PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, Smirnov A, Mylnikov AP, Keeling PJ. 2016 Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. R. Soc. B* **283**: 20152802.
http://dx.doi.org/10.1098/rspb.2015.2802

Received: 24 November 2015
Accepted: 22 December 2015

Subject Areas:
evolution, taxonomy and systematics

Keywords:
phylogenomics, eukaryotes, centrohelids, tree of life, plastid evolution

Authors for correspondence:
Fabien Burki
e-mail: burkif@mail.ubc.ca
Patrick J. Keeling
e-mail: pkeeling@mail.ubc.ca

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2015.2802 or via http://rspb.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista

Fabien Burki[1], Maia Kaplan[1], Denis V. Tikhonenkov[1,2], Vasily Zlatogursky[3], Bui Quang Minh[4], Liudmila V. Radaykina[2], Alexey Smirnov[3], Alexander P. Mylnikov[2] and Patrick J. Keeling[1,5]

[1]Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada
[2]Institute for Biology of Inland Waters, Russian Academy of Sciences, Borok, Russia
[3]Department of Invertebrate Zoology, St Petersburg State University, St Petersburg, Russia
[4]Center for Integrative Bioinformatics, Max F. Perutz Laboratories, University of Vienna, Medical University of Vienna, Vienna, Austria
[5]Canadian Institute for Advanced Research, Integrated Microbial Biodiversity Program, Toronto, Ontario, Canada

Assembling the global eukaryotic tree of life has long been a major effort of Biology. In recent years, pushed by the new availability of genome-scale data for microbial eukaryotes, it has become possible to revisit many evolutionary enigmas. However, some of the most ancient nodes, which are essential for inferring a stable tree, have remained highly controversial. Among other reasons, the lack of adequate genomic datasets for key taxa has prevented the robust reconstruction of early diversification events. In this context, the centrohelid heliozoans are particularly relevant for reconstructing the tree of eukaryotes because they represent one of the last substantial groups that was missing large and diverse genomic data. Here, we filled this gap by sequencing high-quality transcriptomes for four centrohelid lineages, each corresponding to a different family. Combining these new data with a broad eukaryotic sampling, we produced a gene-rich taxon-rich phylogenomic dataset that enabled us to refine the structure of the tree. Specifically, we show that (i) centrohelids relate to haptophytes, confirming Haptista; (ii) Haptista relates to SAR; (iii) Cryptista share strong affinity with Archaeplastida; and (iv) Haptista + SAR is sister to Cryptista + Archaeplastida. The implications of this topology are discussed in the broader context of plastid evolution.

1. Introduction

Reconstructing the tree of life is a challenging task, because the long evolutionary history since the origin of life has often confounded the phylogenetic signal that can be recovered today. Nevertheless, molecular-based phylogenies have made possible profound rearrangements in the tree, most recently using phylogenomics (i.e. the use of genomic-scale datasets with stronger phylogenetic power) [1]. Accordingly, the global tree of eukaryotes has been reshuffled once again, leading to a better understanding of the relationships between the largest assemblages, or supergroups, and the origins of some 'orphan' lineages [2]. However, contentious nodes between supergroups remain, as well as a few lingering 'orphans'. Resolving the positions of these orphans is necessary for understanding their evolution, but also impacts the tree as a whole, because poorly sampled 'orphan' groups may lead to instability in the tree.

One such group lacking proper genomic data is Centrohelida, a monophyletic group of free-living predatory protists mainly found in freshwater and soil habitats, but also increasingly recognized to occur widely in marine environments [3]. With about 90 described species and a vast diversity of environmental sequences [4], centrohelids have traditionally constituted the core of the original phylum Heliozoa,

which included a subset of microbial eukaryotes characterized by a special type of pseudopodia, the axopodia. Heliozoa was shown to be a polyphyletic assemblage, and today several relatively minor lineages are scattered across the tree [5].

Centrohelids, however, have remained one of the last substantially diverse groups of eukaryotes that has eluded phylogenetic placement in the tree of life. Different analyses of the 18S rRNA and a small number of protein-coding genes (actin, α-tubulin, β-tubulin, EF2, HSP70, HSP90) led to placement in various regions of the tree, but never with good statistical support. For example, centrohelids were weakly inferred to branch close to members of the Viridiplantae, specifically glaucophytes [6] or red algae [7]. Other studies showed the centrohelids to share affinities with haptophytes [4,8], or were inconclusive [9]. Even a larger-scale multigene analysis involving 127 genes was unsuccessful at the task, placing centrohelids with low confidence as sister to either haptophytes or the enigmatic telonemids [10]. More recently, the partial transcriptome sequencing for the tiny centrohelid *Oxnerella marina* was included in a 187 genes dataset, which resulted in a less ambiguous monophyletic grouping with haptophytes [11], reinforcing the phylum Haptista originally proposed on weaker evidence [4,12].

Beyond their intrinsic interest as a large group of eukaryotes with unknown evolutionary origin, centrohelids also hold some of the clues to better understand a larger and *a priori* unrelated evolutionary mystery. Owing to their possible link to haptophytes [11], centrohelids may help to shed light on one of the most puzzling aspects of plastid evolution: the origin and evolution of complex red plastids [13,14]. Centrohelids are heterotrophs, and no permanent plastid has ever been observed [15], although kleptoplasty has been reported [16]. Haptophytes, on the other hand, are phototrophs and possess complex plastids derived from an endosymbiotic event with a red alga [17]. Haptophytes represent one of four lineages harbouring such plastids, the others being ochrophytes (photosynthetic stramenopiles), myzozoans (alveolates with plastids: apicomplexans, dinoflagellates and chrompodellids) and cryptophytes (belonging to Cryptista, which also include goniomonads, katablepharids and Palpitia). Whereas the origins of stramenopiles and alveolates are better understood [18,19], haptophytes and Cryptista have notoriously remained challenging to place in the tree. They are sometimes grouped together, along with telonemids and centrohelids [10,11,20–23], which resulted in the establishment of Hacrobia [24]. However, haptophytes and Cryptista have also been shown to have polyphyletic origins in several recent multigene analyses [25–27]. Thus, untangling the controversial phylogenetic positions of these two groups, along with their closely related plastid-lacking lineages such as centrohelids, is a much-needed step to better explain the observed distribution of red plastids in the eukaryotic tree.

In this study, we used a phylogenomic approach including a broad sampling of diversity to investigate the deep evolutionary relationships among eukaryotes, with particular focus on centrohelids, haptophytes and Cryptista. For that purpose, we filled an important gap in genome datasets by sequencing high-quality transcriptomes for four centrohelid species, and combined those with recent transcriptomes for a very large diversity of marine microbial eukaryotes (the MMETSP initiative [28]). Cultures for four species were established, each representing a different centrohelid family, namely *Raphidiophrys heterophryoidea* (Raphidiophryidae), *Raineriophrys erinaceoides* (Pterocystidae), as well as two undescribed species: *Acanthocystis* sp. (Acanthocystidae) and *Choanocystis* sp. (Choanocystidae). Our analyses unambiguously confirm that centrohelids share a common origin with haptophytes. More generally, we present compelling evidence for the phylogenetic position of the centrohelid–haptophyte group and Cryptista, altogether bringing us one step closer to a fully resolved eukaryotic tree of life.

