Chapter 2

The Algal Tree of Life from a Genomics Perspective

Debashish Bhattacharya*
Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ, USA

and

Dana C. Price
Department of Plant Biology, Rutgers University, New Brunswick, NJ, USA

I. Introduction

A large body of molecular, morphological, and fossil data demonstrates that primary plastids are derived from an ancient (≥1 Ga, up to 1.6 Ga; Butterfield 2000; Yoon et al. 2004; Parfrey et al. 2011; Blank 2013; Bengtson et al. 2017; Sánchez-Baracaldo et al. 2017) primary cyanobacterial endosymbiosis. This event occurred in the single common ancestor of three extant photosynthetic lineages collectively known as the Plantae, and more recently, the Archaeplastida (Cavalier-Smith 1981; Margulis 1981; Reyes-Prieto et al. 2007; Adl et al. 2005; Price et al. 2012; Cavalier-Smith 2017). These lineages include the Glaucophyta (glaucophyte algae), the Rhodophyta (red algae), and the Viridiplantae (green algae and land plants) that share a two-membrane bound photosynthetic plastid organelle. Once established in the Archaeplastida ancestor, the primary plastid spread to other lineages, including the SAR clade (stramenopiles [e.g. diatoms, kelps, plastid-lacking oomycetes] + Alveolata [dinoflagellates, ciliates, apicomplexans] + Rhizaria [e.g., chlorarachniophyte algae]), cryptophytes, haptophytes, and the

*Author for correspondence, e-mail: d.bhattacharya@rutgers.edu

A. W. D. Larkum et al. (eds.), *Photosynthesis in Algae: Biochemical and Physiological Mechanisms*, Advances in Photosynthesis and Respiration 45, https://doi.org/10.1007/978-3-030-33397-3_2

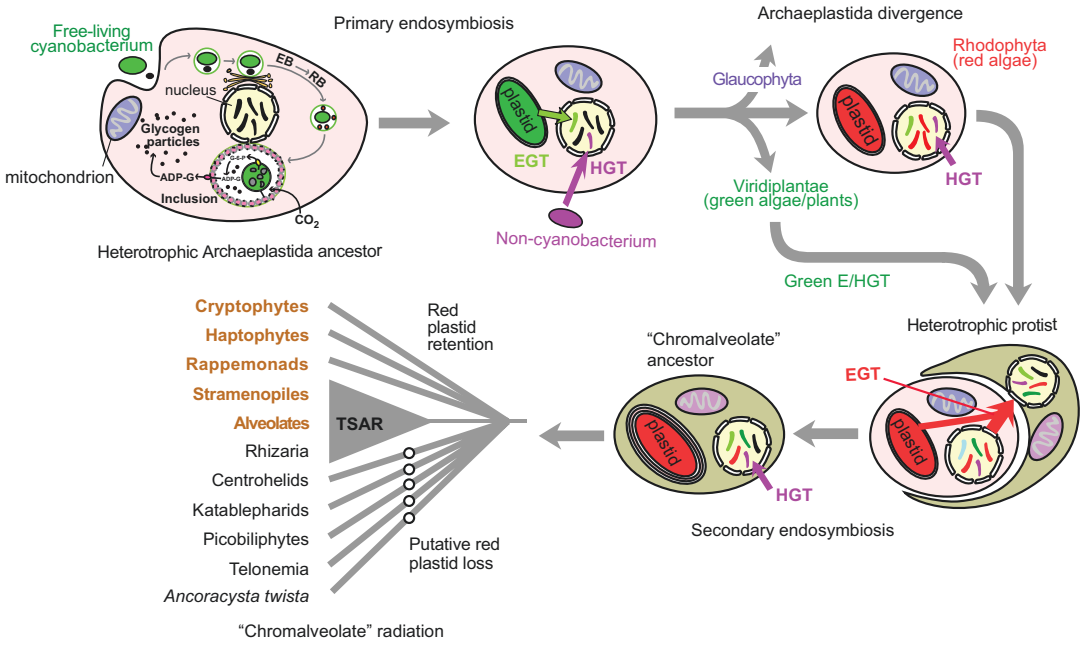

Fig. 2.1. The proposed history of plastid endosymbiosis in photosynthetic Archaeplastida and "chromalveolate" taxa. The primary cyanobacterial endosymbiosis, including contribution by chlamydial cells under the MATH is shown in the top left of this figure. Both EGT and HGT occurred throughout the history of Archaeplastida, prior to and after the split of its constituent phyla. A red alga was captured by the "chromalveolate" ancestor that may have been defined by the telonemid-SAR (TSAR) joint lineage (see Section III, below) and potentially including cryptophytes and haptophytes. There is evidence that this red algal secondary endosymbiosis was preceded by a cryptic green algal capture and subsequent loss of the organelle, leaving behind dozens to hundreds of green genes in the nucleus of diatoms and other chromalveolates (Moustafa et al. 2009; Dorrell et al. 2017). This complex series of gene transfer events was also added to by independent HGTs from external prokaryotes and eukaryotes. Given this scenario for origin of the plastid in most algal groups, it is not surprising that genomic data from these taxa provide reticulate phylogenetic signals when genes are analyzed individually or in groups, as described in the text. Image based on Qiu et al. (2013) and Brodie et al. (2017)

euglenids through multiple secondary and tertiary endosymbioses (Fig. 2.1). Many of these taxa that contain a red alga-derived plastid are colloquially referred to as "chromalveolates", a now defunct (i.e., polyphyletic) taxon that was hypothesized by Cavalier-Smith (1999) to share a single secondary endosymbiosis. Therefore, current data thus suggest that virtually all photosynthetic forms on our planet ultimately owe their photoautotrophic ability to a single cyanobacterial source. The sole exception to this rule is the clade of photosynthetic amoebae, *Paulinella*, to be described below that provides the only known case of an independent plastid primary endosymbiosis.

