Report

Toward Resolving the Eukaryotic Tree: The Phylogenetic Positions of Jakobids and Cercozoans

Naiara Rodríguez-Ezpeleta,[1] Henner Brinkmann,[1]
Gertraud Burger,[1] Andrew J. Roger,[2] Michael W. Gray,[2]
Hervé Philippe,[1,*] and B. Franz Lang[1,*]
[1] Centre Robert Cedergren
Département de Biochimie
Université de Montréal
2900 Boulevard Édouard-Montpetit
Montréal, Québec H3T 1J4
Canada
[2] Department of Biochemistry and Molecular Biology
Dalhousie University
5850 College Street
Halifax, Nova Scotia B3H 1X5
Canada

Summary

Resolving the global phylogeny of eukaryotes has proven to be challenging. Among the eukaryotic groups of uncertain phylogenetic position are jakobids, a group of bacterivorous flagellates that possess the most bacteria-like mitochondrial genomes known [1, 2]. Jakobids share several ultrastructural features with malawimonads and an assemblage of anaerobic protists (e.g., diplomonads and oxymonads) [3, 4]. These lineages together with Euglenozoa and Heterolobosea have collectively been designated "excavates" [5]. However, published molecular phylogenies based on the sequences of nuclear rRNAs [5–7] and up to six nucleus-encoded proteins [8–10] do not provide convincing support for the monophyly of excavates, nor do they uncover their relationship to other major eukaryotic groups [5–10]. Here, we report the first large-scale eukaryotic phylogeny, inferred from 143 nucleus-encoded proteins comprising 31,604 amino acid positions, that includes jakobids, malawimonads and cercozoans [7]. We obtain compelling support for the monophyly of jakobids, Euglenozoa plus Heterolobosea (JEH group), and for the association of cercozoans with stramenopiles plus alveolates. Furthermore, we observe a sister-group relationship between the JEH group and malawimonads after removing fast-evolving species from the dataset. We discuss the implications of these results for the concept of "excavates" and for the elucidation of eukaryotic phylogeny in general.

Results and Discussion

Phylogenetic Analyses with the Complete Dataset

As originally proposed, "excavates" unite five unicellular eukaryote taxa: retortamonads, *Carpediemonas*, *Trimastix*, jakobids, and malawimonads [5]. This circumscription was based on the presence of ultrastructural characters such as a ventral feeding groove, flagellar vanes, and a few other cytoskeletal elements. Later, the group was expanded to include three more taxa: diplomonads (e.g., *Giardia*) and heteroboloseids (e.g., *Naegleria*), both of which possess a feeding groove but lack flagellar vanes, and oxymonads, which lack a feeding groove but possess most other features that define the initial excavate classification [5]. When the monophyly of excavates was tested by molecular phylogenetics, the results lacked coherence: Some typical excavates (e.g., malawimonads) did not cluster with the group, whereas euglenozoans (e.g., *Trypanosoma*) and parabasalids (e.g., Trichomonas) joined as potential new members, although neither of the latter taxa exhibit the distinctive excavate ultrastructure [6, 8, 9, 11]. Notably, key branches in these phylogenies drew only weak statistical support (bootstrap values [BV], much below 95%; e.g., [5, 10, 12]). Potential reasons for the lack of resolution and the observed inconsistencies are (1) the quantity of sequence data (at most six genes), which has proven insufficient to resolve most deep phylogenies, (2) inclusion of data from extremely fast-evolving parasitic species (*Trichomonas vaginalis* and *Giardia lamblia*), which are prone to systematic error such as long-branch attraction (LBA) [13, 14], and (3) the inadvertent use of paralogous rather than orthologous genes.