2. Methods

Details of experimental procedure for culturing, molecular work, sequencing, assembling and gene preparation are described in the electronic supplementary material.

(a) Phylogenomic datasets construction

Following the preparation of 263 genes for phylogenomic analysis (see electronic supplementary material), all taxa were listed with SCaFoS [29], which amounted to 274 taxa. This list was first reduced to 234 taxa after removing all taxa with more than or equal to 20% missing genes. A 234-taxa, 263-gene (234/263) supermatrix was then constructed to infer an initial maximum-likelihood (ML) tree with IQ-TREE v. 1.3.0 [30] under the LG + Γ model. Based on this initial tree, a phylogeny-driven taxon selection approach was applied to reduce further the number of taxa by retaining only representative sequences within strongly supported monophyletic groups (100% bootstrap support), discarding the longest branches and/or least complete sequences. Chimeric concatenated sequences were also allowed by pooling highly incomplete taxa of the same genus (see electronic supplementary material, table S1 for details). This approach led to a final taxon sampling composed of 150 operational taxonomic units (OTUs). Because removal of ambiguously aligned sites is directly influenced by the proportion of gaps, we then re-extracted from the unaligned and untrimmed fasta files the 150 OTUs corresponding to our final selection, which were re-aligned with MAFFT-LINSI v. 7 and trimmed with BMGE v. 1.1 [31] using conservative settings (removal of sites with more than 20%, minimum block size of 8, substitution matrix BLOSUM 75). Finally, from our starting dataset of 263 seed genes, only 250 were retained to enter the final concatenated alignment (55 554 aa positions), which corresponded to genes with less than 50% missing OTUs. See electronic supplementary material, table S1 for details about missing data, and electronic supplementary material, table S2 for complete gene names. These 250 genes, containing up to 150 OTUs, were concatenated into a supermatrix (150/250) with SCaFoS [29].

From the full 150/250 dataset, two reduced datasets were considered. First, *Telonema subtilis* and *Picomonas* sp. were removed (see Results and Discussion for the justification), leading to the 148/250 dataset. This dataset was reduced further by eliminating the 19 047 fastest-evolving positions corresponding to bin10, according to the tree-independent method described in [32]; this dataset was named 148/250-slow.

(b) Phylogenetic analyses

Our supermatrices were analysed by ML and Bayesian tree reconstruction methods. ML analyses were performed with IQ-TREE v. 1.3.0–1.3.10 [30]. Gene-partitioned and unpartitioned alignments were analysed; in all cases, the model that best fits the data was determined by IQ-TREE according to the Bayesian information criterion (BIC). The partitioned analysis was applied to the 150/250 and 148/250 datasets, where the best-fit model was chosen according to a greedy strategy that sequentially merges genes from the fully partitioned alignment (250 partitions) until the model fit stops improving. We opted for the new model selection procedure (-m TESTNEW), which additionally implements the FreeRate

heterogeneity model inferring the site rates directly from the data instead of being drawn from a gamma distribution [33]. Owing to the large size of the partition schemes, only the top 20% was checked using the relaxed clustering algorithm (-rcluster 20), as described in [34]. For both datasets, the best-fit partitioning scheme contained the original 250 partitions, i.e. no merging was deemed necessary. This partitioning scheme was then used to specify a model for each partition, allowing each gene to have its own rate (-spp). For the unpartitioned analyses of both 150/250 and 148/250 supermatrices, the best-fitted model corresponded to the LG matrix with relative rates estimated from the data using the non-parametric FreeRate model with 10 categories and empirical amino acid frequencies (LG + R10 + F). The best-fitted model for the unpartitioned analysis of the 148/250-slow dataset was LG + R6 + F. A more complex empirical mixture model not evaluated by the selection strategy in IQ-TREE was also tested on all datasets: following recommendation in [35], the LG matrix was combined to an amino acid class frequency mixture model with 60 frequency component profiles plus a class of empirical amino acid frequency of the alignment, and four gamma categories to take into account the across-site rate heterogeneity (LG + C60 + F). To assess branch support, all IQ-TREE analyses used the ultrafast bootstrap approximation (UFboot) with 1000 replicates [36] and the SH-like approximate likelihood ratio test (SH-aLRT) also with 1000 bootstrap replicates [37].

Bayesian analyses were performed with PhyloBayes MPI v. 1.5a [38], under a site-heterogeneous mixture model combining infinite profile mixtures and exchange rates inferred from the data with the rates across site drawn from a discrete gamma distribution (CAT + GTR + Γ4). Constant sites were removed to decrease computational time (-dc). Three independent Markov chain Monte Carlo (MCMC) chains were run, for at least 3000 generations but up to 7000 for the smaller 148/250-slow dataset. The burnin period was determined after plotting the evolution of the log-likelihood (Lnl) across the iterations, removing the generations anterior to the stabilization of the Lnl. Convergence between the chains was assessed by examining the difference in frequency between all bipartitions (maxdiff). Owing to the large size of our taxon sampling, convergence was generally not globally achieved (maxdiff $\geq$ 0.46), an issue that has been reported in other taxon-rich phylogenomic studies [11,22]. The discrepancies between the chains mostly concerned nodes not under active discussion in this study, except for the monophyly of Archaeplastida, which was accordingly labelled unsupported; electronic supplementary material, figures S2 and S5 show the trees inferred from each individual chains to allow visual assessment of the discrepancies.

3. Results

(a) Improved dataset and model selection

To place the centrohelids in a broad eukaryotic framework, we took special care to include a very large diversity for all known main lineages. Building on previously published datasets [25,39], we more than doubled the taxon sampling, mostly using recently released high-quality transcriptomes for marine microbial species [28] (electronic supplementary material, table S3). Our carefully curated taxon sampling contained 150 OTUs for 250 genes (55 554 aa positions), globally characterized by only 21% of missing data (electronic supplementary material, table S1). Importantly, the four new centrohelid sequences missed only between 7.4% and 12.9% positions, which corresponded to at least 48 399 aa positions included, representing many fold improvements compared with the 76.3% missing data for the older *Polyplacocystis contractilis* dataset [10].

In total, four models of evolution were tested on the different datasets: ML analyses employed a partition approach with 250 gene partitions allowing each gene to have its own model, the LG + Rx + F model and the LG + C60 + F model (electronic supplementary material, table S4); Bayesian analyses were run under the CAT + GTR + Γ4 model. To select the best-fitting model in ML, we followed the BIC score selection criterion, which showed that the LG + C60 + F model consistently achieved better scores than the other two models (electronic supplementary material, table S4). In Bayesian framework, the CAT + GTR + Γ4 model has been repeatedly shown to have a better fit than simpler models based on empirical exchangeability matrices such as LG, or even CAT + Γ4 alone [40,41]. However, the size of our datasets makes comparing the fit of these complex models computationally prohibitive, and thus topologies corresponding to the best-fitting LG + C60 + F model (ML) and the CAT + GTR + Γ4 model (Bayesian) are discussed in the following sections.