II. Why Inferring the Algal Tree of Life Is Non-trivial

Although of central importance to marine ecosystems and terrestrial life, the conversion of solar energy into carbohydrates and lipids through photosynthesis came at a high cost to photosynthetic cells. Light harvesting can capture excess energy that must be eliminated (mostly as heat), and photosynthetic electron flow is accompanied by the formation of reactive oxygen species (ROS) that can impair cellular functions (Peers et al. 2009; Knoefler et al. 2012). Therefore, the first algae, and every subsequent host of a serial plastid endosymbiosis depicted in Fig. 2.1 had to

cope with these challenges and integrate the flow of fixed carbon across cell compartments (Linka and Weber 2010; Karkar et al. 2015). These cells also needed to adapt to diurnal changes in light intensity, temperature, water and nutrient availability, and competition from other protists and predators to survive. These selective pressures necessitated major innovations, not only through mutation and gene duplication but also the acquisition of foreign genes from the endosymbiont *via* endosymbiotic gene transfer (EGT) as well as from external prokaryotic and eukaryotic sources through horizontal gene transfer (HGT) (Fig. 2.1). In addition, protein domains encoded by cyanobacterial (endosymbiont) genes were mixed and matched with domains from other genes to give rise to chimeric symbiogenetic (S)-genes with novel roles. Many of these S-gene functions evolved to deal with redox stress and light sensing to support the novel organelle (Méheust et al. 2016). An important, and unexpected perspective on how complex biotic interactions underlie plastid origin is offered by recent work done on the contribution of chlamydial genes to Archaeplastida.

The chlamydial connection is summarized under the ménage à trois hypothesis (MATH) that suggests a direct role for Chlamydiales obligate intracellular pathogens in plastid establishment. This idea is supported by the finding of 30–100 genes of chlamydial derivation in Archaeplastida that are involved in a range of key functions such as glycogen, tryptophan, and menaquinone metabolism (Ball et al. 2013, 2016; Qiu et al. 2013; Cenci et al. 2017, 2018). As shown in Fig. 2.1, under the MATH, the chlamydial infectious particle (EB: elementary body, black circle) entered the Archaeplastida host together with a free-living cyanobacterium (green circle). The EB remodeled the phagocytic membrane into a chlamydia-controlled inclusion and differentiated into reticulate bodies (RBs; pink circles) that attached to the inclusion and secreted chlamydial effector proteins corresponding to glycogen metabolism enzymes into both the inclusion and the host cytosol. Within the inclusion, the cyanobacterial endosymbiont is believed to have recruited chlamydial transporters via conjugation with the pathogen to facilitate export of glucose-6-phosphate (G-6-P) through the UhpC transporter of chlamydial origin (orange circle). This sugar phosphate was utilized for glycogen synthesis in the inclusion and excess ADP-G was released to the cytosol *via* a nucleotide sugar transporter (magenta circle) of eukaryotic origin. These processes led to the initial survival of the unprotected cyanobacterial endosymbiont in the chlamydial inclusion, precipitated gene transfers between compartments, and the integration of carbon flux that led to permanent plastid maintenance. Once the chlamydial cell was lost, the only "footprints" that remain of this hypothetical scenario are dozens of pathogen-derived HGTs with plastid-related functions. Consistent with the idea that EGTs, HGTs, and redirection of host-encoded proteins are critical to organellogenesis are the findings regarding evolution of the novel plastid in *Paulinella* spp. This plastidial organelle is a far younger version of the Archaeplastida organelle, having originated ca. 100 Ma (Kim et al. 2014).

Paulinella chromatophora and *P. micropora* are filose amoebae (Bhattacharya et al. 1995) with blue-green chromatophores (plastids). *P. chromatophora* was described in 1895 by Robert Lauterborn (Lauterborn 1895) and the photosynthetic *Paulinella* lineage is the only known case of an independent primary (alpha-cyanobacterial) plastid acquisition (Kies 1974; Marin et al. 2007; Yoon et al. 2009), making them models for understanding plastid establishment. The chromatophore genome is highly reduced in size and gene content (ca. 850 protein coding genes) relative to cyanobacterial genomes (Nowack et al. 2008; Yoon et al. 2009; Reyes-Prieto et al. 2010). Recent work shows that dozens of bacterial genes have been recruited to support lost organelle functions (due to Muller's ratchet acting on this non-

recombining DNA) (Nowack et al. 2016; DB and DCP unpublished data) as well as the retargeting of host proteins through a novel sorting pathway (Nowack and Grossman 2012; Singer et al. 2018). These results demonstrate that foreign gene transfer to the host nucleus is key in compensating for organelle genome reduction and suggests that phagotrophy (i.e., photosynthetic *Paulinella* are derived from a phagotrophic lineage; Bhattacharya et al. 2012) was retained early on in endosymbiosis to facilitate HGT, presumably via the ingestion of prey cells.

A final example of genetic complexity associated with endosymbiosis is provided by the work of Moustafa et al. (2009) who determined the phylogenetic origins of proteins encoded in the model diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*, and found several hundred of green algal provenance shared by SAR and other "chromalveolates". These results suggested a cryptic (again, missing compartment) green algal endosymbiosis in the chromalveolate ancestor prior to the capture of the widespread red algal plastid in these taxa. This idea was tested and found wanting by some (e.g., Deschamps and Moreira 2012) but more recent work using a richer collection of genomes, with a focus on plastid proteomes (Dorrell et al. 2017), provided strong support for the original Moustafa et al. (2009) hypothesis. Therefore, accurately inferring algal relationships with the ETOL is not a trivial problem. Beyond testing Archaeplastida monophyly, the chromalveolate taxa have likely undergone serial plastid endosymbioses (e.g., Stiller et al. 2014) and sporadic HGTs over their >1 billion evolutionary history that will invariably muddy the waters (i.e., due to reticulate gene histories) when inferring phylogenetic relationships. Even in instances where Archaeplastida are found to be monophyletic, which is often the case in multigene trees (e.g., Rodríguez-Ezpeleta et al. 2005; Parfrey et al. 2011; Burki et al. 2016), and more robustly when using biochemical and metabolic pathway data (the MATH; Price et al. 2012), the spread of red and green genes in chromalveolates due to EGT and HGT will pull the Archaeplastida apart when included in multi-gene phylogenies. The issue of ancient algal gene transfer complicating ETOL inference was succinctly described by Hackett et al. (2007) when they first provided evidence of SAR monophyly, and has hounded reconstruction of the algal tree of life ever since.