To clarify the phylogenetic position of jakobids and malawimonads, we sequenced a total of ~30,000 ESTs from five jakobids and two malawimonads. In addition, we included new EST data from two heteroloboseids that are expected to belong to excavates and from the cercomonad *Cercomonas longicauda*, which, together with the chlorarachniophyte *Bigelowiella natans* represents a major eukaryotic lineage ("cercozoans") of similarly uncertain phylogenetic affiliation. Figure 1 shows the maximum likelihood (ML) tree inferred from the concatenation of 143 nucleus-encoded protein sequences (31,604 amino acid positions), including representatives of all major eukaryotic super-groups (Opisthokonta, Amoebozoa, Plantae, Stramenopila + Alveolata, Excavata, and Rhizaria; see [15]). The tree, which was rooted on the basis of a gene fusion [13, 16], has 100% bootstrap support value (BV) for nearly all branches. This analysis confirms the monophyly of Holozoa (animals plus choanoflagellates), Fungi, Amoebozoa, Viridiplantae, Rhodophyta, Glaucophyta, Alveolata (apicomplexans, dinoflagellates plus ciliates), Stramenopila, Cercozoa, Malawimonadozoa, Heterolobosea, Euglenozoa, and Jakobozoa ("core" jakobids). We also confirm the higher-order relationships Opisthokonta (animals and choanoflagellates plus Fungi; 100% BV) and Plantae (the primary photosynthetic red algae, green plants, and glaucophytes; 84% BV). A third, new super-ensemble unites Alveolata and Stramenopila plus Cercozoa (the latter represented by *Bigelowiella* and *Cercomonas*; 100% BV), as also suggested in [17].

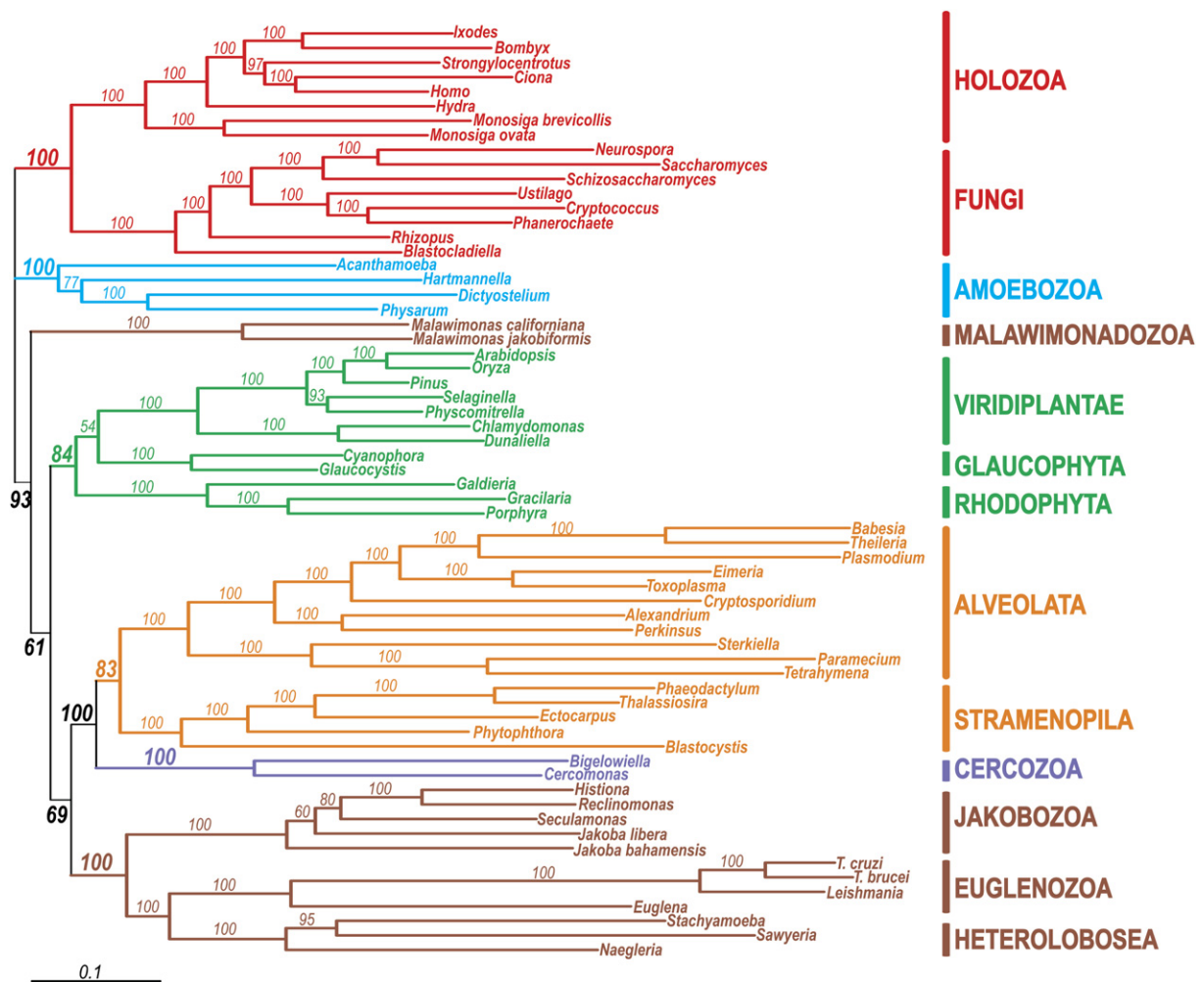*Correspondence: herve.philippe@umontreal.ca (H.P.), franz.lang@umontreal.ca (B.F.L.)

