Genome Biology and Evolution

SMBE

LETTER

Large-Scale Phylogenomic Analyses Reveal That Two Enigmatic Protist Lineages, Telonemia and Centroheliozoa, Are Related to Photosynthetic Chromalveolates

Fabien Burki, Yuji Inagaki,† Jon Bråte,‡ John M. Archibald,§ Patrick J. Keeling,‖ Thomas Cavalier-Smith,¶ Miako Sakaguchi,† Tetsuo Hashimoto,† Ales Horak,‖ Surendra Kumar,‡ Dag Klaveness,‡ Kjetill S. Jakobsen,‡# Jan Pawlowski,* and Kamran Shalchian-Tabrizi‡*

*Department of Zoology and Animal biology, University of Geneva, Geneva, Switzerland; †Center for Computational Sciences, Institute for Biological Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan; ‡Microbial Evolution Research Group, Department of Biology, University of Oslo, Oslo, Norway; §Department of Biochemistry & Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada; ‖Botany Department, University of British Columbia, Vancouver, British Columbia, Canada; ¶Department of Zoology, University of Oxford, South Parks Road, Oxford, UK; and #Department of Biology, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway

Understanding the early evolution and diversification of eukaryotes relies on a fully resolved phylogenetic tree. In recent years, most eukaryotic diversity has been assigned to six putative supergroups, but the evolutionary origin of a few major "orphan" lineages remains elusive. Two ecologically important orphan groups are the heterotrophic Telonemia and Centroheliozoa. Telonemids have been proposed to be related to the photosynthetic cryptomonads or stramenopiles and centrohelids to haptophytes, but molecular phylogenies have failed to provide strong support for any phylogenetic hypothesis. Here, we investigate the origins of *Telonema subtilis* (a telonemid) and *Raphidiophrys contractilis* (a centrohelid) by large-scale 454 pyrosequencing of cDNA libraries and including new genomic data from two cryptomonads (*Guillardia theta* and *Plagioselmis nannoplanctica*) and a haptophyte (*Imantonia rotunda*). We demonstrate that 454 sequencing of cDNA libraries is a powerful and fast method of sampling a high proportion of protist genes, which can yield ample information for phylogenomic studies. Our phylogenetic analyses of 127 genes from 72 species indicate that telonemids and centrohelids are members of an emerging major group of eukaryotes also comprising cryptomonads and haptophytes. Furthermore, this group is possibly closely related to the SAR clade comprising stramenopiles (heterokonts), alveolates, and Rhizaria. Our results link two additional heterotrophic lineages to the predominantly photosynthetic chromalveolate supergroup, providing a new framework for interpreting the evolution of eukaryotic cell structures and the diversification of plastids.

Large-Scale Expressed Sequence Tag Sequencing of Telonemia and Centroheliozoa

The phylum Telonemia encompasses only two formally described heterotrophic zooflagellate species, *Telonema subtilis* and *Telonema antarcticum* (Klaveness et al. 2005), but a study of environmental sequences identified a large number of unknown representatives of this phylum in marine plankton (Shalchian-Tabrizi et al. 2007). Telonemids are of pivotal evolutionary significance because they exhibit a unique combination of cellular structures that have only been found separately in different eukaryotic lineages belonging to the chromalveolate supergroup, suggesting that they may represent a transitional form between deeply diverging eukaryotes (Shalchian-Tabrizi et al. 2006). Thus far, molecular support for the position of telonemids relative to other eukaryotes remains weak, with a 3-gene analysis suggesting a position close to plastid-bearing cryptomonads (Shalchian-Tabrizi et al. 2006) and a 6-gene study favoring an association with stramenopiles (Reeb et al. 2009).

Heliozoans, on the other hand, are a large and ultrastructurally diverse array of axopodia-bearing, largely heterotrophic protists that have recently been confirmed to be polyphyletic (Nikolaev et al. 2004; Sakaguchi et al. 2005). Although some small groups of heliozoans are now known to belong to Rhizaria or stramenopiles (heterokonts), the most distinctive core group of Heliozoa, the invariably heterotrophic Centroheliozoa, is the last substantial eukaryotic group not yet clearly placed on the tree. Based on weakly supported 18S ribosomal RNA trees and some intriguing ultrastructural similarities, a possible relationship between centrohelids and haptophytes was recently suggested (Cavalier-Smith and von der Heyden 2007), but molecular trees (even based on as many as seven genes) are generally inconsistent and inconclusive (Sakaguchi et al. 2007).

In contrast to single-gene phylogenies, phylogenomics has proven very useful for placing important species of uncertain affinity on the tree of eukaryotes (e.g., the breviate amoebae and *Ministeria*; Minge et al. 2009 and Shalchian-Tabrizi et al. 2008). We thus opted for a genomic scale approach to infer the phylogenetic position of the telonemids and centrohelids. In order to generate sufficient amounts of data for this purpose, we used massively parallel 454 pyrosequencing on normalized cDNA libraries from these poorly understood protists. A total of 213,350 and 363,490 sequence reads from *T. subtilis* and *R. contractilis* were obtained, respectively, with read lengths for both data sets averaging 232 bp. Of these, 184,838 (87%) could be assembled into 26,013 contigs for *T. subtilis* and 327,570 (90%) into 30,120 contigs for *R. contractilis* (cutoff for assembling: 100 bp). The average length of the contigs was 302 bp for *T. subtilis* (median = 233; standard

Key words: Telonemia, Centroheliozoa, SAR, CCTH, chromalveolates, plastid evolution.