(b) Evolutionary relationships among major eukaryotic lineages

The LG + C60 + F and CAT + GTR + Γ4 analyses of the complete dataset (150/250) recovered with maximal support (100% UFboot and SH-aLRT; 1.0 PP) a monophyletic assemblage including centrohelids and haptophytes (figure 1; electronic supplementary material, S1). More generally, these analyses recovered many of the major eukaryotic groups, namely Obazoa, Amoebozoa, Excavata and the SAR assemblage (stramenopiles, alveolates, Rhizaria). The association previously suggested between cryptomonads, katablepharids and the marine biflagellate *Palpitomonas bilix* into the Cryptista clade was supported with 100% UFboot and SH-aLRT and 1.0 PP [25,27,43]. In the LG + C60 + F tree, the Archaeplastida lineages (i.e. green algae and land plants, glaucophytes and red algae) were paraphyletic, with Cryptista branching with green plants and glaucophytes (96% UFboot; 88% SH-aLRT). In the CAT + GTR + Γ4 analysis, the position of Cryptista among the Archaeplastida lineages was unresolved owing to incongruent nodes in the independent MCMC chains (electronic supplementary material, figure S2*a–c*). Telonemids were recovered as sister to SAR (93% UFboot; 99% SH-aLRT; 0.78 PP) and Picozoa as sister to the red algae (93% UFboot; 100% SH-aLRT; 1.0 PP).

Following the inference of a close evolutionary link between centrohelids and haptophytes, the next important question is where Haptista goes in the global tree. The analyses of the 150/250 dataset placed Haptista as sister to SAR, a relationship that received no support under the LG + C60 + F model, but 1.0 PP under the CAT + GTR + Γ4 model (figure 1). To investigate this and the deeper structure of the tree in more detail, we reduced our dataset in two successive steps. First, we removed two orphan lineages, *T. subtilis* and Picozoa, leading to the 148/250 dataset. These enigmatic taxa mirror in many ways the problems we sought to solve here for centrohelids. They are still extremely poorly represented in genomic databases, being the sole representatives of a much higher lineage diversity [44,45], which translates into high proportions of missing data (67% for telonemids and 86% for *Picomonas* sp.; electronic supplementary material, table S1). Second, we removed from the 148/250 supermatrix the 19 047 fastest-evolving positions using the similarity between characters as an estimate of the evolutionary rates

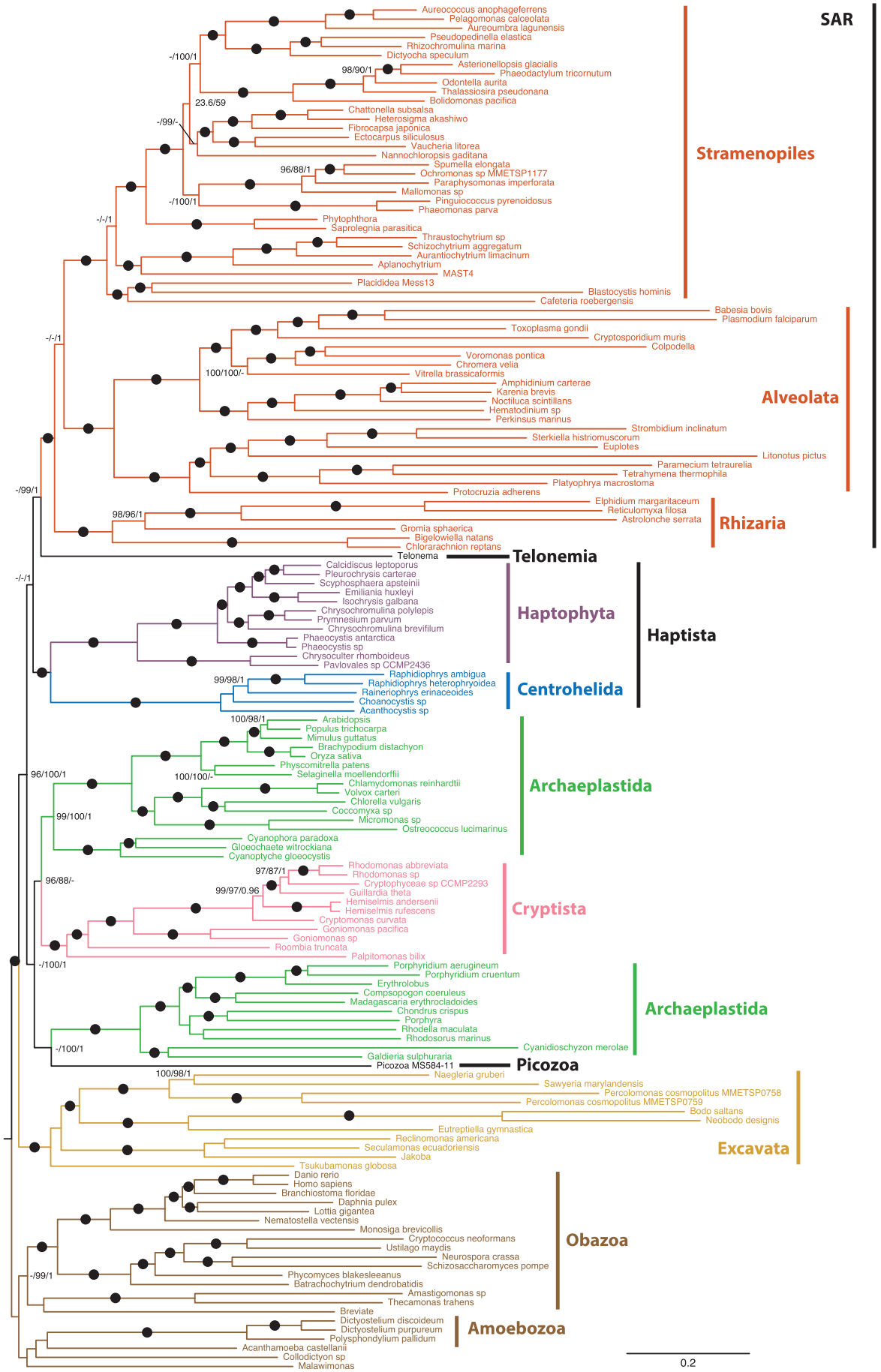

Figure 1. Phylogenetic tree of eukaryotes inferred from the complete dataset (150/250). The topology shown corresponds to the ML tree under the LG + C60 + F model, with both ML and Bayesian support value reported. Black dots on branches mean maximal support (i.e. 100% UFboot and SH-aLRT, and 1.0 Bayesian PP; the Bayesian CAT + GTR + Γ4 topology is shown in electronic supplementary material, figure S1). When not maximal, values are indicated only if deemed robust as follows: UFboot $\geq$ 95%/SH-aLRT $\geq$ 80%/PP $\geq$ 0.9. The tree is drawn rooted between Obazoa, Amoebozoa, Collodictyon, Malawimonas and the rest of eukaryotes after [42], though we note that the position of the root is under active debate.