Figure 1. Maximum-Likelihood Tree of Eukaryotes

The tree includes 64 species and is based on 143 concatenated nucleus-encoded proteins (31,604 amino acid positions). Numbers indicate support values of RaxML analysis (100 replicates) with the WAG + F + Γ model. Posterior probabilities obtained in the Bayesian Inference with MrBayes are 1.0 for all branches. The scale bar denotes the estimated number of amino acid substitutions per site. The tree was rooted according to a gene fusion [13, 16].

Finally, Jakobozoa, Euglenozoa, and Heterolobosea (henceforth referred to as the JEH group) form a monophyletic group (100% BV). The coherence of the JEH group is further supported by a unique insertion in their large-subunit ribosomal protein 24A (Figure 2), an insertion that is absent in the orthologous proteins from other eukaryotes and from Archaea.

Despite the use of a large number of aligned amino acid positions, the phylogenetic affiliation of malwimonads remains uncertain, branching with nonsignificant support (defined here as <95% BV and failure to pass the AU test [18]) at the base of a "bikont" group of species (i.e., Plantae, alveolates, stramenopiles, Cercozoa, and the JEH group; Figure 1). Interestingly, the predominant alternative topology in bootstrap analyses unites the JEH group with malawimonads (30% BV) but, as expected, the AU test does not favor either of these two topologies (Table S1 in the Supplemental Data available online).

Phylogenetic Analyses excluding Fast-Evolving Species

It is well documented that fast-evolving taxa tend to mislead phylogenetic estimation through LBA [19, 20], in many cases inducing high bootstrap support for an incorrect tree topology. For instance, kinetoplastids are known to be indiscriminately attracted to other fast-evolving species in global eukaryote trees [21]. Within the alveolates, stramenopiles and the JEH group are a number of fast-evolving species that might cause LBA artifacts. To test this possibility, we explored removal of fast-evolving species within these groups.

From the taxa shown in Figure 1, we removed the stramenopile *Blastocystis*, the heterolobosean *Sawyeria*, euglenozoans except *Euglena*, and all alveolates except *Toxoplasma* and dinoflagellates. The resulting tree topology and support values remain essentially the same (Figure 3), the only exception being that the JEH group now branches with malawimonads (78% BV), as

	10 34
Archaeoglobus	GYDIEPGTGKMYV-----RRDGRVFYFCSG
Pyrococcus	GKPFEPGTGKMYV-----RNDGRVLFFCSG
Homo	GYKIYPGHGRRYT-----RTDGKVFQFLNA
Hydra	GYKIYPGHGKRYV-----RSDGKLFNFLSK
Monosiga	SYKIHAGHGRRLV-----RVDGKTFYFLGS
Rhizopus	GQKIYPAKGKTYV-----RIDSRTFRFING
Ustilago	QRKIYPGKGRLYV-----RGDNKVFRFVSS
Saccharomyces	GAKIYPGRGTLFV-----RGDSKIFRFQNS
Acanthamoeba	GYKIYPGHGRRYA-----RTDMKTFVFINA
Dictyostelium	EFKIYPARGMKFV-----RGDSKVFHFINT
Tryp. brucei	HFAVHPGHGRRYVP-FAFLSTKPVLTFARP
Tryp. cruzi	HFAVHPGHGRRYVP-FAFLSTKPVLTFARP
Leishmania	HFAVHPGHGRRYVP-FAFLSTKPVLTFSRP
Euglena	GLPVHPGHGKRFVP-TLVQSTRPVLTFVTA
Sawyeria	GVKIYPGHGLSYVPVVSVQATRPVFKFFDQ
Naegleria	GYKIYPGHGVRYVPCTNVQSTRPVFTFVSR
Reclinomonas	GFRIWPGHGIRYVPCVNMQSTKLVYPFINH
Histiona	GFRIFPGHGVRYVPCVNMQSIKLVYPFIIR
Seculamonas	GIKVYPGHGLRYVPTANMQSTKLVYPFLSR
Malawimonas	GFKIYPGHGRRFI-----RGDSKLFQFLNS
Giardia	GRKILPGYGKRMS-----RHDKVLLIFLNR
Spironucleus	GRQILPGYGKRFA-----KLDKSLVIFINR
Trichomonas	GHIFHAGHGRVHI-----REDKHLMAFESR
Cyanophora	GYKIYPGHGMKFV-----RADNRSFMFVSS
Porphyra	GFRIYPGHGSRFI-----RVDGKSYVFANS
Chlamydomonas	GLRIYPGKGMIFI-----RTDGQHYMFLNK
Arabidopsis	GQKIYPGRGIRFI-----RSDSQVFLFLNS
Bigelowiella	ECKVFPGHGIRFV-----RKDGKILTFLNR
Blastocystis	EYKIYPGHGGMYI-----RKDAQPVRYISR
Phytophthora	ESRIYPGHGSRFI-----RRDGSAYVFINS
Tetrahymena	EYRIYPGRGQRFI-----AKDGRGFFFLTK
Alexandrium	EYRIYPGSGQRFI-----AKDGKVSFFISK
Plasmodium	EYRIYPGRGQKYI-----ARDGKVYFYLSS
Cryptosporidium	EYRIYPGRGRKFV-----ARDGRVSTFLNQ