III. Examples of Reticulate Behavior Among Algal Genes

In spite of the issues described above, many nodes in the broader ETOL have been solved, or at least well-supported using a "designer set" of 187 (Cavalier-Smith et al. 2015, 2016) to over 200 concatenated genes that have been manually checked to circumvent EGT/HGT and paralogy artifacts (e.g., 263 genes by Irwin et al. 2018; 248 genes by Strassert et al. 2019). Many of these studies have consolidated algal (and related non-photosynthetic [e.g., *Ancoracysta twista*]) groups (SAR; Burki et al. 2016; Janouškovec et al. 2017), and in others, brought them into question (Archaeplastida; Strassert et al. 2019). Of particular focus in these analyses are the cryptophytes and haptophytes that have been reported in several different positions in trees. Haptophytes were once identified as members of the novel clade 'Hacrobia' (Okamoto et al. 2009) that includes cryptophytes and other lineages such as telonemids and centrohelids (Burki et al. 2009), katablepharids and perhaps picobiliphytes (Okamoto et al. 2009; Yoon et al. 2011). However, the interrelationships of Hacrobia were unresolved (Okamoto et al. 2009), and later phylogenomic analyses refuted Hacrobia monophyly, placing haptophytes as sister to the SAR group (Baurain et al. 2010; Burki et al. 2012) together with telonemids and centrohelids (Burki et al. 2012) or in a later permutation as sister to centrohelids as

part of the 'Haptista' (Burki et al. 2016). The most recent work supports Archaeplastida paraphyly with cryptophytes sister to red algae (Adl et al. 2018), telonemids sister to SAR (the so-called "TSAR" lineage), and haptophytes sister to the Archaeplastida + Cryptophyta clade (Strassert et al. 2019). Based on these results, we can reasonably conclude that the phylogenetic position of SAR is likely to be well-established, yet despite a mountain of biochemical data, Archaeplastida monophyly is surprisingly ambiguous and the position of haptophytes (vis-à-vis cryptophytes and other protist lineages) within the ETOL remains unclear. These issues need to be resolved with the use of significantly more genome data from these taxa and perhaps novel approaches to the ETOL problem. Another obvious question to ask is whether, despite manual checking of designer gene sets, do these alignments contain sufficient phylogenetic power to resolve >1.6 billion-year-old divergences or alternatively, contain hidden signal of algal EGT/HGT (Hackett et al. 2007) that makes the resulting trees unstable?

The latter issue is real and can be shown using the "problematic" cryptophytes as an example of how single nuclear genes in algal genomes may contain highly complex phylogenetic signals. Figures 2.2 and 2.3 show maximum likelihood IQ-TREE analyses (Nguyen et al. 2015) with ultrafast bootstrap (UFB) approximations at nodes (2000 replicates; Minh et al. 2013) of 5 different genes encoded by cryptophytes and other ETOL lineages. The specific methods used to generate these alignments and trees are described in Price and Bhattacharya (2017) and incorporate the extensive MMETSP transcriptome data (Keeling et al. 2014) to expand taxonomic sampling, specifically among chromalveolates. Figure 2.2a presents the tree of a conserved 26S proteasome regulatory complex subunit that is involved in protein degradation. This analysis provides moderate support for the Hacrobia clade (UFB 78%) and most ETOL phyla are well supported in this robust phylogeny. Incidentally, a tree built using 88 concatenated plastid proteins also recovered Hacrobia monophyly with high RAxML bootstrap support (100%; Kim et al. 2017). In contrast, Fig. 2.2b shows the tree of a conserved SURF1 protein that is putatively involved in the biogenesis of cytochrome c oxidase and provides a very different view of cryptophyte evolution. In this tree, cryptophytes are weakly affiliated with red algae (UFB 67%) and haptophytes are sister to stramenopiles (UFB 77%), with several taxon misplacements likely due to cross-contamination in the EST data or mislabeling of MMETSP samples (e.g., *Madagascaria erythrocladiodes*). Regardless, the gene encoding SURF1 might provide an example of EGT from the red algal plastid endosymbiont in cryptophytes. In Fig. 2.2c, we find yet another phylogenetic scenario, whereby the U3 small nucleolar ribonucleoprotein IMP3 involved in 18S rRNA biogenesis splits the cryptophytes into two clades. One includes the aplastidial *Goniomonas pacifica* and green algae (UFB 77%), whereas the second photosynthetic clade is grouped with a red alga (UFB 64%) and other protists. Similar results are reported in Fig. 2.3a, in which a tree made from a hypothetical protein containing a domain of unknown function (DUF866) suggests cryptophyte polyphyly, showing the photosynthetic taxon grouping with red algae (UFB 90%) and haptophytes strongly affiliated with stramenopiles (UFB 97%). The final example tree (Fig. 2.3b) of a putative D123 protein involved in the cell division cycle supports the monophyly of photosynthetic cryptophytes and glaucophytes (UFB 82%) and the affiliation of haptophytes and stramenopiles (UFB 85%). It should be noted that these trees represent only a tiny fraction of the 1000s of phylogenies that we have generated. These examples provide evidence that single genes may each tell a unique story of algal evolution that merits attention, yet are clearly confounded by issues such as insuffi-