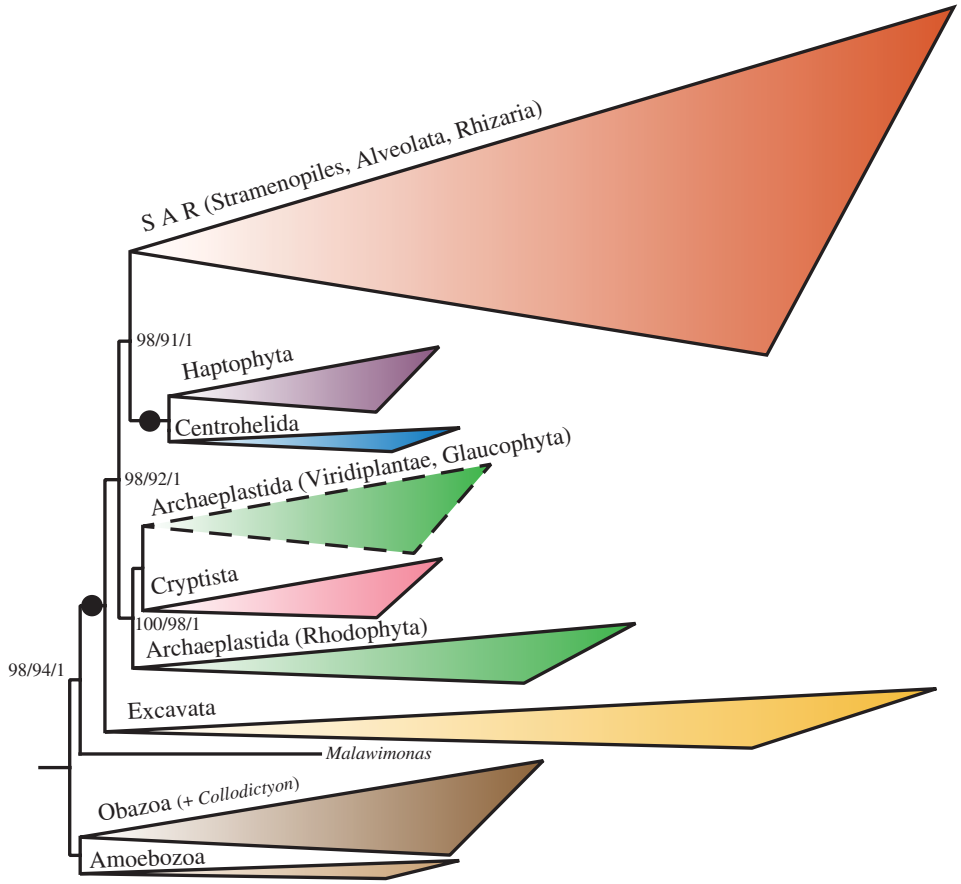

Figure 2. Schematics of the new backbone for the eukaryotic tree, highlighting the relationships among the main groups. The topology is based on the 148/250-slow supermatrix, and corresponds to both ML and Bayesian reconstructions under the LG + C60 + F and CAT + GTR + Γ4 models, respectively. The complete tree is presented in electronic supplementary material, figure SX. Black dots on branches mean maximal support (i.e. 100% UFboot and SH-aLRT, and 1.0 Bayesian PP). When not maximal values are indicated as followed: UFboot/SH-aLRT/PP. All supergroups indicated by the triangles received maximal support, with the exception of the grouping of Viridiplantae and glaucophytes, which was unsupported (shown by dashed lines). The size of the triangles roughly represents the diversity of taxa included in our analyses, as well as the length of the longest branch in each group. The root is placed in the same position as in figure 1.

[32], leading to the 148/250-slow dataset. Fast-evolving positions are more likely to concentrate undetected multiple substitutions, even by advanced models of evolution such as the mixture models used here. Removing these positions from large alignments diminishes the amount of undetected multiple substitutions, but maintains enough phylogenetic information to reconstruct even ancient events, so this approach has shown great potential in other phylogenomic studies [26,46].

The resulting topologies were similar to those based on the full dataset. However, whereas the analyses of the 148/250 dataset did not improve the general statistical support of the tree (electronic supplementary material, figures S3, S4 and S5*a–c*), the reconstructions based on the 148/250-slow dataset led to consistent and more robust topologies (figure 2; electronic supplementary material, S6). Here, Haptista received strong support (98% UFboot; 91% SH-aLRT; 1.0 PP) for its position as sister to SAR, and the Archaeplastida lineages and Cryptista were strongly inferred to share a common ancestor (100% UFboot; 98% SH-aLRT; 1.0 PP). Archaeplastida remained paraphyletic, but this was still unsupported and should be further tested (69% UFboot; 80% SH-aLRT; 0.89 PP). In these analyses, the Archaeplastida–Cryptista grouping branched with SAR + Haptista to the exclusion of all other eukaryotes with maximal support (100% UFboot; 100% SH-aLRT; 1.0 PP).

4. Discussion

(a) Towards resolving the eukaryotic tree

Over the past decade, several phylogenomic studies have attempted to resolve the deep-level relationships among the main lineages of eukaryotes [18,19,25–27,47]. These studies have greatly improved our model for the tree of eukaryotes, but several questions remain unsolved owing to the lack of data from poorly studied groups. Among these unsolved questions, the relationships between centrohelids, haptophytes, Cryptista and the main Archaeplastida lineages (green plants, glaucophytes and red algae) have all proved to be refractory to robust phylogenetic inferences. A combination of three important sources of artefact is most likely to explain the poor resolution for the placement of these lineages: (i) lack of data; (ii) too few representative species with genomic datasets; a (iii) models of evolution that fail to account for homoplasic positions. In this study, we addressed these possible sources of incongruence by (i) sequencing the transcriptome of four centrohelid lineages, (ii) using a considerable amount of newly available taxon diversity, and (iii) reducing non-phylogenetic signal by removing fast-evolving sites and applying site-heterogeneous models of evolution in both ML and Bayesian frameworks.

Our analyses recovered Haptista with maximal support, regardless of the dataset or the model used, strongly

confirming that centrohelids share a direct common ancestry with haptophytes [4,11]. For the deeper relationships among eukaryotic groups, we found that a greater taxon diversity together with the systematic use of site-heterogeneous models, allowing us to take into account site-specific substitution patterns (C60 mixture and CAT models), improves the general statistical confidence of the tree. When combined with a less noisy dataset (removal of the fastest-evolving sites), these models converged towards a similar picture in both ML and Bayesian frameworks (figure 2). In this tree, Haptista are closely related to the SAR assemblage with high support, in agreement with weaker results based on lower taxon diversity and different models [25,48]. Another relationship to receive strong support for the first time is the grouping of Archaeplastida with Cryptista. This affinity between Archaeplastida and Cryptista has been noted before in several nuclear [25–27,48] and mitochondrial-based [42] phylogenomic investigations, as well as in many 18S rDNA molecular studies [6], but unlike here, it never received significant support. Taken together, the affinities of Cryptista to Archaeplastida and of Haptista to SAR further diminish the support for Hacrobia, which was initially a less controversial assemblage when poorer taxon sampling was available [10,20,21,24]. Even though recent phylogenomic analyses continued to show a monophyletic Hacrobia, this was with no support [11,22], or with better confidence only when a large part of the diversity was removed [11].