Figure 2. Amino Acid Insertion Specific to Jakobids, Euglenozoa, and Heterolobosea

A section of the amino acid sequence alignment of Rpl24A is shown. Numbers above the alignment indicate the sequence position of the *Homo* protein. Jakobids, Euglenozoa, and Heterolobosea are highlighted with different gray shades. Because of space constraints, only one or two representative species per group are shown. The complete alignment is available upon request.

a sister clade (96%) of Plantae/Alveolata + Stramenopila + Cercozoa. Yet, alternatives to the JEH + malawimonad grouping are not rejected by the AU test (Table S2). A reliable resolution of these branches will require data from additional slow-evolving sister taxa of malawimonads and the JEH group. The same reservation applies in the case of the extremely fast-evolving diplomonads and parabasalids (Figure S1).

Rooting the Eukaryotic Tree
The phylogenetic trees presented here are unrooted. In principle, archaeal sequences might serve as an obvious outgroup for the rooting of eukaryotes, but these sequences proved to be too distant (data not shown). Alternatively, several bacteria-like features exclusive to the mitochondrial DNA (mtDNA) of jakobids might suggest a root basal to this group. For example, the jakobid mtDNA codes for more genes than any other eukaryote [1], its protein-coding genes exhibit Shine-Dalgarno-like motifs for translation initiation of the corresponding mRNAs, and its encoded rRNA and RNase P RNA secondary structures strikingly resemble those of bacteria [1, 22]. However, although these features apparently derive from the bacterial ancestor of mitochondria, they may have been independently lost on a number of occasions in other eukaryotic groups and thus cannot be used to infer the eukaryotic root.

A character that does suggest a root basal to jakobids is the type of RNA polymerase employed for transcription in mitochondria. Jakobid mtDNAs encode the four subunits (RpoA-D) of a bacteria-like $\alpha_2\beta\beta'\sigma$ RNA polymerase, whereas all other eukaryotes studied to date utilize a nucleus-encoded "T3/T7 phage-type" enzyme instead [23]. Evidently, this enzyme has replaced the mtDNA-encoded *rpo* genes that originated from the bacterial ancestor of mitochondria [1]. A eukaryotic root basal to jakobids is nevertheless contradicted by another rare character in our dataset: an insertion that is in the Rpl24A protein and that is present in species of the JEH group but absent in all other eukaryotes and Archaea (Figure 2). Assuming that this insertion has been gained only once, the root of eukaryotes would be placed prior to the divergence of jakobids, Euglenozoa, and Heterolobosea, (not basal to jakobids). The conflict created by this inconsistency is further amplified by evidence from other rare genetic characters, suggesting that the root lies between Opisthokonts + Amoebozoa and all other major eukaryotic groups [24, 25]. Each of these mutually exclusive rooting scenarios relies on the (most parsimonious) interpretation of one or a few supposedly rare genetic changes; however, genomic changes are as prone to homoplasy as sequence characters (via convergence or reversion) [26]. For example, the bacterial-type or the T3/T7 phage-type mitochondrial RNA polymerases may have been differentially lost in various lineages, after having coexisted over a prolonged period. Indeed, chloroplasts of extant land plants utilize both bacteria-type (chloroplast DNA-encoded) and phage-type (nucleus-encoded) RNA polymerases [27, 28]. Thus, for non-sequence-based characters to be useful for the inference of deep phylogenetic relationships, a much larger number of congruent (preferably complex) characters would be required. In the absence of compelling information of this sort, the rooting of the eukaryotic tree remains an open question.

