corresponds to the comparison of the 4 trees given CA and shows that no topology can be rejected although topology D is just above the limit at the significance level of 0.05. Focusing on NMDA, the AU test significantly rejects topologies B and D (rows 6–9 of table 1), keeping only solutions where Rhizaria are directly related to excavates. As we go further down, the rest of the results in table 1 means that all topologies passed the test (no rejection), except topology D, which is either discarded with the shortest alignments or just above the rejection limit.



E

	Complet alignment	No *R.filosa* missing data	Without 3.80	Without 3.70	Without 3.60	Without 3.50	Without 3.40	Without 3.30	Without 3.20
Aln length	13258	9947	13006	12827	12689	12590	12507	12434	12370
PHYML	B	A	B	C	B	C	B	B	C
TreeFinder	D	A	D	D	D	D	D	B	B
Mrbayes	C	A	-	-	-	-	-	C	-

FIG. 2.—Results of the fast-evolving site removal analysis. (*A*, *B*, *C*, and *D*) The 4 different topologies obtained after successively excluding classes of sites (see Materials and Methods for details). The length of triangles corresponds to the branch length of the faster evolving lineage in that group, and the width is proportional to the number of taxa included in our analyses. (*E*) Summary of the different data sets analyzed with for each class (columns) the length of the alignment in amino acids and the topology obtained with PHYML, TREEFINDER, and MrBayes (rows).

Table 1
Likelihood AU Test of Alternative Tree Topologies

Alignments/Tree Topologies	Δ Ln L[a]	AU Test
Complet (CA)/Rhiz. sister to exc. chromal. (fig. 2*A*)	50.5	0.147
Complet (CA)/Rhiz. sister to stram. (fig. 2*B*)	14.6	0.412
Complet (CA)/Rhiz. sister to exc. (fig. 2*C*)	−14.6	0.802
Complet (CA)/Rhiz. sister to ciliates + exc. (fig. 2*D*)	69.9	0.051
No missing data *R. filosa* (NMDA)/Rhiz. sister to exc. chromal. (fig. 2*A*)	−15.5	0.702
No missing data *R. filosa* (NMDA)/Rhiz. sister to stram. (fig. 2*B*)	63.2	0.028
No missing data *R. filosa* (NMDA)/Rhiz. sister to exc. (fig. 2*C*)	15.5	0.438
No missing data *R. filosa* (NMDA)/Rhiz. sister to ciliates + exc. (fig. 2*D*)	71.1	0.044
Without 3.80/Rhiz. sister to exc. chromal. (fig. 2*A*)	42.4	0.183
Without 3.80/Rhiz. sister to stram. (fig. 2*B*)	15.8	0.386
Without 3.80/Rhiz. sister to exc. (fig. 2*C*)	−15.8	0.816
Without 3.80/Rhiz. sister to ciliates + exc. (fig. 2*D*)	71.4	0.052
Without 3.70/Rhiz. sister to exc. chromal. (fig. 2*A*)	67.6	0.227
Without 3.70/Rhiz. sister to stram. (fig. 2*B*)	18.1	0.360
Without 3.70/Rhiz. sister to exc. (fig. 2*C*)	−18.1	0.821
Without 3.70/Rhiz. sister to ciliates + exc. (fig. 2*D*)	37.6	0.063
Without 3.60/Rhiz. sister to exc. chromal. (fig. 2*A*)	50.7	0.138
Without 3.60/Rhiz. sister to stram. (fig. 2*B*)	11.2	0.420
Without 3.60/Rhiz. sister to exc. (fig. 2*C*)	−11.2	0.793
Without 3.60/Rhiz. sister to ciliates + exc. (fig. 2*D*)	69.2	0.057
Without 3.50/Rhiz. sister to exc. chromal. (fig. 2*A*)	52.5	0.117
Without 3.50/Rhiz. sister to stram. (fig. 2*B*)	14.7	0.401
Without 3.50/Rhiz. sister to exc. (fig. 2*C*)	−14.7	0.805
Without 3.50/Rhiz. sister to ciliates + exc. (fig. 2*D*)	74.7	0.036
Without 3.40/Rhiz. sister to exc. chromal. (fig. 2*A*)	57.0	0.093
Without 3.40/Rhiz. sister to stram. (fig. 2*B*)	15.5	0.374
Without 3.40/Rhiz. sister to exc. (fig. 2*C*)	−15.5	0.818
Without 3.40/Rhiz. sister to ciliates + exc. (fig. 2*D*)	76.8	0.040
Without 3.30/Rhiz. sister to exc. chromal. (fig. 2*A*)	58.1	0.104
Without 3.30/Rhiz. sister to stram. (fig. 2*B*)	12.5	0.409
Without 3.30/Rhiz. sister to exc. (fig. 2*C*)	−12.5	0.775
Without 3.30/Rhiz. sister to ciliates + exc. (fig. 2*D*)	80.1	0.022
Without 3.20/Rhiz. sister to exc. chromal. (fig. 2*A*)	60.3	0.095
Without 3.20/Rhiz. sister to stram. (fig. 2*B*)	14.5	0.372
Without 3.20/Rhiz. sister to exc. (fig. 2*C*)	−14.5	0.794
Without 3.20/Rhiz. sister to ciliates + exc. (fig. 2*D*)	84.1	0.023