E-mail: kamran@bio.uio.no; jan.pawlowski@unige.ch

Genome. Biol. Evol. 1:231–238.
doi:10.1093/gbe/evp022
Advance Access publication July 27, 2009

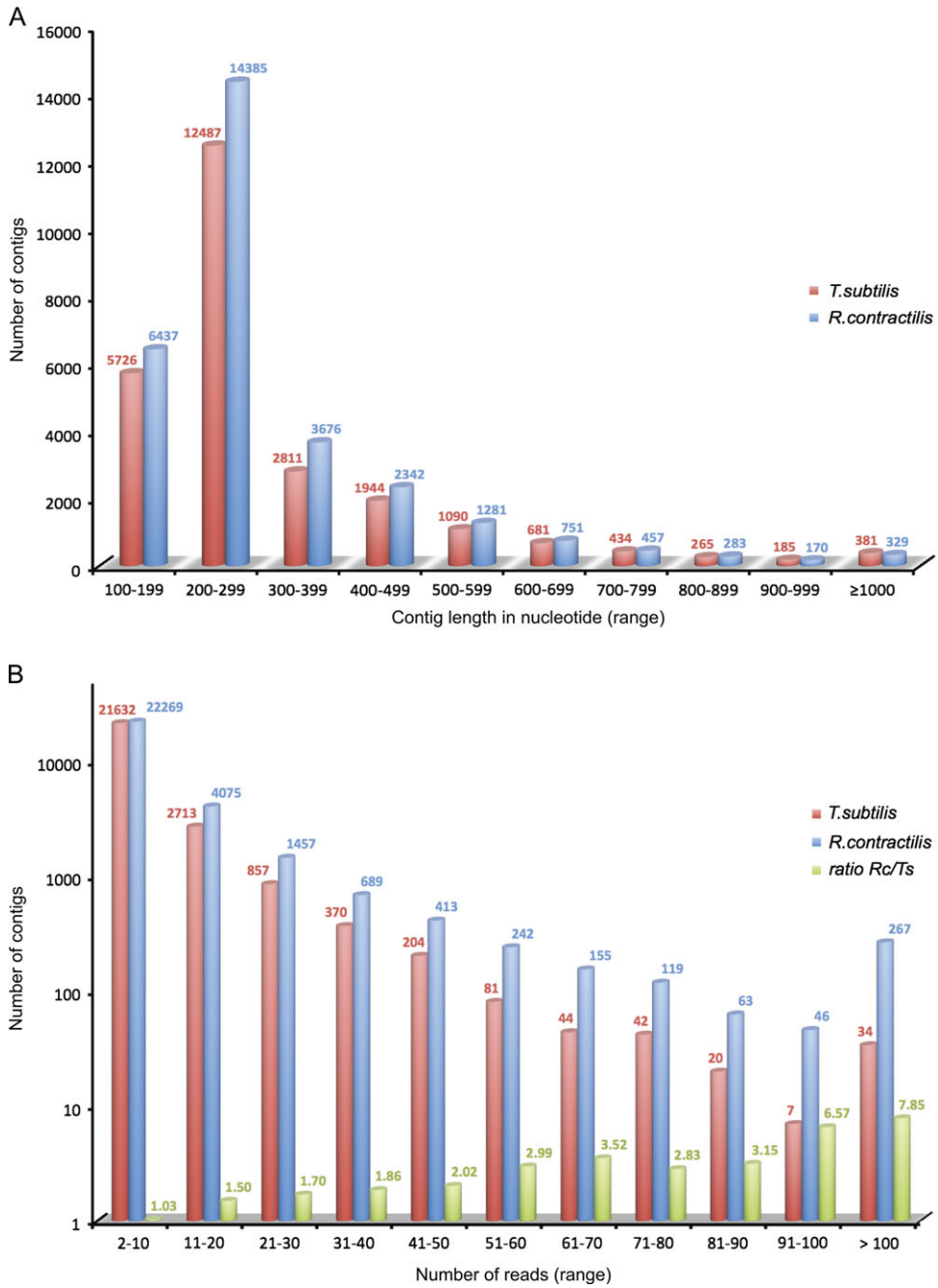

FIG. 1.—Characteristics of the 454 contigs assembled in this study for *Telonema subtilis* (red bars) and *Raphidiophrys contractilis* (blue bars). (*A*) Numbers of contigs of different lengths (in base pair). The majority of contigs were small; yet, significant numbers of contigs larger than 500 bp could be assembled. (*B*) Numbers of contigs with different numbers of reads (from 2 to 10 to over 100), illustrating the depth of coverage. The majority of contigs had low coverage; however, several thousands of contigs were comprised of at least 11 reads. The green bars show the ratio of the number of contigs for *R. contractilis* over *T. subtilis*, revealing an increasing difference between the two species as more reads were assembled into single contigs (e.g., there were 7.85 times more contigs containing at least 100 reads in *R. contractilis*, and about the same amount in the two species for the contigs made of 2–10 contigs). Vertical axis is logarithmic.

deviation [SD] = 197) and 299 bp for *R. contractilis* (median = 236; SD = 180). In *T. subtilis*, 3,041 contigs (12%) were larger than 500 bp, of which 381 were larger than 1,000 bp; in *R. contractilis*, 3,276 contigs (11%) were larger than 500 bp, of which 329 were larger than 1,000 bp (fig. 1*A*). As expected from normalized libraries (see Supplementary Material online), most contigs were comprised of a small number of reads (10 or less); yet, the high throughput pyrosequencing approach yielded 4,372 (17%) and 7,525 contigs (25%) that contained 11 or more reads for *T. subtilis* and *R. contractilis*, respectively (fig. 1*B*). Interestingly, the 150,140 additional reads for *R. contractilis* seemed mostly to increase the number of contigs with many reads (fig. 1*B*). For example, we observed nearly 8 times the number of contigs with 100 reads or more in *R. contractilis* than we did in *T. subtilis*, whereas

both species had a comparable number of contigs with 2–10 reads. This suggests that we may have sampled the majority of the expressed genes for *R. contractilis*. Yet, even for this species, the total number of contigs obtained (30,120) probably substantially overestimates the total number of genes present in the genome as many contigs are so short that many genes are represented by two or more contigs.