Toward a Global Eukaryotic Tree
The analysis presented here is the first phylogenomic study to include members of all six proposed eukaryotic super-groups [15]: Opisthokonta, Amoebozoa, Plantae, Chromalveolata (here represented by Alveolata and Stramenopila), Rhizaria (here represented by Cercozoa), and Excavata. This analysis confirms the monophyly of three supergroups: Opisthokonta, Amoebozoa, and Plantae (the latter with less support). Moreover, we find a strong affiliation between Stramenopila + Alveolata + Cercozoa, an affiliation that was also recovered in a recent analysis based on only 16 proteins [17] (the latter study further included haptophytes and cryptophytes but did not position them with confidence). In light of these findings, the chromalveolate hypothesis (i.e., the grouping of alveolates, cryptophytes, haptophytes, and stramenopiles [15]) remains questionable and, even if demonstrated in future studies, would have to be reformulated to include Cercozoa. It also needs to be seen whether Rhizaria will remain monophyletic, once data from other rhizarians become available [15]. Finally, three excavate taxa (jakobids, Euglenozoa,
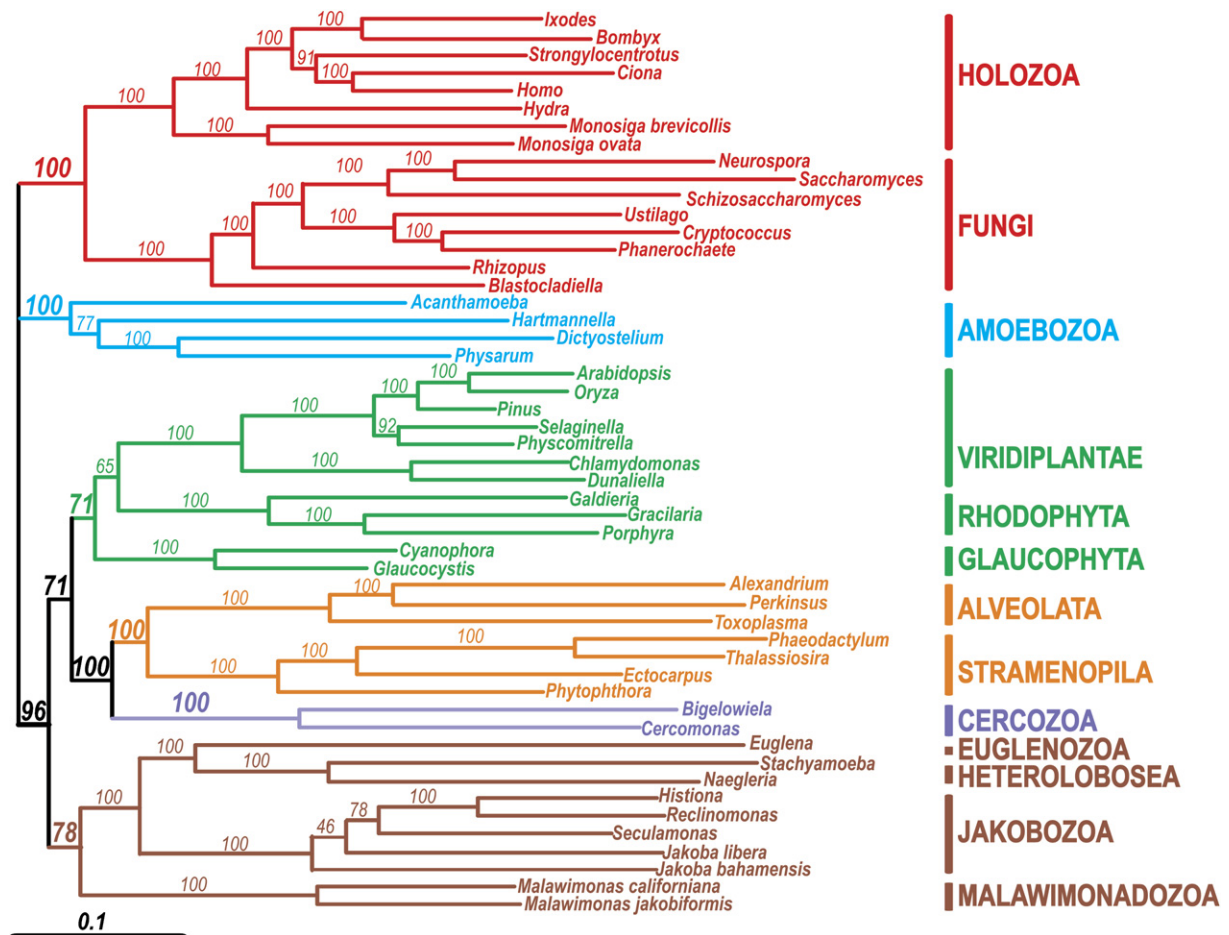
Figure 3. Maximum-Likelihood Tree for Slow-Evolving Eukaryotic Species