NOTE.—Underlined numbers correspond to the significant *P* values of the rejected topologies. Abbreviations are as follows: Rhiz. = Rhizaria; exc. = excavates; chromal. = chromalveolates; stram. = stramenopiles.

[a] Log likelihood difference.

Discussion

Our data bring a new multigenic evidence for the close evolutionary relationships between Foraminifera and Cercozoa. The branching of the foraminiferan *R. filosa* and the chlorarachniophyte *B. natans* receives strong bootstrap support in all our analyses. Besides, these 2 species branch together in all different topologies we obtained (fig. 2). The relationships between these 2 phyla were previously suggested based on analyses of actin (Keeling 2001; Flakowski et al. 2005), polyubiquitin (Archibald, Longet, et al. 2003; Bass et al. 2005), RNA polymerase (Longet et al. 2003), and SSU rRNA gene (Berney and Pawlowski 2003; Cavalier-Smith and Chao 2003). With more than 80 analyzed genes, our study strongly confirms these single-gene analyses, providing a compelling evidence for the monophyly of Rhizaria. However, as this supergroup is very heterogenous (Adl et al. 2005), the phylogenetic position of other putative rhizarians, especially the polycystine and acantharian radiolarians, should still be confirmed by multigene data.

Although the monophyly of Rhizaria (Cercozoa + Foraminifera) was ascertained by our data, their phylogenetic position in the eukaryotic tree remains questionable. Two concurrent hypotheses on the relationships between Rhizaria and other eukaryotes were brought by our analyses, preventing us from a univocal conclusion. According to the first hypothesis, Rhizaria are sister group to excavates. There are several lines of evidence supporting this hypothesis: 1) all phylogenetic reconstruction methods used in this study show this association when an alignment with for *R. filosa* is analyzed; 2) if ciliates are removed from the taxa sampling, this union is also recovered with the alignment of the complete data set; 3) topology comparisons never reject trees where Rhizaria are specifically related to excavates and they are always the best plausible trees examined. Finally, this relationship has been previously suggested based on the presence of secondary symbiosis with green algae in some excavates

(Euglena) and some rhizarians (chlorarachniophytes) and is known as the cabozoan hypothesis (Cavalier-Smith 1999).

More unexpected is the second hypothesis suggesting that Rhizaria are sister group to stramenopiles. The branching of Rhizaria and stramenopiles is shown by many of ML analyses (fig. 2) and none of these trees can be statistically rejected (table 1). Moreover, Rhizaria also branch with stramenopiles when fast-evolving excavate sequences are removed as well as when the less divergent *B. natans* sequence in isolation is kept. If this configuration turns out to be correct with additional evidence such as discrete characters or phylogenomic analyses of other less rapidly evolving rhizarians, it would have important implications on the chromalveolates hypothesis (Harper et al. 2005). This hypothesis is based, among others, on a specific model of plastid evolution suggesting that both stramenopiles and alveolates (with the exception of ciliates) have a plastid derived from a single endosymbiotic event with a red algae in their common ancestor (Cavalier-Smith 1999; Harper et al. 2005). A putative sister relationship between Rhizaria and stramenopiles would complicate the situation suggesting that either stramenopiles have acquired their secondary plastid in an independent event of endosymbiosis or the single engulfment of a red algae occurred in a very early stage of chromalveolates evolution and the resulting plastid was secondarily lost in certain lineages, such as ciliates and Rhizaria. Although such a scenario is certainly less parsimonious than the chromalveolates or cabozoan hypotheses, none of them are actually strongly supported by multigenic data.