One group of Cryptista (the cryptomonads) includes lineages with plastids of red algal origin (see below), which may confound our ability to discriminate vertically inherited genes from endosymbiotically derived ones. Indeed, it is at face value possible that the phylogenetic relationship between Cryptista and Archaeplastida observed here and elsewhere [25–27,48] is due to undetected red algal genes in phylogenomic datasets, rather than common ancestry. This is formally possible, because endosymbiotic gene transfer (EGT) is common during endosymbiosis, but there are several reasons to suggest this is not affecting our results. First, if it was the case that large numbers of unrecognized red algal genes invaded eukaryotic genomes after endosymbiosis, then one would expect all red algal plastid-containing lineages to contain many such genes, and accordingly, all be attracted to Archaeplastida, not only cryptomonads. Second, large-scale investigations of EGT in various eukaryotes (including the whole genome of the cryptomonad *Guillardia theta*) have shown that the endosymbiotic contribution to the host genome, although real, is probably less substantial than originally envisioned [49–52]. Third, phylogenomic datasets usually consist of highly expressed housekeeping genes that show no sign of widespread red algal signal. Careful inspection of our dataset allowed us to detect various contamination in different lineages, but not specifically from red algae, and suspicious topologies were not included, as in the case of the translation elongation factor 2 [53]. Overall, we observed no genes in our dataset that individually showed a strong affinity between Cryptista and red algae, suggesting that this relationship is a reflection of vertical inheritance rather than owing to a cryptic contamination of endosymbiont genes.

(b) Implications for plastid evolution

Beyond these taxonomic considerations, the positions of centrohelids, haptophytes and Cryptista in the tree of eukaryotes have important implications for how we interpret some major evolutionary and ecological transitions in eukaryotic history. The groups investigated here and their relationships to the SAR and Archaeplastida supergroups represent a complex mixture of photosynthetic and heterotrophic eukaryotes, as well as lineages for which we have little evidence as to whether they harbour a plastid or not [13,54]. Many of these lineages possess plastids bounded by three or four membranes, which are the result of eukaryote-to-eukaryote endosymbioses where heterotrophic organisms acquired plastids from red algae [55]. What makes the evolution of complex red plastids so hard to decipher is the apparent discrepancy between plastid and host phylogenies. Plastid phylogenies have generally been consistent with the notion that all red plastids are the product of a single secondary endosymbiosis [17,56–58]. This idea of a single origin was first formalized in the chromalveolate hypothesis, which posited that there was a single engulfment of a red alga in a common ancestor of stramenopiles, haptophytes, cryptophytes and alveolates [59]. Host-derived phylogenies, on the other hand, have generally failed to provide any strong evidence that all red-algal-containing lineages (and their associated plastid-lacking relatives) are monophyletic, which is required under the single endosymbiotic origin scenario. However, host phylogenies have thus far not provided any convincing alternative topologies either, making it difficult to see how plastid and host data can be best reconciled.

In this context, our work can help us understand the evolution of red plastids. Specifically, the strongly supported grouping of Archaeplastida and Cryptista *de facto* rules out the scenario of a single red plastid origin in a hypothetical ancestor of a unified chromalveolate assemblage (figure 3*a*). As stated above, the lack of support for the monophyletic origin of red plastids from host data is not new, but this is the first time, to the best of our knowledge, that a phylogenetic tree strongly argues against it. Indeed, had Cryptista branched elsewhere in the tree, a single origin of chromalveolate plastids could be explained by positing additional plastid loss events, however likely that may be. However, because Cryptista branches with the same lineage from which the plastid is derived (i.e. Archaeplastida), a single origin of red plastids is formally impossible, because those plastids would have needed to travel backwards in time to result in this topology.

With what are now robust relationships for both plastids and hosts, how can we best reconcile their apparent conflictual topologies? Two main scenarios exist to explain the origin and present distribution of complex red plastids: (i) independent secondary endosymbioses (figure 3*b*) and (ii) a unique secondary endosymbiosis followed by additional layers of endosymbioses (i.e. tertiary or quaternary; figure 3*c*). Even though the first scenario of independent endosymbioses involving different red algae could explain the tree topologies, such a model is unlikely in the light of several other pieces of evidence showing that all or substantial subsets of the 'chromalveolate' plastids trace back to a single secondary red algal endosymbiont [23,60–62]. Lately, the second scenario of serial endosymbioses (figure 3*c*) has received increased attention, being now supported by a growing body of empirical data [48,63]. Several versions of this serial endosymbiotic framework for red plastid evolution have been proposed, all involving the idea of one secondary endosymbiosis with a red alga, followed by subsequent eukaryote-to-eukaryote endosymbioses [48,62,64,65]. Recently, an explicit model was