The analysis was performed as shown in Figure 1, except that taxon sampling was restricted to slowly evolving alveolates, Euglenozoa, Heterolobosea, and stramenopiles.

and Heterolobosea) form a convincingly supported monophyletic group, and the fourth one (malawimonads) appears to be weakly associated. Still, the grouping with the other six proposed excavate groups remains to be demonstrated once sufficient data become available.

In conclusion, we demonstrate here that EST projects targeting poorly studied protist groups are an effective means to generate data for robust phylogenomic analyses. In turn, these results permit selection of early diverging and slowly evolving taxa for complete genome sequencing, providing an ever-widening window on early eukaryotic evolution.

Experimental Procedures

Construction of cDNA Libraries and EST Sequencing

We generated cDNA libraries from five jakobids (*Reclinomonas americana*, *Jakoba libera*, *Jakoba bahamensis*, *Histiona aroides*, and *Seculamonas ecuadoriensis*), two malawimonads (*Malawimonas californiana* and *Malawimonas jakobiformis*), a Cercozoa (*Cercomonas longicauda*), and two Heterolobosea (*Stachyamoeba lipophora* and *Sawyeria marylandensis*) as described [29]. Plasmids were purified with the QIAprep 96 Turbo Miniprep Kit (QIAGEN), and sequencing reactions were performed with the ABI Prism BigDyeTM Terminators version 3.0/3.1 (Perkin-Elmer, Wellesley, MA) and sequenced on an MJ BaseStation. Trace files were imported into

the TBestDB database (http://tbestdb.bcm.umontreal.ca/searches/login.php) [30] for automated processing including assembly as well as automated annotation by AutoFact [31]. Table S3 lists details about the number of ESTs and clusters obtained for each species.

Dataset Construction

Data from jakobids, malawimonads, *Cercomonas*, *Sawyeria*, and *Stachyamoeba*, and additional sequences from GenBank (http://www.ncbi.nlm.nih.gov/) were added to an existing multiple alignment of protein sequences [32] as described earlier [33–35]. To represent the putative chromalveolate super-group, we selected alveolates and stramenopiles because only they have a sufficiently rich sampled taxonomic diversity (including photosynthetic and nonphotosynthetic organisms) to allow efficient detection of possible endosymbiotic gene transfers (see below). The selection of species, genes, and orthologous sequences was performed with SCaFoS [36]. In brief, we chose (1) species that represent all major eukaryotic groups for which genomic data are available from more than one member, (2) genes that are present in at least 24 of the selected species, and (3) among these, species exhibiting the most slowly evolving orthologous sequences. The latter sequences were identified as described in [37].

Species evolving at accelerated rates are known to induce tree-reconstruction artifacts. Therefore, fast-evolving taxa were excluded when more slowly evolving relatives were available (e.g., the red alga *Cyanidioschyzon merolae* [21]). In the case of fungi, animals, and embryophyte plants, only representative (preferentially slowly evolving) members were used. Chimaeras were constructed in some instances either to increase the number of sequence positions or to obtain slow-evolving proteins (see [36]): *Homo*

(*H. sapiens*, *Mus musculus*, and *Canis familiaris*), *Theileria* (*T. annulata* and *T. parva*), Florideophyceae (*Gracilaria changii* and *Chondrus crispus*). The dataset used contains 143 proteins (31,604 amino acid positions). Overall, 23% of the theoretical total number of amino acids in the alignment were unavailable (resulted in missing data). Tables S4 and S5 list the 143 proteins as well as the distribution of the missing data across the selected genes and species.