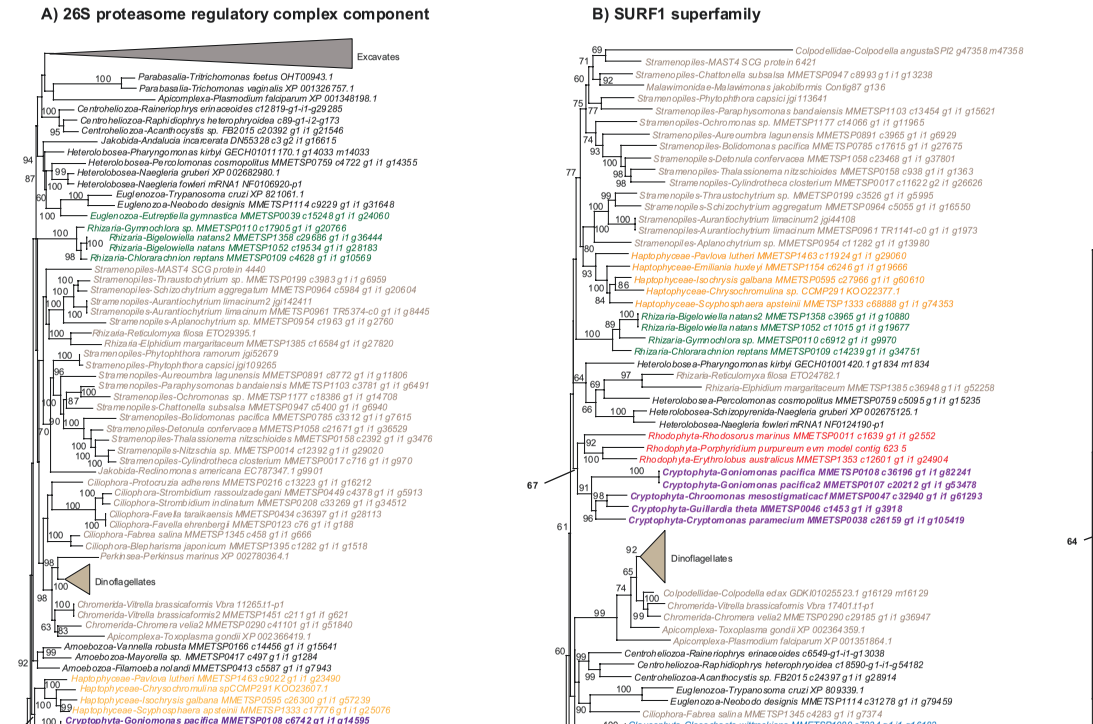

Fig. 2.2. Phylogenies of three proteins implicated in alga-derived EGT or HGT in "chromalveolates". (**a**) Phylogeny of a 26S proteasome regulatory complex component, (**b**) SURF1 superfamily protein, and (**c**) U3 small nucleolar ribonucleoprotein IMP3, inferred using IQ-TREE. The results of 1000 ultrafast bootstraps are shown at the branch nodes (when ≥60%), and the legends for substitution rates on branches are shown. Archaeplastida are shown in red (Rhodophyta), green (Viridiplantae), and light blue (Glaucophyta) text. SAR members are in brown text, cryptophytes in purple, haptophytes in orange, and photosynthetic chlororachniophytes in dark green. Dinoflagellates are summarized with the brown triangle. NCBI or MMETSP identifications are shown for each of the sequence entries

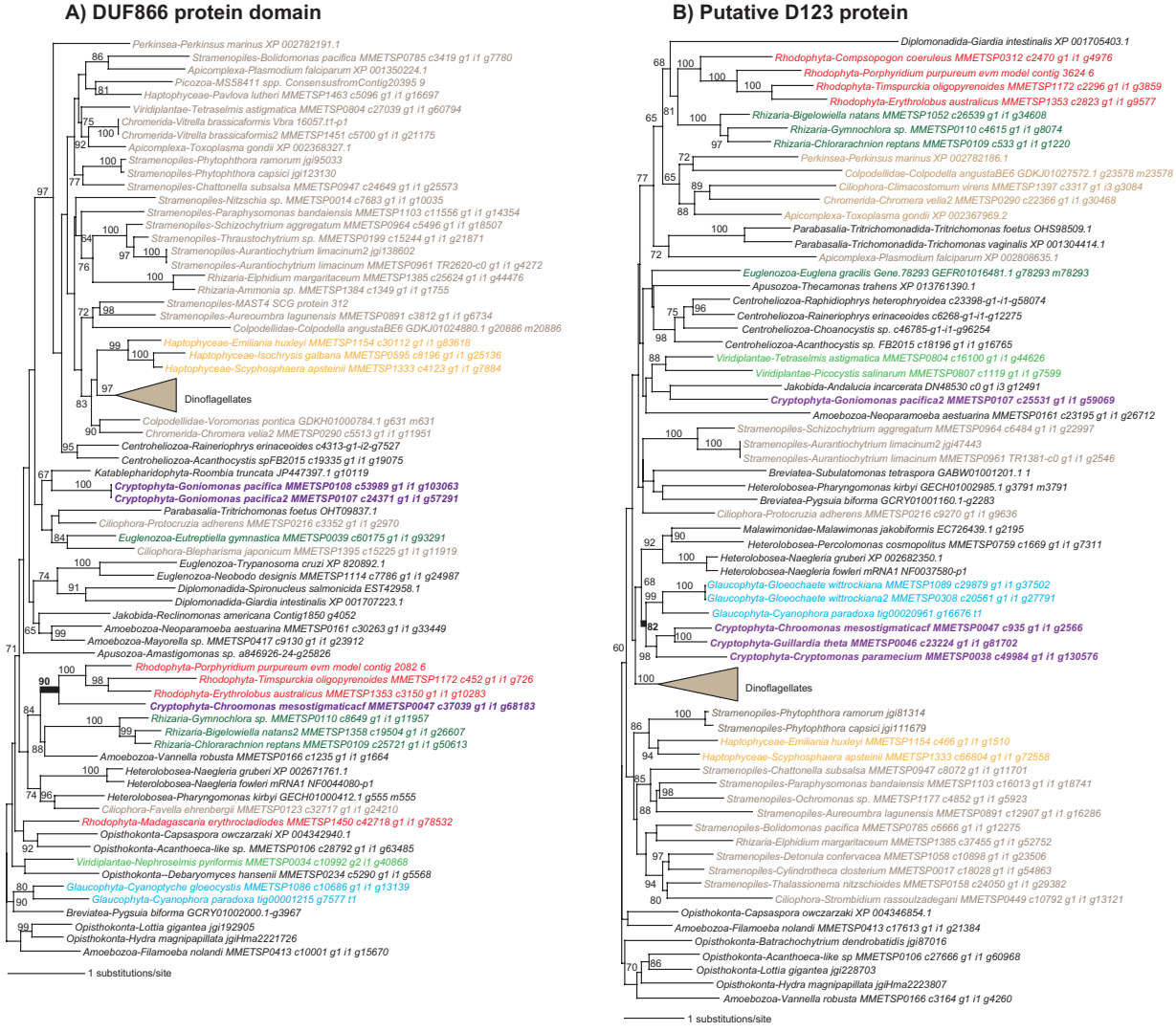

Fig. 2.3. Phylogenies of two proteins implicated in alga-derived EGT or HGT in "chromalveolates". (**a**) Phylogeny of a DUF866 protein domain containing sequence, and (**b**) putative D123 protein, inferred using IQ-TREE. The results of 1000 ultrafast bootstraps are shown at the branch nodes (when ≥60%), and the legends for substitution rates on branches are shown. Archaeplastida are shown in red (Rhodophyta), green (Viridiplantae), and light blue (Glaucophyta) text. SAR members are in brown text, cryptophytes in purple, haptophytes in orange, and photosynthetic chlororachniophytes in dark green. Dinoflagellates are summarized with the brown triangle. NCBI or MMETSP identifications are shown for each of the sequence entries

cient phylogenetic signal in single proteins, incomplete taxon sampling, paralog gains/losses, contamination, MMETSP taxon mislabeling (obvious cases are shown), or a combination of these factors. Nonetheless, single genes are the basis of phylogenetic inference and it is important to recognize their limitations prior to generating complex concatenated datasets to infer the ETOL.