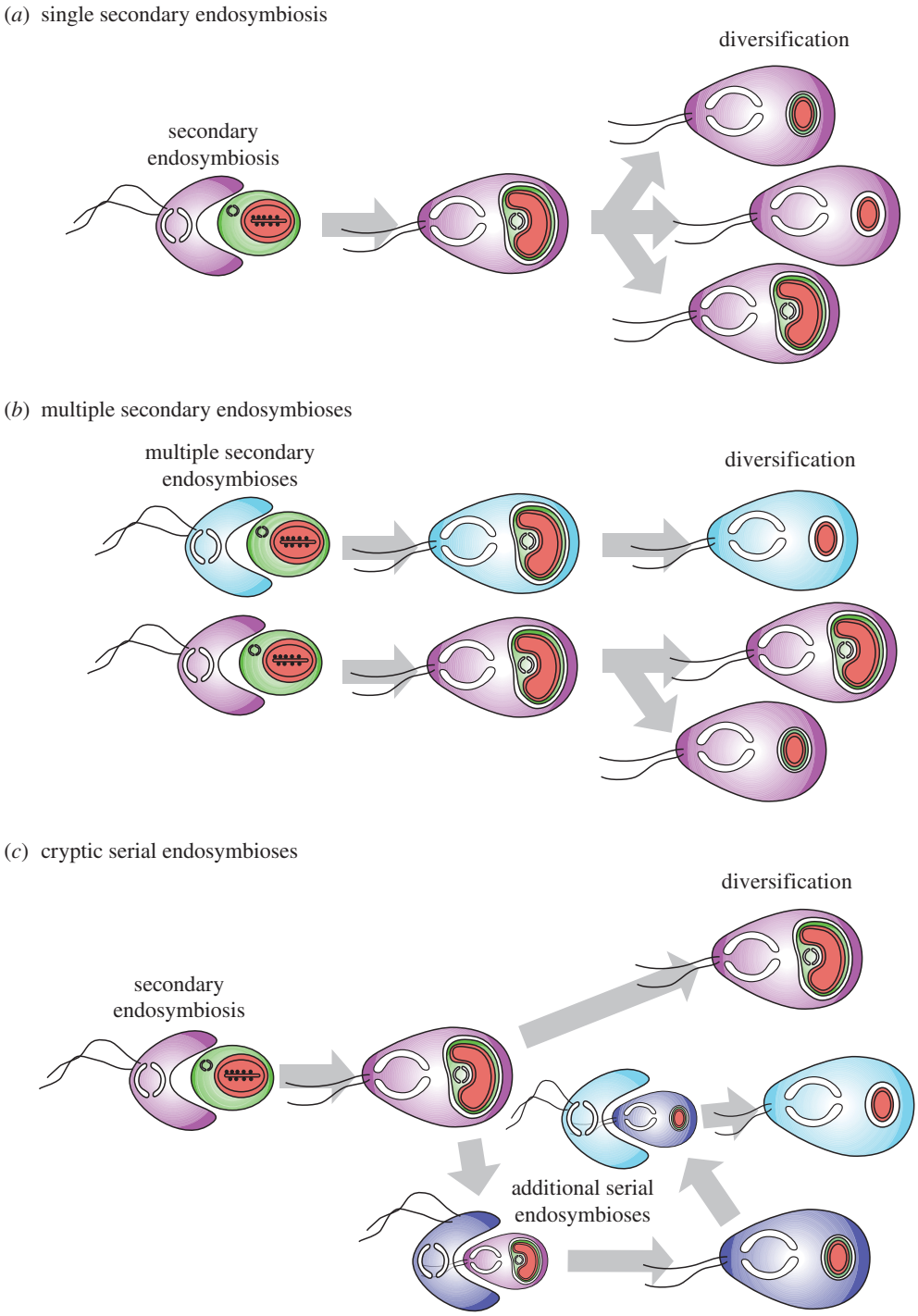

Figure 3. Scenarios for the origin and evolution of complex red plastids. These scenarios do not refer to any specific taxa, but rather illustrate the various possibilities discussed in the text, and show that the same diversity of plastid types can be generated by different combinations of events. (*a*) A single secondary endosymbiosis in the ancestor of all red plastid-bearing eukaryotes was followed only by descent with modification, as formalized in the chromalveolate hypothesis; this scenario is not supported by the current analyses. (*b*) Multiple independent secondary endosymbioses take place with different red algal symbionts, followed by descent with modification; this is compatible with current phylogenetic evidence from hosts, but not with evidence from plastids. (*c*) A single secondary endosymbiosis takes place, but is followed by serial eukaryote-to-eukaryote endosymbioses; several versions of this scenario have been proposed (see text for references), and they are consistent with current phylogenetic data.

devised using regression analyses to measure the expected similarity between genomes of various 'chromalveolate' lineages [63]. This approach resulted in a model where cryptophytes first engulfed a red alga, which was then transferred to the ochrophytes by tertiary endosymbiosis, and to the haptophytes by quaternary endosymbiosis [63].

In this context, our results are compatible with such a 'cryptophyte-first' model, although we note that phylogenetic lines of evidence are not compelling by themselves. More generally, our results will need to stand the test of time, as even strongly supported trees can be shown to be misleading with additional data. Moreover, the breadth for plastid genome data has now been far exceeded by nuclear data, so that it is likely that changes to the plastid tree will occur after the addition of new sequences, as recently demonstrated [58]. All of this could ultimately affect our interpretation, but more importantly various kinds of data, not only phylogenetics, will be needed to validate a particular model. Plastids are cellular structures of great complexity that have integrated with their hosts in many ways [54,66]. Serial endosymbiosis is

currently known for certain only in a few dinoflagellate lineages, whose endosymbionts display peculiar ways of integrating that are very different from what we observe in lineages like haptophytes, ochrophytes or most alveolates [67–69]. Thus, an integrative model of plastid evolution will need to explain many aspects to be comprehensive, from phylogeny to genetics to fine cellular processes.

5. Concluding remarks

Our centrohelid transcriptomes fill an important diversity gap in genomic sequencing. In the near future, effort should be made to provide better-quality datasets for taxa that are still evolutionary mystery but are essential to further resolve the tree; telonemids and Picozoa represent obvious targets near to the organisms studied here, but many other enigmatic microbial eukaryotes probably affect other parts of the tree in similar ways. More work is also necessary to determine the relative position of Cryptista to the Archaeplastida lineages in order to assess the monophyletic origin of the primary plastids.

Data accessibility. Raw reads are available through GenBank sequence read archive: SRR2170621, SRR2170625, SRR2170626, SRR2170627, SRR2170634.

Assembled transcriptomes: Dryad data depository (http://datadryad.org) accession http://dx.doi.org/10.5061/dryad.rj87v.

Untrimmed sequences, trimmed alignments and single-gene trees: Dryad data depository (http://datadryad.org) accession http://dx.doi.org/10.5061/dryad.rj87v.

Authors' contributions. F.B. designed the study, participated in the dataset construction, carried out the phylogenetic analyses and drafted the manuscript; M.K. participated in the dataset construction; D.V.T. and V.Z. collected field samples, established cultures, carried out molecular laboratory work and drafted the manuscript; L.V.R. collected field samples and established cultures; B.Q.M. carried out phylogenetic analyses and drafted the manuscript; A.S. and A.P.M. participated in the design of the study and critically revised the manuscript; P.J.K. designed the study and drafted the manuscript. All authors gave final approval for publication.

Competing interests. The authors declare no competing interests.

Funding. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada, and by a grant from the Tula Foundation to the Centre for Microbial Diversity and Evolution. This work was also partially supported by the Russian Foundation for Basic Research (no. 14-04-00554, 15-34-20065, 15-29-02518, 15-04-18101_a) and by a grant from the President of Russian Federation MK-7436.2015.4. The work of D.V.T. was supported by the Russian Science Foundation (no 14-14-00515). B.Q.M. acknowledges financial support to Arndt von Haeseler from the University of Vienna and the Medical University Vienna.

Acknowledgements. We thank Compute/Calcul Canada for computing resources and assistance, in particular WestGrid's Orcinus and Calcul Quebec's Guillimin and Colosse facilities.