Phylogenetic Analyses

We used protein sequences for phylogenetic analyses. Maximum likelihood (ML) and Bayesian analyses were performed with RaxML v.VI-HPC [38] and MrBayes v.3.1 [39], respectively, with the WAG amino acid replacement matrix, stationary amino acid frequencies estimated from the dataset, and four categories of gamma-distributed rates across sites (WAG + F + Γ4). Statistical support was evaluated on the basis of the ML analyses by 100 bootstrap replicates. Three independent RaxML analyses were performed, each with a different starting tree (maximum parsimony, Figures 1 and 3), because for genome-scale datasets the probability is high of being trapped in a local maximum [40]. The tree with the best log likelihood was selected for each replicate, and the 100 trees obtained were used to compute bootstrap proportions.

The selected 143 proteins were used for single-gene and concatenated phylogenies. In single-gene phylogenies, 75 genes supported (with BV >70%) one or more groupings that conflict with the tree obtained from the concatenated protein sequences, potentially indicating undetected horizontal gene transfer or paralogy. However, a concatenated tree generated from these 75 proteins (data not shown) is in agreement with the result of the concatenation of all 143 proteins, assuring the high quality of our dataset.

Supplemental Data

One figure and five tables are available at http://www.current-biology.com/cgi/content/full/17/16/1420/DC1/.

References

1. Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M., and Gray, M.W. (1997). An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature *387*, 493–497.
2. Gray, M.W., Lang, B.F., and Burger, G. (2004). Mitochondria of protists. Annu. Rev. Genet. *38*, 477–524.
3. O'Kelly, C.J. (1993). The jakobid flagellates: Structural features of *Jakoba*, *Reclinomonas* and *Histiona* and implications for the early diversification of eukaryotes. J. Eukaryot. Microbiol. *40*, 627–636.
4. O'Kelly, C.J., and Nerad, T.A. (1999). *Malawimonas jakobiformis* n. gen., n. sp. (Malawimonadidae n. fam): A *Jakoba*-like heterotrophic nanoflagellate with discoidal mitochondrial cristae. J. Eukaryot. Microbiol. *46*, 522–531.
5. Simpson, A.G. (2003). Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). Int. J. Syst. Evol. Microbiol. *53*, 1759–1777.
6. Simpson, A.G., Roger, A.J., Silberman, J.D., Leipe, D.D., Edgcomb, V.P., Jermiin, L.S., Patterson, D.J., and Sogin, M.L. (2002). Evolutionary history of "early-diverging" eukaryotes: The excavate taxon *Carpediemonas* is a close relative of *Giardia*. Mol. Biol. Evol. *19*, 1782–1791.
7. Cavalier-Smith, T. (2004). Only six kingdoms of life. Proc. Biol. Sci. *271*, 1251–1262.
8. Archibald, J.M., O'Kelly, C.J., and Doolittle, W.F. (2002). The chaperonin genes of jakobid and jakobid-like flagellates: Implications for eukaryotic evolution. Mol. Biol. Evol. *19*, 422–431.
9. Simpson, A.G., Inagaki, Y., and Roger, A.J. (2006). Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. Mol. Biol. Evol. *23*, 615–625.
10. Edgcomb, V.P., Roger, A.J., Simpson, A.G., Kysela, D.T., and Sogin, M.L. (2001). Evolutionary relationships among "jakobid" flagellates as indicated by alpha- and beta-tubulin phylogenies. Mol. Biol. Evol. *18*, 514–522.
11. Cavalier-Smith, T. (2003). The excavate protozoan phyla Metamonada Grasse emend. (*Anaeromonadea, Parabasalia, Carpediemonas, Eopharyngia*) and Loukozoa emend. (*Jakobea, Malawimonas*): Their evolutionary affinities and new higher taxa. Int. J. Syst. Evol. Microbiol. *53*, 1741–1758.
12. Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., and Doolittle, W.F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. Science *290*, 972–977.
13. Philippe, H., Lopez, P., Brinkmann, H., Budin, K., Germot, A., Laurent, J., Moreira, D., Muller, M., and 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*, 1213–1221.
14. Embley, T.M., and Hirt, R.P. (1998). Early branching eukaryotes? Curr. Opin. Genet. Dev. *8*, 624–629.
15. Keeling, P., Burger, G., Durnford, D., Lang, B., Lee, R., Pearlman, R., Roger, A., and Gray, M. (2006). The tree of eukaryotes. Trends Ecol. Evol. *20*, 670–676.
16. Stechmann, A., and Cavalier-Smith, T. (2002). Rooting the eukaryote tree by using a derived gene fusion. Science *297*, 89–91.
17. Hackett, J., Yoon, H., Li, S., Reyes-Prieto, A., Rümmele, S., and Bhattacharya, D. (2007). Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of 'Rhizaria' with chromalveolates. Mol. Biol. Evol., in press.
18. Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. Syst. Biol. *51*, 492–508.
19. Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. *27*, 401–410.
20. Brinkmann, H., van der Giezen, M., Zhou, Y., Poncelin de Raucourt, G., and Philippe, H. (2005). An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. Syst. Biol. *54*, 743–757.
21. Rodríguez-Ezpeleta, N., Brinkmann, H., Roure, B., Lartillot, N., Lang, B.F., and Philippe, H. (2007). Detecting and overcoming systematic errors in genome-scale phylogenies. Syst. Biol. *56*, 1–11.
22. Seif, E., Cadieux, A., and Lang, B.F. (2006). Hybrid *E. coli*–mitochondrial ribonuclease P RNAs are catalytically active. RNA *12*, 1661–1670.
23. Cermakian, N., Ikeda, T.M., Cedergren, R., and Gray, M.W. (1996). Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. Nucleic Acids Res. *24*, 648–654.
24. Stechmann, A., and Cavalier-Smith, T. (2003). The root of the eukaryote tree pinpointed. Curr. Biol. *13*, R665–R666.
25. Richards, T., and Cavalier-Smith, T. (2005). Myosin domain evolution and the primary divergence of eukaryotes. Nature *436*, 1113–1118.
26. Bapteste, E., and Philippe, H. (2002). The potential value of indels as phylogenetic markers: Position of trichomonads as a case study. Mol. Biol. Evol. *19*, 972–977.