IV. From Designer Datasets to Whole Genomes

Given the uncertainty associated with algal placements in the ETOL shown in Figs. 2.2 and 2.3 and previous studies, we chose to use another approach to this problem. Rather than trying to identify the "best set" of genes based on parameters such as length, conservation, paralogy, absolute distribution, evidence of EGT or HGT, we built a bioinformatic pipeline that follows a few simple rules and is fed predicted proteins from over a hundred genomic data sets from which a massive alignment is built, and an IQ-TREE inferred. The approach is described in Price and Bhattacharya (2017) and involves deriving de novo ortholog groups (OGs) to construct, in the example shown here, a 3000 OG dataset from 115 publicly available eukaryote proteomes. In brief, EST and/or predicted proteome data were retrieved for the target species and OrthoFinder (Emms and Kelly 2015) was used to construct OGs from the total data. Each group (or putative gene family) was parsed and we retained those that had low levels of paralogy (>80% of taxa were single-copy). Taxa with multi-copy representative proteins were removed from these groups, and the protein sequences corresponding to each individual group were aligned with MAFFT v. 7.3 (Katoh and Standley 2013). These alignments (summing to 2,458,432 aligned amino acids) were used to construct a maximum-likelihood phylogeny using IQ-TREE *via* a partitioned analysis in which each OG alignment represented a single partition with unlinked models of evolution chosen by IQ-TREE. Consensus tree branch support was determined by 2000 rapid bootstraps.

The phylogeny that resulted from this genome-wide approach is shown in Fig. 2.4. The position of telonemids (data not yet publicly available) is marked with an arrow based on Strassert et al. (2019). Several things relating to algae in the ETOL fall out. First, most phyla, including non-algal taxa receive strong UFB support. Archaeplastida monophyly is well-supported, with red algae as the earliest divergence. This latter result is consistent with the work of Lee et al. (2016) who studied the history of EGT among Archaeplastida and found 23 shared OGs in the plastid genomes of glaucophytes and Viridiplantae that were transferred to the nucleus from their putative common ancestor, versus only four such OGs being common to all three lineages, and only one shared OG being common to the Viridiplantae and rhodophytes. This pattern of intracellular gene movement supports the "red early" hypothesis, as depicted in Fig. 2.4. This tree also supports SAR monophyly and a common ancestry of cryptophytes, katablepharids, and picozoans with haptophytes sister to SAR. The broader story depicted in this genome-based perspective on the ETOL is that all algal groups and their non-photosynthetic sisters form a single clade in the tree (UFB 99%) that is distinct from opisthokonts, excavates, and their allies. The presence of plastid-lacking taxa at the base of cryptophytes suggests that this algal group may have undergone an independent algal secondary endosymbioses as suggested by Figs. 2.2b and 2.3a. This idea merits additional analysis given that some trees (both nuclear [Fig. 2.2a] and plastid based) favor Hacrobia monophyly. It is clear that the gene

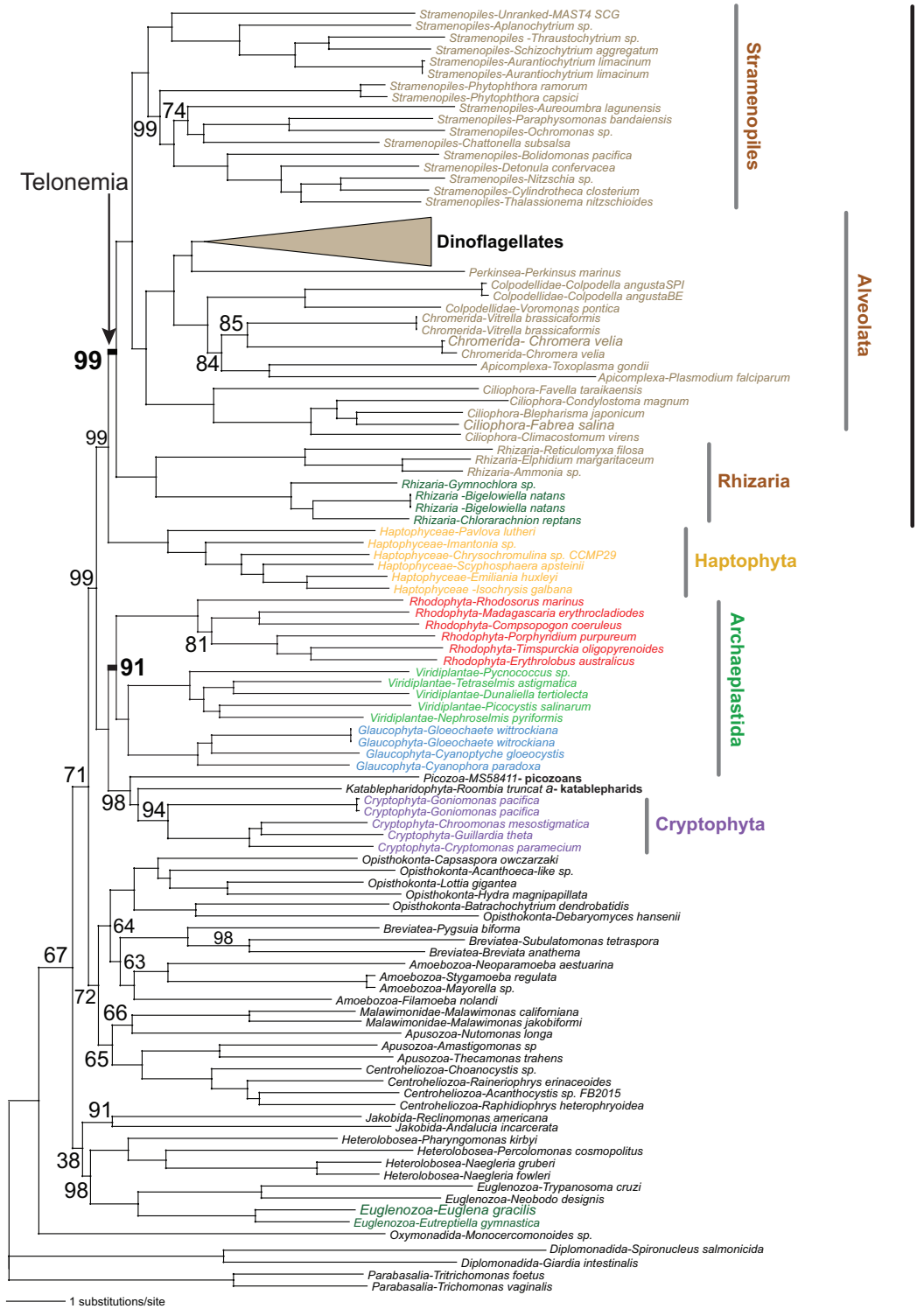

Fig. 2.4. Phylogeny built using IQ-TREE showing the positions of different algal groups in the ETOL. This tree was constructed using a partitioned 3000 OG dataset from 115 publicly available eukaryote proteomes. All nodes have 100% bootstrap support unless shown otherwise. Major algal groups are identified in the image. The putative position of telonemids is based on Strassert et al. (2019)

inventory of cryptophytes is highly chimeric in origin with *Goniomonas* species perhaps having the most complex pattern of algal EGT/HGT. This complexity notwithstanding, our current best estimate in this regard is that haptophytes are sister to SAR and telonemids.