In addition to *T. subtilis* and *R. contractilis*, new data for the cryptomonads *Guillardia theta* and *Plagioselmis nannoplanctica* as well as the haptophyte *Imantonia rotunda* were also generated (see Supplementary Material online), leading to a much improved sampling for the cryptomonad/haptophyte group compared with earlier phylogenomic studies (Patron et al. 2007; Burki et al. 2008). These new sequences, together with publicly available data, were used to construct a multigene alignment (supermatrix) containing 127 genes (29,235 amino acid positions) and a taxon-rich sampling of 72 species belonging to all supergroups of eukaryotes. Importantly, all species were carefully selected to minimize the impact of heterogeneity in evolutionary rates by excluding long-branched taxa.

Evolutionary Origins of Telonemia and Centroheliozoa

Our concatenated data set was first analyzed by Bayesian (phylobayes—CAT model; Lartillot and Philippe 2004) and maximum likelihood (ML) methods (RAxML—RTREV model; Stamatakis 2006). Figure 2 shows an unrooted Bayesian consensus tree, with ML bootstrap values (BPs) and Bayesian posterior probabilities (PPs) indicated. All main eukaryotic assemblages were recovered with moderate to maximum support and are consistent with the most recent published studies of eukaryote evolution (Burki et al. 2007, 2008; Rodriguez-Ezpeleta et al. 2007; Hampl et al. 2009; Minge et al. 2009). Amoebozoa and opisthokonts robustly grouped together (the unikonts) to the exclusion of excavates (only monophyletic in ML analyses) and a megagroup composed of all other eukaryotes (78% BP; 1.0 PP). Within this megagrouping, Plantae were monophyletic, as was the stramenopile, alveolate, and Rhizaria clade (the SAR group; Burki et al. 2007) and the haptophyte/cryptomonad clade. Remarkably, in all analyses, *T. subtilis* and *R. contractilis* grouped with cryptomonads and haptophytes in a moderately supported clade (node 1; fig. 2: 70% BP; 0.88 PP). Within this group (henceforth the group composed of cryptomonads, centrohelids, telonemids, and haptophytes is referred to as the CCTH group), relationships were essentially unresolved and the Bayesian and ML methods yielded different but unsupported branching patterns (fig. 2). In addition, all analyses placed the CCTH group as sister to the SAR group with moderate support (node 2; fig. 2: 65% BP; 0.99 PP), and this major assemblage branched with the plants (78% BP; 1.0 PP). In an attempt to evaluate further alternative positionings of *T. subtilis* and *R. contractilis*, a procedure that randomly sampled among the 127 genes to construct 200 bootstrap replicates was applied. Each of these concatenated alignments was then analyzed by the ML method. A clade containing telonemids, centrohelids, and haptophytes was obtained in the majority rule consensus tree and recovered in 41% of the replicates (supplementary fig. 1, Supplementary Material online). The CCTH group was inferred in 35 trees (18%); this low number was notably due to the tendency of cryptomonads to branch as sister to some excavates, a relationship that was not observed after excluding *T. subtilis* and *R. contractilis* from the analysis (cryptomonads grouped with haptophytes in 61% of the trees; supplementary fig. 2, Supplementary Material online). Of much interest, *T. subtilis* and *R. contractilis* branched together in 133 trees (67%) regardless of their association with other members of the CCTH group, supporting the affinity that exists between these lineages in our "standard" supermatrix approach. With the exception of cryptomonads, this analysis also placed *T. subtilis*, *R. contractilis*, and haptophytes as sister to the SAR group (66 trees, 33%). Thus, the bootstrapping of genes approach showed overall similar relationships compared with the original data set and did not reveal supported alternative placements in the tree for the CCTH members.

Previous studies have shown that phylogenomic analyses treating multigene data sets as concatenated alignments may not sufficiently account for the evolutionary specificities of each gene and potentially introduce tree reconstruction artifacts (Bapteste et al. 2002; Philippe et al. 2004; Patron et al. 2007). We therefore conducted a "separate" analysis that takes into account the difference in evolutionary tempos and modes across genes. This analysis specifically examined the relationships among 8 major groups: 1) *T. subtilis*, 2) *R. contractilis*, 3) cryptomonads, 4) haptophytes, 5) the SAR group, 6) Plantae, 7) excavates, and 8) unikonts (opisthokonts + Amoebozoa). Because not all genes in the original selection contained at least one representative taxon for each group of interest, a subset of 87 genes (amounting to 19,270 amino acids) was selected from the total 127 genes used in the concatenation (supplementary table S1, Supplementary Material online). This analysis resulted in the same relationships observed in the Bayesian analysis of the supermatrix (RELL BPs indicated on fig. 2), notably recovering a *T. subtilis* plus *R. contractilis* clade (65% RELL BP) that formed a group with haptophytes and cryptomonads (69% RELL BP; fig. 2). Furthermore, this approach was consistent with the concatenated analysis in positioning the CCTH group together with the SAR group (92% RELL BP). Topology comparisons using the approximate unbiased (AU) test strongly confirmed the monophyletic association between the CCTH and SAR lineages. Indeed, only 19 of 351 test trees were not rejected at the 5% level, among which 18 trees contained a clade comprising *T. subtilis*, *R. contractilis*, haptophytes, cryptomonads, and the SAR group, to the exclusion of all other eukaryotes (table 1).

An Emerging Major Group of Eukaryotes and the Expansion of Chromalveolates

Telonema subtilis and *R. contractilis* are two heterotrophic unicellular eukaryotes that represent groups which are among the most difficult to place within the tree of eukaryotes (Cavalier-Smith and Chao 2003; Nikolaev et al.

