学位論文

Evolutionary history of the green secondary eukaryotes chlorarachniophytes based on phylogenetic analyses of nuclear-encoded genes

（核コード遺伝子の系統解析に基づく緑色二次植物
クロララクニオン藻類の進化的系譜）

平成 25 年 12 月　博士（理学）

申請

東京大学大学院理学系研究科

生物科学専攻

楊　億

Evolutionary history of the green secondary eukaryotes chlorarachniophytes based on phylogenetic analyses of nuclear-encoded genes

Yi   Yang

Department of Biological Sciences

Graduate School of Science

University of Tokyo

December, 2013

.

Contents

Abbreviations

EGT	endosymbiotic gene transfer
PRK	phosphoribulokinase
OTU	operational taxonomic unit
CR	Chromalveolata and Rhizaria
CC	Calvin cycle
SBP	sedoheptulose-bisphosphatase
SAR	Stramenopiles, Alveolates and Rhizaria
HGT	horizontal gene transfer
AU test	approximate unbiased test

Abstract

Chlorarachniophyta is a group of photosynthetic eukaryotes harboring secondary plastids of green algal origin. Although previous phylogenetic analyses of genes encoding Calvin cycle (CC) enzymes demonstrated the presence of genes apparently not derived from green algal endosymbionts in the nuclear genomes of *Bigelowiella natans* (Chlorarachniophyta), the origins of these "non-green" genes in the "green" secondary phototrophs were unclear due to the limited taxon sampling. In the second chapter, I sequenced five new *PRK* genes (from one euglenophyte, two chlorarachniophytes, and two glaucophytes) and performed an extended phylogenetic analysis of the genes based on a wide taxon sampling. My phylogenetic analyses demonstrated that the phosphoribulokinase (*PRK*) sequences form three genera of Chlorarachniophyta were placed within the red algal clade. This "non-green" affiliation was supported by the taxon-specific insertion/deletion sequences in the *PRK* alignment. My results suggest that *PRK* genes may have been transferred from a red algal ancestor to Chlorarachniophyta before the divergence of this lineage of "green" secondary phototrophs.

The plastids of chlorarachniophytes were derived from an ancestral green alga via secondary endosymbiosis. Hence, if *PRK* was one of the traces left by a cryptic endosymbiotic gene transfer (EGT) from "red" lineage, more genes besides *PRK* from this "red" lineage via secondary EGT should be found in the nuclear genomes of the Chlorarachniophyta. In order to elucidate the origin of such "red" genes in the chlorarachniophyte nuclear genomes, in the third chapter, I carried out exhaustive single-gene phylogenetic analyses including two operational taxonomic units (OTUs) that represent two divergent sister lineages of the Chlorarachniophyta, *Amorphochlora amoeboformis* (=*Lotharella amoeboformis*; based on RNA sequences newly determined here) and *B. natans* (based on the published genome sequence). I identified eleven genes of cyanobacterial origin phylogenetic analysis of which showed the chlorarachiophytes to branch with the red algae and/or "red" secondary or tertiary phototrophs such as stramenopiles. Among the 11, nine demonstrated robust monophyly of the two chlorarachniophyte OTUs, suggesting multiple horizontal gene transfers from the red lineage before divergence of extant chlorarachniophytes. Five of the nine gene trees demonstrated weak to moderate statistical support for the affiliation of the chlorarachniophyte genes with those of the "red" secondary/tertiary phototrophs. Although chlorarachniophyte *PRK* genes apparently evolved directly from a red algal ancestor, OTUs from eukaryotes with red algal derived secondary and/or tertiary plastids (e.g., stramenopiles) did not belong to the red algal lineage, suggesting that the red secondary/tertiary phototrophs might have experienced a gene replacement after the typical secondary EGT scenario from the red algal ancestor. Thus, cryptic endosymbiosis of a secondary or tertiary plastid of the red lineage could be speculated in or before the common ancestor of the extant Chlorarachiophyta. Because most of such possible "red" secondary or tertiary EGT genes are related to photosynthesis and plastid function, the extant secondary green plastids of the Chlorarachniophyta may originate from replacement of the ancestral "red" plastid by the endosymbiotic green alga.

Combining available data on other secondary phototroph (euglenophytes, stamnopiles, cryptophytes, haptophytes) , an idea surfaced itself: secondary plastid endosymbiosis seems unlikely in the plastid lacking, heterotrophic host, and might have been supported by the pre-existing plastid-targeting system in the host nucleus.

Chapter 1

General Introduction

According to the enodosymbiotic theory, the acquisition of endosymbiotic bacteria by a protist were dated back to more than one billion years (Hedges 2004; Yoon et al. 2004; Blair et al. 2005). Evidence shows that mitochondria were developed from ancient Rickettsiales or a close relative (proteobacteria), while primary plastids from an ancient cyanobacterium (Thrash et al. 2011). The endosymbiotic theory has been considered a type of saltational evolution (Syvanen et al. 2002).