References

1. Delsuc F, Brinkmann H, Philippe H. 2005 Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* **6**, 361–375. (doi:10.1038/nrg1603)
2. Burki F. 2014 The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb. Perspect. Biol.* **6**, a016147. (doi:10.1101/cshperspect.a016147)
3. Mikrjukov KA. 2000 System and phylogeny of Heliozoa: should this taxon exist in modern systems of protists? *Zool Z.* **79**, 883–897.
4. Cavalier-Smith T, Heyden von der S. 2007 Molecular phylogeny, scale evolution and taxonomy of centrohelid heliozoa. *Mol. Phylogenet. Evol.* **44**, 1186–1203. (doi:10.1016/j.ympev.2007.04.019)
5. Cavalier-Smith T. 2003 Protist phylogeny and the high-level classification of Protozoa. *Eur. J. Protistol.* **39**, 338–348. (doi:10.1078/0932-4739-00002)
6. 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**, 8066–8071. (doi:10.1073/pnas.0308602101)
7. Sakaguchi M, Nakayama T, Hashimoto T, Inouye I. 2005 Phylogeny of the Centrohelida inferred from SSU rRNA, tubulins, and actin genes. *J. Mol. Biol.* **61**, 765–775. (doi:10.1007/s00239-005-0006-6)
8. Nikolaev SI, Berney C, Petrov NB, Mylnikov AP, Fahrni JF, Pawlowski J. 2006 Phylogenetic position of *Multicilia marina* and the evolution of Amoebozoa. *Int. J. Syst. Evol. Microbiol.* **56**, 1449–1458. (doi:10.1099/ijs.0.63763-0)
9. 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. (doi:10.1016/j.gene.2007.09.003)
10. Burki F *et al.* 2009 Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, telonemia and centroheliozoa, are related to photosynthetic chromalveolates. *Genome Biol. Evol.* **1**, 231–238. (doi:10.1093/gbe/evp022)
11. Cavalier-Smith T, Chao EE, Lewis R. 2015 Multiple origins of Heliozoa from flagellate ancestors: new cryptist subphylum Corbihelia, superclass Corbistoma, and monophyly of Haptista, Cryptista, Hacrobia and Chromista. *Mol. Phylogenet. Evol.* **93**, 331–362. (doi:10.1016/j.ympev.2015.07.004)
12. Cavalier-Smith T, Chao EE-Y. 2003 Molecular phylogeny of centrohelid heliozoa, a novel lineage of bikont eukaryotes that arose by ciliary loss. *J. Mol. Biol.* **56**, 387–396. (doi:10.1007/s00239-002-2409-y)
13. Archibald JM. 2015 Genomic perspectives on the birth and spread of plastids. *Proc. Natl Acad. Sci. USA* **112**, 10 147–10 153. (doi:10.1073/pnas.1421374112)
14. Keeling PJ. 2013 The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu. Rev. Plant Biol.* **64**, 27.1–27.25. (doi:10.1146/annurev-arplant-050312-120144)
15. Mikrjukov KA, Siemensma FJ, Patterson DJ. 2000 Phylum Heliozoa. In *The Illustrated guide to the protozoa* (eds JJ Lee, GF Leedale, P Bradbury), pp. 860–871. Lawrence, KS: Society of Protozoologists.
16. Patterson DJ, Dürrschmidt M. 1987 Selective retention of chloroplasts by algivorous heliozoa: fortuitous chloroplast symbiosis? *Eur. J. Protistol.* **23**, 51–55. (doi:10.1016/S0932-4739(87)80007-X)
17. Yoon HS, Hackett JD, Pinto G, Bhattacharya D. 2002 The single, ancient origin of chromist plastids. *Proc. Natl Acad. Sci. USA* **99**, 15 507–15 512. (doi:10.1073/pnas.242379899)
18. Burki F, Shalchian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI, Jakobsen KS, Pawlowski J. 2007 Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE* **2**, e790. (doi:10.1371/journal.pone.0000790)
19. Rodriguez-Ezpeleta N, Brinkmann H, Burger G, Roger AJ, Gray MW, Philippe H, Lang BF. 2007 Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. *Curr. Biol.* **17**, 1420–1425. (doi:10.1016/j.cub.2007.07.036)
20. Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rümmele SE, Bhattacharya D. 2007 Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. *Mol. Biol. Evol.* **24**, 1702–1713. (doi:10.1093/molbev/msm089)
21. Patron NJ, Inagaki Y, Keeling PJ. 2007 Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr. Biol.* **17**, 887–891. (doi:10.1016/j.cub.2007.03.069)
22. Katz LA, Grant JR. 2014 Taxon-rich phylogenomic analyses resolve the eukaryotic tree of life and reveal the power of subsampling by sites. *Syst. Biol.* **64**, 406–415. (doi:10.1093/sysbio/syu126)