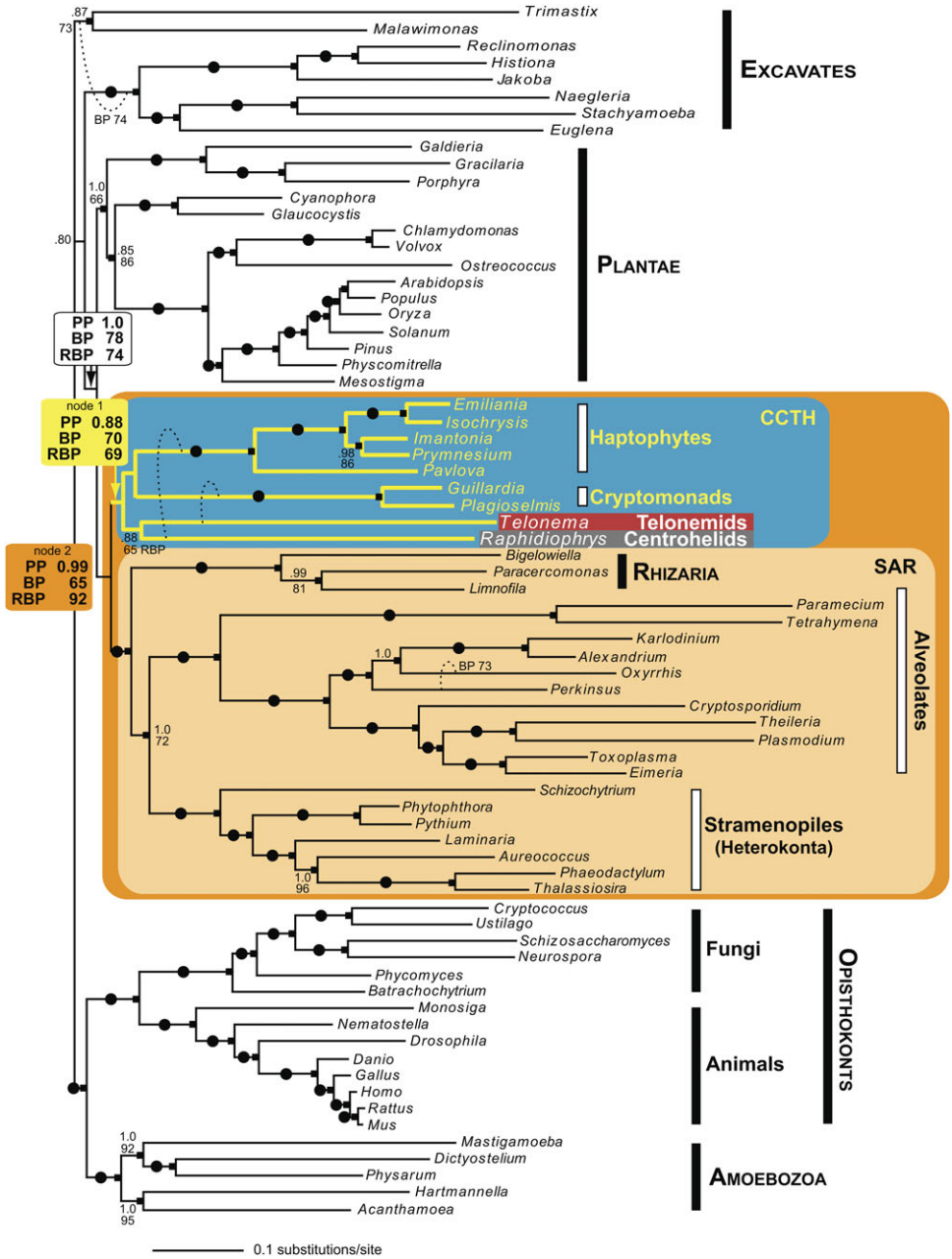

FIG. 2.—Unrooted Bayesian phylogeny of eukaryotes. The tree was obtained from the consensus between two independent Markov chains, run under the CAT model implemented in phylobayes. Identical relationships were obtained in our separate analysis (see text for Discussion). The curved dashed lines indicate the alternative branchings recovered in the ML analysis of the concatenated data set. Black dots correspond to 1.0 PP and 100% ML BP. Values at node represent PP (above) and BP (below) when not maximal. The RELL BPs calculated in the separate analysis are also shown for the three main nodes (RBP). Black squares indicate the constrained bifurcations used in the separate analysis. The white thick bars are the groups that were originally included in the chromalveolates. Assemblages indicated by capitalized names correspond to the hypothetical supergroups of eukaryotes. The scale bar represents the estimated number of amino acid substitutions per site.

2004; Shalchian-Tabrizi et al. 2006; Cavalier-Smith and von der Heyden 2007; Sakaguchi et al. 2007; Reeb et al. 2009). By generating large molecular data sets from both lineages, we have provided convincing and congruent evidence suggesting that telonemids and centrohelids both have evolutionary affinities with haptophytes and cryptomonads and more generally with the SAR group. Nevertheless, uncertainties remain. Despite use of a very large data set, there are three important reasons why we failed to recover a more highly resolved topology. First, our analyses strongly confirm earlier indications from multi-gene trees that *T. subtilis* and *R. contractilis* are not closely related to any known eukaryotic lineage (Shalchian-Tabrizi et al. 2006; Sakaguchi et al. 2007; Reeb et al. 2009) and may have diverged soon after the origin of the CCTH-SAR grouping. If true, such early divergence would have resulted in relatively few sequence synapomorphies to have arisen during their brief period of shared common ancestry and in the loss of much of that phylogenetic signal during the far longer period of subsequent evolution. Second, one