The uncertainty concerning the phylogenetic position of Rhizaria reflects the general difficulties in resolving the phylogeny in this part of the eukaryotic tree. Except for plants, whose position seems to be well established, the relations between all other groups of bikonts remained unresolved. This is not surprising given that even the analyses of larger data sets, with more than 100 proteins, failed to properly resolve the phylogeny of bikonts (Bapteste et al. 2002). For example, chromalveolates were strongly supported in multigene phylogenies only when no other unicellular bikonts were present in the analyses (Rodriguez-Ezpeleta et al. 2005) and other phylogenetic analyses provided only mixed support for this plastid-based view of eukaryotic relationships (Yoon et al. 2002, 2004). Despite this lack of clear support, the union of chromalveolate taxa has been potentially confirmed by the existence of a gene replacement in which the cytosolic GAPDH gene was duplicated and retargeted to the plastid uniquely in these taxa (Fast et al. 2001; Harper and Keeling 2003). Nevertheless, none of these studies was directly concerned by the overall phylogeny of bikonts, which resulted in a relatively limited taxon sampling of unicellular bikonts and a lack of detailed analysis of their relationships. By adding Rhizaria and all available sequence data on stramenopiles, alveolates, and excavates, we included in our analyses all major bikont phyla, except haptophytes, cryptophytes, and centrohelids. However, even with such exhaustive sampling, we were unable to resolve the relationships between these taxa.

The obvious question is why multigene analyses cannot reliably resolve the phylogeny of unicellular bikonts? It has been proposed that this lack of resolution observed in other EST-based phylogenies is due to the mutational saturation, phylogenetic incongruence, or rapid diversification (Philippe et al. 2004). Indeed, it has been demonstrated by single-gene phylogenies that some excavates (Philippe et al. 2000), foraminifers (Pawlowski et al. 1996), and ciliates (Philippe and Adoutte 1998) can evolve exceptionally rapidly, and it cannot be excluded that most part of these genomes show accelerate rates of evolution. In our trees, this is particularly well illustrated by the case of ciliates (*Tetrahymena* + *Paramecium*). Although there are several evidences that ciliates share a common ancestor with apicomplexans and dinoflagellates (Cavalier-Smith 1993; Fast et al. 2002; Leander and Keeling 2003, 2004), in our analyses, they often branch as sister group to excavates (fig. 2*D*), but this branching is systematically rejected by the AU test, suggesting an artifactual position.

The accelerated rates of evolution in some unicellular bikonts, which potentially erase the phylogenetic signal, are probably the main source of problems when inferring their evolutionary relationships. However, other possible causes cannot be discarded. One of them could be the rapid diversification of eukaryotes, suggested by some authors (Cavalier-Smith 2002). In fact, the lack of resolution in early animal phylogeny compared with the well-resolved phylogeny of fungi (observed also in our data, see fig. 1) has been interpreted as an indirect evidence for Cambrian explosion (Rokas et al. 2005). However, it is not clear why such rapid diversification would occur in the unicellular bikonts but not in other eukaryotes. Alternatively, it may be that the position of the root for the eukaryotic tree between unikonts and bikonts, principally based on a single genomic fusion (Stechmann and Cavalier-Smith 2002), is not correct. Some authors indeed suggest that this root could rather be on the branch leading to Opisthokonts or to the common ancestor of diplomonads/parabasalids (Arisue et al. 2005). If this is true, then the unicellular bikonts would be paraphyletic and their phylogeny will be particularly difficult to resolve.