To understand how this massive, genome-wide analysis compares to a tree inferred from a designer set of our making, we limited our dataset to OGs comprising the highly conserved odb9 (65 species; 303 OGs; 491,224 aligned amino acids) ortholog set implemented in BUSCO Eukaryota (Simão et al. 2015). The tree generated from this alignment is shown in Fig. 2.5. The topology is similar to Fig. 2.4, but appears to be more highly impacted by long-branch artifacts due to the high divergence rates among some excavates and ciliates, making these sister taxa to alveolates. Archaeplastida are again monophyletic but with glaucophytes as the earliest divergence in this clade. Hacrobia are paraphyletic with some plastid-lacking taxa being sister to haptophytes. In general, this tree receives 100% UFB support for most branches (as in Fig. 2.4) but appears to be more sensitive to divergence rate variation. This particular issue is not readily apparent in the genome-based tree.

V. Conclusions

Inferring algal phylogenetic relationships within the ETOL and generating a stable taxonomy is a vital but challenging frontier in photosynthesis research and more broadly in evolutionary biology. Once considered to be only a matter of time until all nodes are unequivocally established, the ETOL has only remained a significant problem for the fields of phylogenetics and genomics. This is because additional data have uncovered ever more complex behavior such as mixtures of photosynthetic and non-photosynthetic taxa suggesting massive plastid losses or multiple plastid gains that need to be accounted for before a "simple" framework of vertical evolution could be espoused. It is however clear that most algae are now securely placed within monophyletic groups and higher phyla such as Archaeplastida and SAR are well-established. Other orphan taxa such as cryptophytes and haptophytes continue to be difficult to place with confidence in the ETOL because of their history of endosymbioses and HGTs. This suggests that significantly more genomic data are needed to elucidate these processes and more robustly comprehend how photosynthetic ability and nuclear genomes have intersected over >1 billion years of eukaryotic history. From our perspective, the ETOL is best inferred using genome wide bioinformatic approaches that do not rely heavily on human intervention. Given the inherent biases associated with the field of phylogenetics, we surmise that allowing genomes to educate us is the more plausible approach to ETOL reconstruction and understanding how algae have evolved. Finally, a study was published after the submission of this manuscript that identified nonphotosynthetic, phagotrophic *Rhodelphis* species as sister organisms to the Rhodophyta within Archaeplastida (Gawryluk et al. 2019). These findings suggest that both phototrophy and predation were key components of the evolutionary history of this lineage.

Acknowledgments

We are grateful to the New Jersey Agricultural Experiment Station and the Rutgers University School of Environmental and Biological Sciences Genome Cooperative for supporting our genomics research. We also thank our many lab colleagues and collaborators for inspiring and nurturing our research in algal evolution.

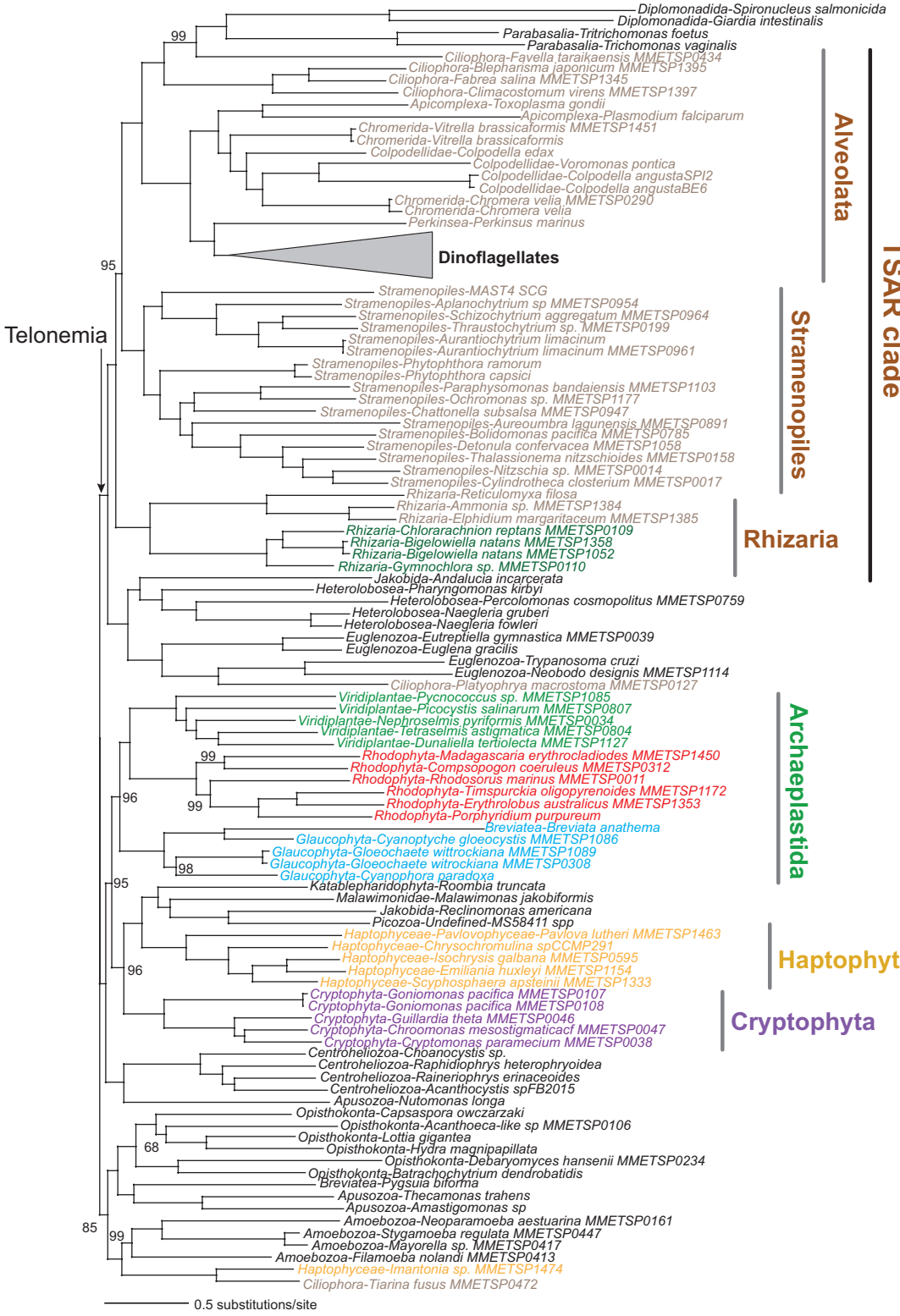

Fig. 2.5. Phylogeny built using IQ-TREE showing the positions of different algal groups in the ETOL. This tree was constructed using a partitioned 303 OG set based on the BUSCO Eukaryota dataset from 114 publicly available eukaryote proteomes. All nodes have 100% bootstrap support unless shown otherwise. Major algal groups are identified in the image. The putative position of telonemids is based on Strassert et al. (2019)