Table 1
The Details of the Test Trees Not Rejected at 5% Level in AU Test

ID[a]	Tree Topology[b]	Δln*L*	Distance from the ML Tree (SE)	*P* Value
1*	(Telo,Raph,((Cryp,Hapt),(SAR,(Plan,(Uni + Ex)))))	(ML)	—	0.899
2*	(Telo,Hapt,((Cryp,Raph),(SAR,(Plan,(Uni + Ex)))))	21.2	0.46	0.597
3*	(Telo,Raph,((SAR,(Cryp,Hapt)),(Plan,(Uni + Ex))))	21.9	0.77	0.413
4*	(Telo,Cryp,((SAR,(Raph,Hapt)),(Plan,(Uni + Ex))))	68.1	1.22	0.254
6*	(Telo,Raph,(Hapt,(Cryp,(SAR,(Plan,(Uni + Ex))))))	46.8	1.48	0.177
10*	(Telo,Raph,(Hapt,(SAR,(Cryp,(Plan,(Uni + Ex))))))	69.8	1.65	0.148
8*	(Telo,Cryp,((Raph,Hapt),(SAR,(Plan,(Uni + Ex)))))	75.4	1.57	0.127
11*	(Telo,Raph,(Hapt,((Cryp,SAR),(Plan,(Uni + Ex)))))	70.3	1.71	0.125
9*	(Telo,Hapt,(Raph,((Cryp,SAR),(Plan,(Uni + Ex)))))	80.6	1.59	0.100
12*	(Telo,Hapt,(Cryp,(Raph,(SAR,(Plan,(Uni + Ex))))))	76.6	1.72	0.089
15*	(Telo,Raph,(Cryp,(Hapt,(SAR,(Plan,(Uni + Ex))))))	58.5	1.86	0.089
20*	(Telo,Cryp,(Hapt,(Raph,(SAR,(Plan,(Uni + Ex))))))	86.8	1.95	0.088
23*	(Telo,Cryp,(Hapt,(SAR,(Raph,(Plan,(Uni + Ex))))))	99.2	2.04	0.081
16*	(Telo,Raph,(SAR,((Hapt,Cryp),(Plan,(Uni + Ex)))))	56.5	1.89	0.080
26*	(Telo,Cryp,(Hapt,((Raph,SAR),(Plan,(Uni + Ex)))))	114.2	2.22	0.078
5*	(Telo,Hapt,(SAR,((Raph,Cryp),(Plan,(Uni + Ex)))))	75.8	1.36	0.074
19*	(Telo,Hapt,(Cryp,((Raph,SAR),(Plan,(Uni + Ex)))))	97.6	1.91	0.065
17*	(Telo,Cryp,(Raph,((Hapt,SAR),(Plan,(Uni + Ex)))))	100.6	1.89	0.053
190	(Telo,Cryp,((Hapt,SAR),((Raph,Plan),(Uni + Ex))))	253.3	4.07	0.050

NOTE.—SE, standard error. In all, 87 genes were used for the test (19,270 amino acid positions in total). We subjected 347 test trees with the Unikonta–Excavata group that were distant from the ML tree by <5 SE units and 4 extra trees of particular interest.

[a] Trees with the monophyly of *Telonema*, *Raphidiophrys*, cryptomonads, haptophytes, and SAR are highlighted by asterisks.

[b] Telo, *Telonema*; Raph, *Raphidiophrys*; Cryp, cryptomonads; Hapt, haptophytes; SAR, stramenopiles + alveolates + Rhizaria; Plan, Plantae; and Uni + Ex, the grouping of Unikonta plus Exavata.

expects to observe a decrease in statistical support when early diverging species are added to a phylogeny (Sanderson and Wojciechowski 2000). In keeping with this expectation, after removing *T. subtilis* and *R. contractilis* from our multigene alignment, support for the haptophyte/cryptomonad/SAR group and its sister grouping with Plantae both increased substantially (supplementary fig. 3, Supplementary Material online). Finally, *T. subtilis* and *R. contractilis* are the only species from telonemids and centrohelids for which genomic data are available; similar samples from additional representatives of these and other related lineages are needed before their phylogenetic position can be determined conclusively. Of particular importance are the heterotrophic flagellate katablepharids that have been proposed to be sister to haptophytes or cryptomonads on the basis of a handful of genes (Okamoto and Inouye 2005; Kim and Graham 2008) or classified with cryptomonads and telonemids based on ultrastructure (Cavalier-Smith 2004) and the as-yet uncultured biliphytes that might be related to cryptomonads (Not et al. 2007; Cuvelier et al. 2008).

If the relationships between these lineages and the CCTH-SAR clade are confirmed, a new major assemblage is emerging with important implications for understanding the early evolution of eukaryotes. Recently, multigene analyses (Hackett et al. 2007; Patron et al. 2007) and a shared lateral transfer of bacterial *rpl36* to their plastid genomes (Rice and Palmer 2006) suggested that cryptomonads and haptophytes form a clade. Taken together, the evidence that katablepharids and possibly biliphytes are related to cryptomonads, and our demonstration that telonemids and centrohelids may also be part of the cryptomonad/haptophyte clade, substantially increase the organismal diversity and importance of this novel phylogenetic group. Our results also provide additional confirmation of the monophyly of the SAR clade, and, because of our carefully chosen taxon sampling, the possibility that this grouping was a long-branch artifact (Cavalier-Smith 2009) is now reduced. Moreover, the monophyly of SAR was recently strengthened by the discovery of a shared paralogy in the Rab1A gene family, representing the first synapomorphy associated with the origin of the group (Elias et al. 2009). The Rab1A paralogy is found in each completely sequenced genome and several expressed sequence tag (EST) data sets of stramenopiles, alveolates, and chlorarachniophytes (Elias et al. 2009), as well as in at least two other rhizarians (*Reticulomyxa filosa* and *Gromia sphaerica*, data not shown).