27. Gray, M.W., and Lang, B.F. (1998). Transcription in chloroplasts and mitochondria: A tale of two polymerases. Trends Microbiol. *6*, 1–3.

28. Richter, U., Kiessling, J., Hedtke, B., Decker, E., Reski, R., Borner, T., and Weihe, A. (2002). Two RpoT genes of *Physcomitrella patens* encode phage-type RNA polymerases with dual targeting to mitochondria and plastids. Gene *290*, 95–105.

29. Rodríguez-Ezpeleta, N., Teijeiro, S., Forget, L., Burger, G., and Lang, B.F. (2007). Generation of cDNA libraries: Protists and fungi. In Methods in Molecular Biology: Methods in ESTs, J. Parkinson, ed. (Totowa, NJ: Humana Press).

30. O'Brien, E., Koski, L., Zhang, Y., Yang, L., Wang, E., Gray, M., Burger, G., and Lang, B. (2007). TBestDB: A taxonomically broad database of expressed sequence tags (ESTs). Nucleic Acids Res. *35*, D445–D451.

31. Koski, L.B., Gray, M.W., Lang, B.F., and Burger, G. (2005). AutoFACT: An automatic functional annotation and classification tool. BMC Bioinformatics *6*, 151.

32. Rodríguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Loffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. (2005). Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. Curr. Biol. *15*, 1325–1330.

33. Philippe, H., Snell, E.A., Bapteste, E., Lopez, P., Holland, P.W., and Casane, D. (2004). Phylogenomics of eukaryotes: Impact of missing data on large alignments. Mol. Biol. Evol. *21*, 1740–1752.

34. Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. *17*, 540–552.

35. Philippe, H. (1993). MUST, a computer package of management utilities for sequences and trees. Nucleic Acids Res. *21*, 5264–5272.

36. Roure, B., Rodríguez-Ezpeleta, N., and Philippe, H. (2007). SCaFoS: Selection, concatenation and fusion of sequences for phylogenomics. BMC Evol. Biol. *7* (Suppl. 1), S2.

37. Philippe, H., Lartillot, N., and Brinkmann, H. (2005). Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa and Protostomia. Mol. Biol. Evol. *22*, 1246–1253.

38. Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.

39. Ronquist, F., and Huelsenbeck, J. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics *19*, 1572–1574.

40. Salter, L.A. (2001). Complexity of the likelihood surface for a large DNA dataset. Syst. Biol. *50*, 970–978.