To conclude, resolving the phylogeny of bikonts will probably require several additional efforts. As illustrated by our study, the addition of new higher-level taxa, such as Rhizaria, is not sufficient but may help to solidify the relationships within particular supergroups. It is doubtful whether better resolution can be achieved only by increasing the number of analyzed genes (more EST or whole-genome data). In fact, the analysis of selected slowly evolving genes may be more informative than the analysis of large databases, as it has been shown in case of chromalveolates (Harper et al. 2005). Also, searching for new genomic signatures may be an essential complement to multigene analyses. Finally, proper rooting of the eukaryotic tree will be crucial for an accurate interpretation of the relationships between unicellular bikonts and a better understanding of the deep phylogeny of eukarytotes.

Supplementary Material

Supplementary Tables S1 and S2 and Figures S1, S2, S3, S4, and S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors would like to thank José Fahrni, Jackie Guiard, Ignacio Bolivar, and Sergey Nikolaev for technical assistance and constructive discussions; Juan Montoya and Naiara Rodriguez-Ezpeleta for useful suggestions; Bruno Nyffeler and Jacques Rougemont for help with the Vital-IT server; Kamran Shalchian-Tabrizi for help with the Bioportal server; and Vassilios Ioannidis for help with programming. This research was supported by the Swiss National Science Foundation grant 3100A0-100415 and 3100A0-112645 (J.P.).

Literature Cited

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104–5.

Adl SM, Simpson AGB, Farmer MA, et al. (28 co-authors). 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol 52: 399–451.

Archibald JM, Longet D, Pawlowski J, Keeling PJ. 2003. A novel polyubiquitin structure in Cercozoa and Foraminifera: evidence for a new eukaryotic supergroup. Mol Biol Evol 20:62–6.

Archibald JM, Rogers MB, Toop M, Ishida KI, Keeling PJ. 2003. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga Bigelowiella natans. Proc Natl Acad Sci 100:7678–83.

Arisue N, Hasegawa M, Hashimoto T. 2005. Root of the Eukaryota tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. Mol Biol Evol 22:409–20.

Baldauf SL, Palmer JD. 1993. Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. Proc Natl Acad Sci USA 90:11558–62.

Baldauf SL, Roger AJ, Wenk-Siefert I, Doolitle WF. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290:972–7.

Bapteste E, Brinkmann H, Lee JA, et al. (11 co-authors). 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. Proc Natl Acad Sci USA 99:1414–9.

Bass D, Moreira D, Purificacion L, Polet S, Chao E, von der Heyden S, Pawlowski J, Cavalier-Smith T. 2005. Polyubiquitin insertions and the phylogeny of Cercotoa and Rhizaria. Protist 156:149–61.

Berney C, Pawlowski J. 2003. Revised small subunit rRNA analysis provides further evidence that Foraminifera are related to Cercozoa. J Mol Evol 57:S120–7.

Bhattacharya D, Helmchen T, Melkonian M. 1995. Molecular evolutionary analyses of nuclear-encoded small subunit ribosomal RNA identify an independent Rhizopod lineage containing the Euglyphida and the Chlorarachniophyta. J Eukaryot Microbiol 42:65–9.

Brinkmann H, Giezen M, Zhou Y, Poncelin de Raucourt G, Philippe H. 2005. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. Syst Biol 54:743–57.

Burki F, Berney C, Pawlowski J. 2002. Phylogenetic position of Gromia oviformis Dujardin inferred from nuclear-encoded small subunit ribosomal DNA. Protist 153:251–60.

Burki F, Nikolaev S, Bolivar I, Guiard J, Pawlowski J. 2006. Analysis of expressed sequence tags (ESTs) from a naked foraminiferan Reticulomyxa filosa. Genome. Forthcoming.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540–52.

Cavalier-Smith T. 1993. Kingdom protozoa and its 18 phyla. Microbiol Rev 57:953–94.

Cavalier-Smith T. 1998. A revised six-kingdom system of life. Biol Rev Camb Philos Soc 73:203–66.

Cavalier-Smith T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. J Eukaryot Microbiol 46:347–66.