Altogether, the phylogenomic data presented here suggest that the chromalveolate assemblage should be expanded to include Rhizaria and at least two additional poorly known lineages for which plastids have never been reported, telonemids and centrohelids. However, more data are urgently needed to exclude completely that these relationships are not affected by undetected endosymbiotic gene transfers and replacements (Lane and Archibald 2008), which, given the very limited genomic data available for red algae, are currently hard to identify. In particular, the affinities between some members of the CCTH group and Plantae, as observed in several recent multigene phylogenies (Patron et al. 2007; Burki et al. 2008; Hampl et al. 2009; Minge et al. 2009), need to be further tested. This is particularly important because AU tests based on our supermatrix (contrasting with the separate analysis, see above) failed to reject that alternative relationship at the 5% level ($P = 0.072$ and $P = 0.06$ for the branching pattern within CCTH corresponding to the Bayesian tree or the ML tree, respectively) or a hypothetical grouping of haptophytes and cryptomonads alone with Plantae ($P = 0.079$). Other scenarios, such as the recent Plastidophila hypothesis (Kim and Graham 2008), also need to be specifically

addressed, essentially based on the phylogenetic signal present in just one protein (eukaryotic translation elongation factor 2, eEF2), that hypothesis challenged the monophyly of both Plantae and chromalveolates. However, Kim and Graham's interpretation of a two amino acid signature (SA) in the eEF2 protein as evidence supporting a grouping of green plants, red algae, haptophytes, cryptomonads, and katablepharids (to the exclusion of glaucophytes, alveolates, stramenopiles, and Rhizaria, which have GS) is weakened by its absence in both *T. subtilis* and *R. contractilis*, which possess GS and GA amino acid residues, respectively, as well as by other contradictions to the oversimplified signature sequence distribution they noted (i.e., *Ustilago*, *Rhizopus*, *Schizosaccharomyces*, *Reclinomonas*, *Jakoba*, and an *Acanthamoeba* all have AS instead of GS; *Malawimonas californiana* has AL and *Spironucleus* has GA).

Implications for Plastid Evolution

Many members of cryptomonads, haptophytes, stramenopiles, and alveolates possess chlorophyll-*c*–containing plastids that are, under the chromalveolate hypothesis (Cavalier-Smith 1999), postulated to have originated by a single secondary endosymbiosis of a red alga in the ancestor of all these lineages. A photosynthetic ancestry for all chromalveolates is suggested by the history inferred from plastid phylogenies (e.g., Iida et al. 2007; Khan et al. 2007) and rare genomic events such as endosymbiotic gene replacements (Fast et al. 2001; Harper and Keeling 2003; Patron et al. 2004). The unexpected phylogenetic position of Rhizaria, for which no red algal-derived plastid-bearing lineages are known, most closely related to alveolates and stramenopiles caused some controversy over the chromalveolate hypothesis (Burki et al. 2007, 2008; Hackett et al. 2007; Rodriguez-Ezpeleta et al. 2007). Taking into account these new relationships as well as general difficulties associated with the chromalveolate hypothesis, in particular concerning the lack of clear evidence for the relative difficulty of plastid gain versus plastid loss, alternative models for the origin and spread of red algal-derived plastids have been proposed (Sanchez-Puerta and Delwiche 2008; Archibald 2009; Bodyl et al. 2009).

In this context, an important question raised by our results is whether expanding the group of eukaryotes with red algal-derived plastids to include additional nonphotosynthetic lineages is still compatible with the chromalveolate hypothesis. Indeed, the addition of telonemids, centrohelids, and Rhizaria involves additional photosynthesis and/or plastid loss events to explain the observed distribution of photosynthesis under the assumption that the common ancestor of all chromalveolates was photosynthetic. However, recent discoveries of cryptic plastids and genes of putative red algal origin in nonphotosynthetic chromalveolates illustrated the difficulty in distinguishing between the absence of a plastid and the absence of photosynthesis (Tyler et al. 2006; Reyes-Prieto et al. 2008; Slamovits and Keeling 2008), indicating that the chromalveolate hypothesis remains reasonable in spite of the numerous heterotrophic lineages it comprises.

In conclusion, the genomes of telonemids and centrohelids, as well as other nonphotosynthetic lineages such as Rhizaria, katablepharids, or early divergent stramenopiles potentially bear important information to evaluate the chromalveolate hypothesis and other scenarios. Indeed, remnant algal-derived genes might still persist in these nuclear genomes, which would be very helpful to favor either a photosynthetic ancestry for the chromalveolates or the possibility that plastids were transferred between the CCTH and SAR groups by serial endosymbioses.

Materials and Methods

Contig Assembly and Sequence Alignment

All reads were assembled into contigs using the Newbler assembler with default parameters. We searched among contigs larger than 200 bp for sequences with significant similarity to genes recently used in multigene phylogenies using the following rigorous procedure (Burki et al. 2007, 2008): 1) BlastP searches against the translated set of *T. subtilis* and *R. contractilis* contigs using as queries the single-gene sequences composing our multiple alignments; 2) retrieving (with a stringent *e* value cutoff at 10^{-40}) and adding of the new homologous copies to existing single-gene alignments; 3) automatic alignments using Mafft (Katoh et al. 2002), followed by manual inspection to remove unambiguously aligned positions; 4) testing the orthology, in particular possible lateral or ancestral endosymbiotic gene transfer, for each of the selected genes by performing single-gene ML reconstructions using Treefinder (WAG substitution matrix and six gamma categories; Jobb et al. 2004) and visually inspecting the resulting individual trees. We retained a set of 127 genes (29,235 amino acid positions) that did not show any obvious problem of deep paralogy or nonvertical transmission and 72 species excluding fast-evolving taxa used previously (Burki et al. 2008; the rhizarians *Reticulomyxa* and *Quinqueloculina*; the stramenopile *Blastocystis*; and the excavates *Sawyeria*, *Leishmania*, and *Trypanosoma*) when more slowly evolving lineages were available. Importantly, careful attention was made to correctly distinguish the *Imantonia* sequences from the *Telonema* library and the *Chlorogonium* sequences from the *Raphidiophrys* library. These species were kept in the single genes only when unambiguous sequence attributions were recovered, so for the trees shown here there was no confusion between the genes of the centrohelid and telonemid and those of their cocultured algal food. Monophyletic groups corresponding to haptophytes (including *Imantonia* but excluding *Telonema*) and green algae (including *Chlorogonium* but excluding *Raphidiophrys*) were mandatory in order to consider sequences from these species for concatenation. The final concatenation of all single-gene alignments was done using Scafos (Roure et al. 2007). Because of the limited data available for certain groups and to maximize the number of genes for each taxonomic assemblage, some lineages were represented by different closely related species always belonging to the same genus.