23. 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. (doi:10.1186/1741-7007-4-31)
24. Okamoto N, Chantangsi C, Horak A, Leander BS, Keeling PJ. 2009 Molecular phylogeny and description of the novel katablepharid *Roombia truncata* gen. et sp. Nov., and establishment of the *Hacrobia taxon* nov. *PLoS ONE* **4**, e7080. (doi:10.1371/journal.pone.0007080)
25. Burki F, Okamoto N, Pombert J-F, Keeling PJ. 2012 The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc. R Soc. B* **279**, 2246–2254. (doi:10.1098/rspb.2011.2301)
26. Brown MW, Sharpe SC, Silberman JD, Heiss AA, Lang BF, Simpson AGB, Roger AJ. 2013 Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proc. R. Soc. B* **280**, 20131755. (doi:10.1016/j.protis.2010.06.004)
27. Yabuki A, Kamikawa R, Ishikawa SA, Kolisko M, Kim E, Tanabe AS, Kume K, Ishida K-I, Inagki Y. 2014 *Palpitomonas bilix* represents a basal cryptist lineage: insight into the character evolution in Cryptista. *Sci. Rep.* **4**, 4641. (doi:10.1038/srep04641)
28. Keeling PJ *et al.* 2014 The marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **12**, e1001889. (doi:10.1371/journal.pbio.1001889)
29. 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. (doi:10.1186/1471-2148-7-S1-S2)
30. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)
31. Criscuolo A, Gribaldo S. 2010 BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210. (doi:10.1186/1471-2148-10-210)
32. Cummins CA, McInerney JO. 2011 A method for inferring the rate of evolution of homologous characters that can potentially improve phylogenetic inference, resolve deep divergence and correct systematic biases. *Syst. Biol.* **60**, 833–844. (doi:10.1093/sysbio/syr064)
33. Soubrier J, Steel M, Lee MSY, Der Sarkissian C, Guindon S, Ho SYW, Cooper A. 2012 The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol. Biol. Evol.* **29**, 3345–3358. (doi:10.1093/molbev/mss140)
34. Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. 2014 Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol. Biol.* **14**, 82. (doi:10.1186/1471-2148-14-82)
35. Wang H-C, Susko E, Roger AJ. 2014 An amino acid substitution–selection model adjusts residue fitness to improve phylogenetic estimation. *Mol. Biol. Evol.* **31**, 779–792. (doi:10.1093/molbev/msu044)
36. Minh BQ, Nguyen MAT, von Haeseler A. 2013 Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195. (doi:10.1093/molbev/mst024)
37. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010 New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321. (doi:10.1093/sysbio/syq010)
38. Lartillot N, Lepage T, Blanquart S. 2009 PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288. (doi:10.1093/bioinformatics/btp368)
39. Burki F, Corradi N, Sierra R, Pawlowski J, Meyer GR, Abbott CL, Keeling PJ. 2013 Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in rhizaria. *Curr. Biol.* **23**, 1541–1547. (doi:10.1016/j.cub.2013.06.033)
40. Philippe H, Brinkmann H, Copley RR, Moroz LL, Nakano H, Poustka AJ, Wallberg A, Peterson KJ, Telford MJ. 2011 Acoelomorph flatworms are deuterostomes related to Xenoturbella. *Nature* **470**, 255–258. (doi:10.1038/nature09676)
41. Lartillot N, Philippe H. 2008 Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. *Proc. R. Soc. B* **363**, 1463–1472. (doi:10.1098/rstb.2007.2236)
42. Derelle R, Torruella G, Klimeš V, Brinkmann H, Kim E, Vlček Č, Lang BF, Eliáš M. 2015 Bacterial proteins pinpoint a single eukaryotic root. *Proc. Natl Acad. Sci. USA* **112**, E693–E699. (doi:10.1073/pnas.1420657112)
43. Cavalier-Smith T. 2013 Symbiogenesis: mechanisms, evolutionary consequences, and systematic implications. *Annu. Rev. Ecol. Evol. Syst.* **44**, 145–172. (doi:10.1146/annurev-ecolsys-110411-160320)
44. Bråte J, Klaveness D, Rygh T, Jakobsen KS, Shalchian-Tabrizi K. 2010 Telonemia-specific environmental 18S rDNA PCR reveals unknown diversity and multiple marine-freshwater colonizations. *BMC Microbiol.* **10**, 168. (doi:10.1186/1471-2180-10-168)
45. Seenivasan R, Sausen N, Medlin LK, Melkonian M. 2013 *Picomonas judraskeda* gen. et sp. nov.: the first identified member of the Picozoa phylum nov., a widespread group of picoeukaryotes, formerly known as 'picobiliphytes'. *PLoS ONE* **8**, e59565. (doi:10.1371/journal.pone.0059565.s006)
46. Rodriguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007 Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst. Biol.* **56**, 389–399. (doi:10.1080/10635150701397643)
47. Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger AJ. 2009 Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic 'supergroups'. *Proc. Natl Acad. Sci. USA* **106**, 3859–3864. (doi:10.1073/pnas.0807880106)
48. Baurain D *et al.* 2010 Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol. Biol. Evol.* **27**, 1698–1709. (doi:10.1093/molbev/msq059)
49. Burki F, Imanian B, Hehenberger E, Hirakawa Y, Maruyama S, Keeling PJ. 2014 Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. *Eukaryot. Cell* **13**, 246–255. (doi:10.1128/EC.00299-13)
50. Burki F, Flegontov P, Obornik M, Cihlar J, Pain A, Lukes J, Keeling PJ. 2012 Reevaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. *Genome Biol. Evol.* **4**, 626–635. (doi:10.1093/gbe/evs049)
51. Deschamps P, Moreira D. 2012 Re-evaluating the green contribution to diatom genomes. *Genome Biol. Evol.* **4**, 795–800. (doi:10.1093/gbe/evs053)
52. Curtis BA *et al.* 2012 Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* **492**, 59–65. (doi:10.1038/nature11681)
53. Kim EE, Graham LE. 2008 EEF2 analysis challenges the monophyly of Archaeplastida and Chromalveolata. *PLoS ONE* **3**, e2621. (doi:10.1371/journal.pone.0002621)
54. Archibald JM. 2009 The puzzle of plastid evolution. *Curr. Biol.* **19**, R81–R88. (doi:10.1016/j.cub.2008.11.067)
55. Bhattacharya D, Archibald JM, Weber APM, Reyes-Prieto A. 2007 How do endosymbionts become organelles? Understanding early events in plastid evolution. *Bioessays* **29**, 1239–1246. (doi:10.1002/bies.20671)
56. Qiu H, Yang EC, Bhattacharya D, Yoon HS. 2012 Ancient gene paralogy may mislead inference of plastid phylogeny. *Mol. Biol. Evol.* **29**, 3333–3343. (doi:10.1093/molbev/mss137)
57. Janouskovec J, Horak A, Obornik M, Lukes J, Keeling PJ. 2010 A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc. Natl Acad. Sci. USA* **107**, 10 949–10 954. (doi:10.1073/pnas.1003335107)
58. Ševčíková T *et al.* 2015 Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Sci. Rep.* **5**, 10134. (doi:10.1038/srep10134)
59. 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. (doi:10.1111/j.1550-7408.1999.tb04614.x)
60. Zimorski V, Ku C, Martin WF, Gould SB. 2014 Endosymbiotic theory for organelle origins. *Curr. Opin. Microbiol.* **22**, 38–48. (doi:10.1016/j.mib.2014.09.008)
61. Stork S, Moog D, Przyborski JM, Wilhelmi I, Zauner S, Maier UG. 2012 Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent

translocation from degradation. *Eukaryot Cell* **11**, 1472–1482. (doi:10.1128/EC.00183-12)

62. Petersen J, Ludewig A-K, Michael V, Bunk B, Jarek M, Baurain D, Brinkmann H. 2014 *Chromera velia*, endosymbioses and the rhodoplex hypothesis: plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). *Genome Biol. Evol.* **6**, 666–684. (doi:10.1093/gbe/evu043)
63. Stiller JW, Schreiber J, Yue J, Guo H, Ding Q, Huang J. 2014 The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat. Commun.* **5**, 5764. (doi:10.1038/ncomms6764)
64. Bodyl A, Stiller JW, Mackiewicz P. 2009 Chromalveolate plastids: direct descent or multiple endosymbioses? *Trends Ecol. Evol.* **24**, 119–121; author reply 121–122. (doi:10.1016/j.tree.2008.11.003)
65. Sanchez Puerta MV, Delwiche CF. 2008 A hypothesis for plastid evolution in chromalveolates. *J. Phycol.* **44**, 1097–1107. (doi:10.1111/j.1529-8817.2008.00559.x)
66. Bolte K, Bullmann L, Hempel F, Bozarth A, Zauner S, Maier UG. 2009 Protein targeting into secondary plastids. *J. Eukaryot. Microbiol.* **56**, 9–15. (doi:10.1111/j.1550-7408.2008.00370.x)
67. Keeling PJ. 2010 The endosymbiotic origin, diversification and fate of plastids. *Phil. Trans. R. Soc. B* **365**, 729–748. (doi:10.1098/rstb.2009.0103)
68. Wisecaver JH, Hackett JD. 2011 Dinoflagellate genome evolution. *Annu. Rev. Microbiol.* **65**, 369–387. (doi:10.1146/annurev-micro-090110-102841)
69. Patron NJN, Waller RF. 2007 Transit peptide diversity and divergence: a global analysis of plastid targeting signals. *Bioessays* **29**, 1048–1058. (doi:10.1002/bies.20638)