Cavalier-Smith T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int J Syst Evol Microbiol 52:297–354.

Cavalier-Smith T, Chao EE. 1997. Sarcomonad ribosomal RNA sequences, rhizopod phylogeny and the origin of euglyphid amoebae. Arch Protistenkd 147:227–36.

Cavalier-Smith T, Chao EE. 2003. Phylogeny and classification of phylum Cercozoa (Protozoa). Protist 154:341–58.

de Koning A, Tartar A, Boucias D, Keeling P. 2005. Expressed sequence tag (EST) survey of the highly adapted green algal parasite, Helicosporidium. Protist 156:181–90.

Fast NM, Kissinger JC, Roos DS, Keeling PJ. 2001. Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. Mol Biol Evol 18:418–26.

Fast NM, Xue L, Bingham S, Keeling P. 2002. Re-examining alveolate evolution using multiple protein molecular phylogenies. J Eukaryot Microbiol 49:30–7.

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27:401–10.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 40:783–91.

Felsenstein J. 2004. PHYLIP (phylogeny inference package). Version 3.6. Seattle, WA: University of Washington.

Flakowski J, Bolivar I, Fahrni J, Pawlowski J. 2005. Actin phylogeny of Foraminifera. J Foraminifer Res 35:93–102.

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

Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–8.

Hampl V, Horner DS, Dyal P, Kulda J, Flegr J, Foster PG, Embley TM. 2005. Inference of the phylogenetic position of oxymonads based on nine genes: support for Metamonada and Excavata. Mol Biol Evol 22:2508–18.

Harper JT, Keeling PJ. 2003. Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for the chromalveolate plastids. Mol Biol Evol 20:1730–5.

Harper JT, Waanders E, Keeling PJ. 2005. On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. Int J Syst Evol Microbiol 55:487–96.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? Trends Genet 22:225–31.

Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol 4:18.

Keeling PJ. 2001. Foraminifera and Cercozoa are related in actin phylogeny: two orphans find a home? Mol Biol Evol 18: 1551–7.

Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE, Roger AJ, Gray MW. 2005. The tree of eukaryotes. Trends Ecol Evol 20:670–6.