Phylogenetic Analyses

RAxML (Stamatakis 2006) was used in combination with the RTREV amino acid replacement matrix. The best

ML tree was determined with the PROTGAMMA implementation in multiple inferences using 10 randomized maximum parsimony starting trees. Statistical support was evaluated with 200 bootstrap replicates. Four independent runs from different starting trees were performed on each replicate in order to prevent the analysis from getting trapped in a local maximum. The tree with the best log likelihood was selected for each replicate, and the 200 resulting trees were used to calculate the bootstrap proportions. To save computational burden, the PROTMIX solution was chosen with 25 distinct rate categories. Phylobayes (Lartillot and Philippe 2004) was run using the site-heterogeneous mixture CAT model and two independent Markov chains with a total length of 8,000 cycles, discarding the first 1,000 points as burn-in, and calculating the posterior consensus on the remaining trees. Convergence between the two chains was ascertained by comparing the frequency of their bipartitions. The bootstrapping of genes analysis used a new perl script that allows individual genes to be sampled (with replacement) to create new matrices containing the same number of genes as the original concatenated alignment. Specifically, 200 replicates containing 127 concatenated genes were constructed by sampling from within the initial pool of 127 genes and analyzed by ML as described above. To assess the robustness of the phylogenetic position of the CCTH group (or the position of the haptophytes and cryptomonads in the absence of *T. subtilis* and *R. contractilis*), we conducted topology comparisons using the AU test based on our supermatrix. Alternative topologies were generated by moving the CCTH group (or haptophytes and cryptomonads alone) either as sister to Plantae or at any possible positions within Plantae. Site likelihoods were then calculated using RAxML, and the AU test was performed using CONSEL v.0.1 (Shimodaira and Hasegawa 2001) with default scaling and replicate values.

In the separate analysis, we exhaustively examined the 10,395 test trees resulting from the applied constraints on eight major groups of eukaryotes. Log likelihoods for each test tree were calculated under the RTREV + Γ model using RAxML. RELL BPs were calculated using TotalML in Molphy v.2.3 (Adachi and Hasegawa 1996). Out of these 10,395 possible topologies, we subjected 351 test trees to the AU test. Specifically, we considered the 347 trees possessing the unikonts–excavates bifurcation that were closer than five standard error units from the ML tree (this restricted number of tested trees was due to computational burden) and four additional trees that were constructed by 1) moving *T. subtilis* to the branch leading to unikonts, 2) moving *T. subtilis* to the branch leading to excavates, 3) moving *R. contractilis* to the branch leading to unikonts, and 4) moving *R. contractilis* to the branch leading to excavates.

Funding

Research Council of Norway (the FUGE and FRIBIO programs) to K.S.J; Swiss National Science Foundation (3100A0–112645 and 31003A–125372); Centre for Microbial Diversity and Evolution from the Tula Foundation to A.H.; and Japanese Society for the Promotion of Sciences (201242) to Y.I.

Supplementary Material

Supplementary figures 1–3 and table S1 are available at *Genome Biology and Evolution* online (http://www.oxfordjournals.org/our_journals/gbe/). New sequences that were generated in this study and included in the concatenated alignment were deposited in EMBL nucleotide sequence database (GenBank accession numbers: *Telonema subtilis* FN392439–FN392557; *Raphidiophrys contractilis* FN392328–FN392438; *Plagioselmis nannoplanctica* FN392589–FN392620; *Imantonia rotunda* FN392558–FN392588) and GenBank (accession numbers: *Guillardia theta* GO787145–GO787665).

Acknowledgments

We thank Daniel Vaulot for providing the *Telonema subtilis* culture. All analyses were performed on the freely available Bioportal at the University of Oslo (http://www.bioportal.uio.no) and the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the Swiss Institute of Bioinformatics. We thank Ave Tooming-Klunderud and Lex Nederbragt at the Ultra-high throughput sequencing laboratory at the University of Oslo for help with 454 sequencing of *T. subtilis*. And we thank Tom Andersen for assistance with bootstrapping of genes. K.S.T. thanks the University of Oslo for starting grants and EMBIO for funding of the Bioportal. F.B. is grateful to the Fondation Ernst and Lucie Schmidheiny for funding part of the *R. contractilis* sequencing. The *G. theta* ESTs used in this study were generated by the Protist EST Program and the Joint Genome Institute's Community Sequencing Program (http://www.jgi.doe.gov/sequencing/why/50026.
html). We thank Kerrie Barry and Erika Lindquist of the JGI for project management and data availability.

Literature Cited

Adachi J, Hasegawa M. 1996. MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. Comput Sci Monographs. 28:1–150.

Archibald JM. 2009. The puzzle of plastid evolution. Curr Biol. 19:R81–R88.

Bapteste E, et al. 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. Proc Natl Acad Sci USA. 99:1414–1419.

Bodyl A, Stiller JW, Mackiewicz P. 2009. Chromalveolate plastids: direct descent or multiple endosymbioses? Trends Ecol Evol. 24:119–121.

Burki F, et al. 2007. Phylogenomics reshuffles the eukaryotic supergroups. PLoS ONE. 2:e790.

Burki F, Shalchian-Tabrizi K, Pawlowski J. 2008. Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. Biol Lett. 4:366–369.

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. 46:347–366.

Cavalier-Smith T. 2004. Chromalveolate diversity and cell megaevolution: interplay of membranes, genomes and cytoskeleton.