References

Adl SM, Simpson AG, Farmer MA, Andersen R, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, McCourt RM, Mendoza L, Moestrup O, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW, Taylor MF (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol 52:399–451

Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW, Burki F, Cárdenas P, Čepička I, Chistyakova L, Del Campo J, Dunthorn M, Edvardsen B, Eglit Y, Guillou L, Hampl V, Heiss AA, Hoppenrath M, James TY, Karpov S, Kim E, Kolisko M, Kudryavtsev A, Lahr DJG, Lara E, Le Gall L, Lynn DH, Mann DG, Massana I Molera R, Mitchell EAD, Morrow C, Park JS, Pawlowski JW, Powell MJ, Richter DJ, Rueckert S, Shadwick L, Shimano S, Spiegel FW, Torruella I Cortes G, Youssef N, Zlatogursky V, Zhang Q (2018) Revisions to the classification, nomenclature, and diversity of eukaryotes. J Eukaryot Microbiol. https://doi.org/10.1111/jeu.12691

Ball SG, Subtil A, Bhattacharya D, Moustafa A, Weber AP, Gehre L, Colleoni C, Arias MC, Cenci U, Dauvillée D (2013) Metabolic effectors secreted by bacterial pathogens: essential facilitators of plastid endosymbiosis? Plant Cell 25:7–21

Ball SG, Bhattacharya D, Weber AP (2016) EVOLUTION. Pathogen to powerhouse. Science 351:659–660

Baurain D, Brinkmann H, Petersen J, Rodriguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H (2010) Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. Mol Biol Evol 27:1698–1709

Bengtson S, Sallstedt T, Belivanova V, Whitehouse M (2017) Three-dimensional preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae. PLoS Biol 15:e2000735

Bhattacharya D, Helmchen T, Melkonian M (1995) Molecular evolutionary analyses of nuclear-encoded small subunit ribosomal RNA identify an independent rhizopod lineage containing the Euglyphidae and the Chlorarachniophyta. J Eukaryot Microbiol 42:65–69

Bhattacharya D, Price DC, Yoon HS, Yang EC, Poulton NJ, Andersen RA, Das SP (2012) Single cell genome analysis supports a link between phagotrophy and primary plastid endosymbiosis. Sci Rep 2:356

Blank CE (2013) Origin and early evolution of photosynthetic eukaryotes in freshwater environments: reinterpreting proterozoic paleobiology and biogeochemical processes in light of trait evolution. J Phycol 49:1040–1055

Brodie J, Ball SG, Bouget FY, Chan CX, De Clerck O, Cock JM, Gachon C, Grossman AR, Mock T, Raven JA, Saha M, Smith AG, Vardi A, Yoon HS, Bhattacharya D (2017) Biotic interactions as drivers of algal origin and evolution. New Phytol 216:670–681

Burki F, Inagaki Y, Brate J, Archibald JM, Keeling PJ, Cavalier-Smith T, Sakaguchi M, Hashimoto T, Horak A, Kumar S, Klaveness D, Jakobsen KS, Pawlowski J, Shalchian-Tabrizi K (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

Burki F, Okamoto N, Pombert JF, Keeling PJ (2012) The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. Proc Biol Sci 279:2246–2254

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 Biol Sci 283:1823

Butterfield NJ (2000) *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. Paleobiology 26:386–404

Cavalier-Smith T (1981) Eukaryote kingdoms: seven or nine? Biosystems 14:461–481

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 (2017) Kingdom Chromista and its eight phyla: a new synthesis emphasizing periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. Protoplasma 255:297–357

Cavalier-Smith T, Fiore-Donno AM, Chao E, Kudryavtsev A, Berney C, Snell EA, Lewis R (2015) Multigene phylogeny resolves deep branching of Amoebozoa. Mol Phylogenet Evol 83:293–304

Cavalier-Smith T, Chao EE, Lewis R (2016) 187-gene phylogeny of protozoan phylum Amoebozoa reveals a new class (Cutosea) of deep-branching, ultrastructurally unique, enveloped marine Lobosa and clarifies amoeba evolution. Mol Phylogenet Evol 99:275–296

Cenci U, Bhattacharya D, Weber APM, Colleoni C, Subtil A, Ball SG (2017) Biotic host-pathogen interactions as major drivers of plastid endosymbiosis. Trends Plant Sci 22:316–328

Cenci U, Qiu H, Pillonel T, Cardol P, Remacle C, Colleoni C, Kadouche D, Chabi M, Greub G, Bhattacharya D, Ball SG (2018) Host-pathogen biotic interactions shaped vitamin K metabolism in Archaeplastida. Sci Rep 8:15243

Deschamps P, Moreira D (2012) Reevaluating the green contribution to diatom genomes. Genome Biol Evol 4:683–688

Dorrell RG, Gile G, McCallum G, Méheust R, Bapteste EP, Klinger CM, Brillet-Guéguen L, Freeman KD, Richter DJ, Bowler C (2017) Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. elife 6:e23717

Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16:157

Gawryluk RMR, Tikhonenkov DV, Hehenberger E, Husnik F, Mylnikov AP, Keeling PJ (2019) Non-photosynthetic predators are sister to red algae. Nature 572(7768):240–243

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

Irwin NAT, Tikhonenkov DV, Hehenberger E, Mylnikov AP, Burki F, Keeling PJ (2018) Phylogenomics supports the monophyly of the Cercozoa. Mol Phylogenet Evol 130:416–423.