Keeling P, Palmer J. 2001. Lateral transfer at the gene and subgenic levels in the evolution of eukaryotic enolase. Proc Natl Acad Sci USA 98:10745–50.
Kumar S, Rzhetsky A. 1996. Evolutionary relationships of eukaryotic kingdoms. J Mol Evol 42:183–93.
Leander B, Keeling P. 2003. Morphostasis in alveolate evolution. Trends Ecol Evol 18:498–504.
Leander B, Keeling P. 2004. Early evolutionary history of dinoflagellates and apicomplexans (Alveolata) as inferred from HSP90 and actin phylogenies. J Phycol 40:341–50.
Longet D, Archibald JM, Keeling PJ, Pawlowski J. 2003. Foraminifera and Cercozoa share a common origin according to RNA polymerase II phylogenies. Int J Syst Evol Microbiol 53:1735–9.
Longet D, Burki F, Flakowski J, Berney C, Polet S, Fahrni J, Pawlowski J. 2004. Multigene evidence for close evolutionary relations between Gromia and Foraminifera. Acta Protozool 43:303–11.
Moreira D, Le Guyader H, Philippe H. 1999. Unusually high evolutionary rate of the elongation factor 1 alpha genes from the Ciliophora and its impact on the phylogeny of eukaryotes. Mol Biol Evol 16:234–45.
Morin L. 2000. Long branch attraction effects and the status of "basal eukaryotes": phylogeny and structural analysis of the ribosomal RNA gene cluster of the free-living diplomonad Trepomonas agilis. J Eukaryot Microbiol 47:167–77.
Nikolaev SI, Berney C, Fahrni JF, Bolivar I, Polet S, Mylnikov AP, Aleshin VV, Petrov NB, Pawlowski J. 2004. The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. Proc Natl Acad Sci USA 101:8066–71.
Pawlowski J, Bolivar I, Fahrni J, Cavalier-Smith T, Gouy M. 1996. Early origin of foraminifera suggested by SSU rRNA gene sequences. Mol Biol Evol 13:445–50.
Philippe H. 2000. Opinion: long branch attraction and protists phylogeny. Protist 151:307–16.
Philippe H, Adoutte A. 1998. The molecular phylogeny of Eukaryote: solid facts and uncertainties. In: Coombs G, Vickerman K, Sleigh M, Warren A, editors. Evolutionary relationships among Protozoa. London: Chapman and Hall. p 25–56.
Philippe H, Germot A. 2000. Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution. Mol Biol Evol 17:830–4.
Philippe H, Lartillot N, Brinkmann H. 2005. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. Mol Biol Evol 22:1246–53.
Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J, Moreira D, Muller M, Le Guyader H. 2000. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc R Soc Lond B Biol Sci 267:1213–21.
Philippe H, Snell EA, Bapteste E, Lopez P, Holland PWH, Casane D. 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. Mol Biol Evol 21:1740–52.
Polet S, Berney C, Fahrni J, Pawlowski J. 2004. Small-subunit ribosomal RNA gene sequences of Phaeodarea challenge the monophyly of Haeckel's Radiolaria. Protist 155:53–63.
Posada D, Buckley T. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst Biol 53:793–808.
Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Loffelhardt W, Bohnert HJ, Philippe H, Lang BF. 2005. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr Biol 15:1325–30.
Rogers MB, Archibald JM, Field M, Li C, Strieped B, Keeling PJ. 2004. Plastid-targeting peptides from the Chlorarachniophyte Bigelowiella natans. J Eukaryot Microbiol 51:529–35.
Rokas A, Kruger D, Caroll SB. 2005. Animal evolution and the molecular signature of radiations compressed in time. Science 310:1933–8.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–4.
Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol 51:492–508.
Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246–7.
Simpson AGB, Inagaki Y, Roger AJ. 2006. Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. Mol Biol Evol 23:615–25.
Simpson AGB, Radek R, Dacks J, O'Kelly C. 2002. How oxymonads lost their groove: an ultrastructural comparison of Monocercomonoides and excavate taxa. J Eukaryot Microbiol 49:239–48.
Simpson AGB, Roger AJ, Silberman JD, Leipe DD, Edgcomb VP, Jermiin LS, Patterson DJ, Sogin ML. 2002. Evolutionary history of "early-diverging" eukaryotes: the excavate taxon Carpediemonas is a close relative of Giardia. Mol Biol Evol 19:1782–91.
Sogin M. 1991. Early evolution and the origin of eukaryotes. Curr Opin Genet Dev 1:457–63.
Sogin ML, Gunderson J, Elwood H, Alonso R, Peattie D. 1989. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from Giardia lamblia. Science 243:75–7.
Sogin ML, Silberman JD. 1998. Evolution of the protists and protistan parasites from the perspective of molecular systematics. Int J Parasitol 28:11–20.
Stechmann A, Cavalier-Smith T. 2002. Rooting the eukaryote tree by using a derived gene fusion. Science 297:89–91.
Stechmann A, Cavalier-Smith T. 2003. The root of the eukaryote tree pinpointed. Curr Biol 13:R665–6.
Steenkamp ET, Wright J, Baldauf SL. 2006. The protistan origins of animals and fungi. Mol Biol Evol 23:93–106.
Stiller JW, Hall BD. 1999. Long-branch attraction and the rDNA model of early eukaryotic evolution. Mol Biol Evol 16: 1270–9.
Thompson J, Higgins D, Gibson T. 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 22:4673–80.
Yamamoto A, Hashimoto T, Asaga E, Hasegawa M, Goto N. 1997. Phylogenetic position of the mitochondrion-lacking protozoan Trichomonas tenax, based on amino acid sequences of elongation factors 1. J Mol Evol 44:98–105.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13: 555–6.
Yoon HS, Hackett JD, Pinto G, Bhattacharya D. 2002. The single, ancient origin of chromist plastids. Proc Natl Acad Sci USA 99:15507–12.
Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol 21:809–18.

Martin Embley, Associate Editor

Accepted July 4, 2006