In: Hirt RP, Horner DS, editors. Organelles, genomes and eukaryote phylogeny. London: Taylor & Francis. p. 75–108.
Cavalier-Smith T. 2009. Megaphylogeny, cell body plans, adaptive zones: causes and timing of eukaryote basal radiations. J Eukaryot Microbiol. 56:26–33.
Cavalier-Smith T, Chao E. 2003. Molecular phylogeny of centrohelids heliozoa, a novel lineage of bikont eukaryotes that arose by ciliary lost. J Mol Evol. 56:387–396.
Cavalier-Smith T, von der Heyden S. 2007. Molecular phylogeny, scale evolution and taxonomy of centrohelid heliozoa. Mol Phylogenet Evol. 44:1186–1203.
Cuvelier ML, et al. 2008. Widespread distribution of a unique marine protistan lineage. Environ Microbiol. 10:1621–1634.
Elias M, Patron NJ, Keeling PJ. 2009. The RAB family GTPase Rab1A from *Plasmodium falciparum* defines a unique paralog shared by chromalveolates and rhizaria. J Eukaryot Microbiol. 56:348–356.
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. 18:418–426.
Hackett JD, et al. 2007. Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. Mol Biol Evol. 24:1702–1713.
Hampl V, et al. 2009. Phylogenomic analyses support the monophyly of excavata and resolve relationships among eukaryotic "supergroups". Proc Natl Acad Sci USA. 106:3859–3864.
Harper JT, Keeling PJ. 2003. Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. Mol Biol Evol. 20:1730–1735.
Iida K, Takishita K, Ohshima K, Inagaki Y. 2007. Assessing the monophyly of chlorophyll-*c* containing plastids by multi-gene phylogenies under the unlinked model conditions. Mol Phylogenet Evol. 45:227–238.
Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol. 4:18.
Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.
Khan H, et al. 2007. Plastid genome sequence of the cryptophyte alga *Rhodomonas salina* CCMP1319: lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny. Mol Biol Evol. 24:1832–1842.
Kim E, Graham LE. 2008. EEF2 analysis challenges the monophyly of Archaeplastida and Chromalveolata. PLoS ONE. 3:e2621.
Klaveness D, Shalchian-Tabrizi K, Thomsen HA, Eikrem W, Jakobsen KS. 2005. *Telonema antarcticum* sp. nov., a common marine phagotrophic flagellate. Int J Syst Evol Microbiol. 55:2595–2604.
Lane CE, Archibald JM. 2008. The eukaryotic tree of life: endosymbiosis takes its TOL. Trends Ecol Evol. 23:268–275.
Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol. 21:1095–1109.
Minge MA, et al. 2009. Evolutionary position of breviate amoebae and the primary eukaryote divergence. Philos Trans R Soc Lond B Biol Sci. 276:597–604.
Nikolaev SI, Berney C, Fahrni JF, Bolivar I. 2004. The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. Proc Natl Acad Sci USA. 101:8066–8071.
Not F, et al. 2007. Picobiliphytes: a marine picoplanktonic algal group with unknown affinities to other eukaryotes. Science. 315:253–255.
Okamoto N, Inouye I. 2005. The katablepharids are a distant sister group of the Cryptophyta: a proposal for Katablepharidophyta divisio nova/Kathablepharida phylum novum based on SSU rDNA and beta-tubulin phylogeny. Protist. 156:163–179.
Patron NJ, Inagaki Y, Keeling PJ. 2007. Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. Curr Biol. 17:887–891.
Patron NJ, Rogers MB, Keeling PJ. 2004. Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. Eukaryot Cell. 3:1169–1175.
Philippe H, et al. 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. Mol Biol Evol. 21:1740–1752.
Reeb VC, et al. Forthcoming. 2009. Interrelationships of chromalveolates within a broadly sampled tree of photosynthetic protists. Mol Phylogenet Evol. doi: 10.1016/j.ympev.2009.04.012.
Reyes-Prieto A, Moustafa A, Bhattacharya D. 2008. Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic. Curr Biol. 18:956–962.
Rice DW, Palmer JD. 2006. An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. BMC Biol. 4:31.
Rodriguez-Ezpeleta N, et al. 2007. Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. Curr Biol. 17:1420–1425.
Roure B, Rodriguez-Ezpeleta N, Philippe H. 2007. SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. BMC Evol Biol. 7(Suppl 1):S2.
Sakaguchi M, Inagaki Y, Hashimoto T. 2007. Centrohelida is still searching for a phylogenetic home: analyses of seven *Raphidiophrys contractilis* genes. Gene. 405:47–54.
Sakaguchi M, Nakayama T, Hashimoto T, Inouye I. 2005. Phylogeny of the Centrohelida inferred from SSU rRNA, tubulins, and actin genes. J Mol Evol. 61:765–775.
Sanchez-Puerta MV, Delwiche CF. 2008. A hypothesis for plastid evolution in chromalveolates. J Phycol. 1097–1107.
Sanderson MJ, Wojciechowski MF. 2000. Improved bootstrap confidence limits in large-scale phylogenies, with an example from Neo-Astragalus (Leguminosae). Syst Biol. 49:671–685.
Shalchian-Tabrizi K, et al. 2006. Telonemia, a new protist phylum with affinity to chromist lineages. Philos Trans R Soc Lond B Biol Sci. 273:1833–1842.
Shalchian-Tabrizi K, et al. 2008. Multigene phylogeny of Choanozoa and the origin of animals. PLoS ONE. 3:e2098.
Shalchian-Tabrizi K, Kauserud H, Massana R, Klaveness D, Jakobsen KS. 2007. Analysis of environmental 18S ribosomal RNA sequences reveals unknown diversity of the cosmopolitan phylum Telonemia. Protist. 158:173–180.
Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics. 17:1246–1247.
Slamovits CH, Keeling PJ. 2008. Plastid-derived genes in the non-photosynthetic alveolate *Oxyrrhis marina*. Mol Biol Evol. 25:1297–1306.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22:2688–2690.
Tyler BM, et al. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science. 313:1261–1266.

Gertraud Burger, Associate Editor

Accepted July 20, 2009