Janouškovec J, Tikhonenkov DV, Burki F, Howe AT, Rohwer FL, Mylnikov AP, Keeling PJ (2017) A new lineage of eukaryotes illuminates early mitochondrial genome reduction. Curr Biol 27:3717–3724

Karkar S, Facchinelli F, Price DC, Weber AP, Bhattacharya D (2015) Metabolic connectivity as a driver of host and endosymbiont integration. Proc Natl Acad Sci U S A 112:10208–10215

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780

Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, Beszteri B, Bidle KD, Cameron CT, Campbell L, Caron DA, Cattolico RA, Collier JL, Coyne K, Davy SK, Deschamps P, Dyhrman ST, Edvardsen B, Gates RD, Gobler CJ, Greenwood SJ, Guida SM, Jacobi JL, Jakobsen KS, James ER, Jenkins B, John U, Johnson MD, Juhl AR, Kamp A, Katz LA, Kiene R, Kudryavtsev A, Leander BS, Lin S, Lovejoy C, Lynn D, Marchetti A, McManus G, Nedelcu AM, Menden-Deuer S, Miceli C, Mock T, Montresor M, Moran MA, Murray S, Nadathur G, Nagai S, Ngam PB, Palenik B, Pawlowski J, Petroni G, Piganeau G, Posewitz MC, Rengefors K, Romano G, Rumpho ME, Rynearson T, Schilling KB, Schroeder DC, Simpson AG, Slamovits CH, Smith DR, Smith GJ, Smith SR, Sosik HM, Stief P, Theriot E, Twary SN, Umale PE, Vaulot D, Wawrik B, Wheeler GL, Wilson WH, Xu Y, Zingone A, Worden AZ (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

Kies L (1974) Electron microscopical investigations on *Paulinella chromatophora* Lauterborn, a thecamoeba containing blue-green endosymbionts (Cyanelles). Protoplasma 80:69–89

Kim KM, Park JH, Bhattacharya D, Yoon HS (2014) Applications of next-generation sequencing to unravelling the evolutionary history of algae. Int J Syst Evol Microbiol 64:333–345

Kim JI, Moore CE, Archibald JM, Bhattacharya D, Yi G, Yoon HS, Shin W (2017) Evolutionary dynamics of cryptophyte plastid genomes. Genome Biol Evol 9:1859–1872

Knoefler D, Thamsen M, Koniczek M, Niemuth NJ, Diederich AK, Jakob U (2012) Quantitative in vivo redox sensors uncover oxidative stress as an early event in life. Mol Cell 47:767–776

Lauterborn R (1895) Protozoenstudien II. *Paulinella chromatophora* nov. gen., nov. spec., ein beschalter Rhizopode des Sußwassers mit blaugrunen chromatophorenartigen Einschlussen. Z Wiss Zool 59:537–544

Lee J, Cho CH, Park SI, Choi JW, Song HS, West JA, Bhattacharya D, Yoon HS (2016) Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. BMC Biol 14:75

Linka N, Weber APM (2010) Intracellular metabolite transporters in plants. Mol Plant 3:21–53

Margulis L (1981) Symbiosis in cell evolution. WH Freeman and Company, San Francisco

Marin B, Nowack EC, Glöckner G, Melkonian M (2007) The ancestor of the *Paulinella* chromatophore obtained a carboxysomal operon by horizontal gene transfer from a *Nitrococcus*-like γ-proteobacterium. BMC Evol Biol 7:85

Méheust R, Zelzion E, Bhattacharya D, Lopez P, Bapteste E (2016) Protein networks identify novel symbiogenetic genes resulting from plastid endosymbiosis. Proc Natl Acad Sci U S A 113:3579–3584

Minh BQ, Nguyen MA, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol 30:1188–1195

Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D (2009) Genomic footprints of a cryptic plastid endosymbiosis in diatoms. Science 324:1724–1726

Nguyen LT, 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

Nowack ECM, Grossman AR (2012) Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. Proc Natl Acad Sci U S A 109:5340–5345

Nowack ECM, Melkonian M, Glöckner G (2008) Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. Curr Biol 18:410–418

Nowack EC, Price DC, Bhattacharya D, Singer A, Melkonian M, Grossman AR (2016) Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of *Paulinella chromatophora*. Proc Natl Acad Sci U S A 113:12214–12219

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

Parfrey LW, Lahr DJ, Knoll AH, Katz LA (2011) Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc Natl Acad Sci U S A 108:13624–13629

Peers G, Truong TB, Ostendorf E, Busch A, Elrad D, Grossman AR, Hippler M, Niyogi KK (2009) An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. Nature 462:518–521

Price DC, Bhattacharya D (2017) Robust Dinoflagellata phylogeny inferred from public transcriptome databases. J Phycol 53:725–729

Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, Weber AP, Schwacke R, Gross J, Blouin NA, Lane C, Reyes-Prieto A, Durnford DG, Neilson JA, Lang BF, Burger G, Steiner JM, Löffelhardt W, Meuser JE, Posewitz MC, Ball S, Arias MC, Henrissat B, Coutinho PM, Rensing SA, Symeonidi A, Doddapaneni H, Green BR, Rajah VD, Boore J, Bhattacharya D (2012) *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. Science 335:843–847

Qiu H, Price DC, Weber AP, Facchinelli F, Yoon HS, Bhattacharya D (2013) Assessing the bacterial contribution to the plastid proteome. Trends Plant Sci 18:680–687

Reyes-Prieto A, Weber AP, Bhattacharya D (2007) The origin and establishment of the plastid in algae and plants. Annu Rev Genet 41:147–168

Reyes-Prieto A, Yoon HS, Moustafa A, Yang EC, Andersen RA, Boo SM, Nakayama T, Ishida K, Bhattacharya D (2010) Differential gene retention in plastids of common recent origin. Mol Biol Evol 27:1530–1537

Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H, Lang BF (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr Biol 15:1325–1330

Sánchez-Baracaldo P, Raven JA, Pisani D, Knoll AH (2017) Early photosynthetic eukaryotes inhabited low-salinity habitats. Proc Natl Acad Sci U S A 114:E7737–E7745

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212

Singer A, Poschmann G, Mühlich C, Valadez-Cano C, Hänsch S, Hüren V, Rensing SA, Stühler K, Nowack ECM (2018) Massive protein import into the early-evolutionary-stage photosynthetic organelle of the amoeba *Paulinella chromatophora*. Curr Biol 27:2763–2773

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

Strassert JFH, Jamy M, Mylnikov AP, Tikhonenkov DV, Burki F (2019) New phylogenomic analysis of the enigmatic phylum Telonemia further resolves the eukaryote tree of life. Mol Biol Evol 36:757–765.

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D (2004) A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol 21:809–818

Yoon HS, Nakayama T, Reyes-Prieto A, Andersen RA, Boo SM, Ishida K, Bhattacharya D (2009) A single origin of the photosynthetic organelle in different *Paulinella* lineages. BMC Evol Biol 9:98

Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, Wilson WH, Yang EC, Duffy S, Bhattacharya D (2011) Single-cell genomics reveals organismal interactions in uncultivated marine protists. Science 332:714–717