It is thought that endosymbionts transferred a large number of their genes to the nucleus of the host cell during its serial endosymbiosis, which is one form of horizontal gene transfer: EGT. Thus, the traditional way to distinguish organelles from endosymbionts is to judge the reduced genome size (Keeling et al. 2008). In the case of plastid primary endosymbiosis, the eukaryotic host acquiring photosynthetic activity diversified to the extant members of the supergroup Archaeplastida, which consists of the Chloroplastida (green algae and land plants), Rhodophyceae (red algae), and Glaucophyta. Freshwater cercozoan amoeba *Paulinella chromatophora* contains endosymbionts termed as chromatophores, which are acquired 60 MYA, and can be regarded as nascent plastid-like organelles. Chromatophores' genome is almost one third (1.02Mb) of their closest genus *Synechococus* (3Mb) and contains deficient genes accountable for photosynthesis (Nowack et al. 2008). Thirty expressed genes were transferred to the nuclear genome by EGT and some expressed proteins without signal peptides are routed to chromatophores by Golgi apparatus (Nowack et al. 2012).

In the case of plastids' evolution, despite primary endosymbiosis, secondary endosymbiosis, host cell with primary endosymbionts itself engulfed and retained by another free living eukaryote, occurred multi-times among different lineages of eukaryotes. One possible ongoing secondary endosymbiosis is the heterotrophic protist *Hatena* (Okamoto et al. 2005).

Unique morphology of plastid membranes was left after secondary endosymbiosis. Three plastid membranes can be observed among some dinoflagellates and euglenophytes and four membranes among haptophytes, stramenopiles, cryptophytes and chlorarachniophytes. The additional two layers of plastid membranes are thought to be the former plasma membrane of engulfed alga and the phagosomal membrane of the host cell.

Although the number of the red secondary endosymbiosis has been debated

(McFadden et al. 2008; Gould et al. 2004), all of the eukaryotes with red algal derived secondary and tertiary plastids (hereafter CR group) [stramenopiles (Heterokontophyta), alveolates, haptophytes and cryptophytes] were assigned to the supergroup Chromalveolata in 2005 (Cavalier-Smith 1999; Adl et al. 2005). On the other hand, green secondary plastids are found in two separate lineages, chlorarachniophytes (supergroup Rhizaria) and euglenophytes (supergroup Excavata) (Archibald 2012). Recent studies suggested that the green secondary plastids in these lineages may originate from two independent endosymbioses of green algae (Rogers et al. 2007; Takahashi et al. 2007). Multigene phylogenetic analyses of various eukaryotic lineages demonstrated that Rizaria, stramenopiles and alveolates form a monophyletic group (Archibald 2012; Burki et al. 2007; Frommolt et al. 2008; Hampl et al. 2009; Archibald 2012) that was very recently reclassified as the supergroup "SAR" (Stramenopiles, Alveolates, Rhizaria)(Adl et al. 2012). For cryptophytes in "red" secondary lineage and chlorarachniophytes in "green" secondary lineage, remnant nucleus of the algal symbiont, termed as nucleomorph, is present between the second and third layer of plastid membranes (Reyes-Prieto et al. 2007; Archibald 2009). During the secondary endosymbiotic process of engulfing, reducing and integrating primary endosymbiotic plastid ancestors, EGT also took place and a bulk of the endosymbiont genes were transferred to the host genomes to remain functional as the host nuclear genes, which works as relics of the past endosymbiotic events that we can now trace back (Martin et al 2002; Timmis et al. 2004).

Several studies on EGT events in various secondary endosymbiotic eukaryots have been carried out and EGT genes from cryptic sources, obviously not from the plastids they possess, have been disclosed. For example, Moustafa et al. (2009) found out that approximately 16% of diatom nuclear coding genes are transferred from green origins.

Archibald et al. (2003) suggested that eight *B.natans* (Chlorarachniophyta) genes encoding plastid enzymes were derived from red algae or secondary algae harboring red algal plastids and that these "red" lineage genes in *B. natans* were acquired via horizontal gene transfer (HGT) through the feeding of red algal prey organisms by mixotrophic host chlorarachniophytes. One key gene among these eight is *PRK* (EC 2.7.1.19), a Calvin cycle enzyme that catalyzes conversion of ATP and D-ribulose 5-phosphate into ADP and D-ribulose 1,5-bisphosphate. Phylogenetic analysis of *PRK* by Petersen et al. (2006) shed light on the unusual origins of the genes from two lineages of "green" secondary phototrophs,

Chlorarachniophyta and Euglenophyta. *B. natans* has a "red alga-like" *PRK* gene while *Euglena gracilis* (Euglenophyta) has a "stramenopile-like" *PRK*. However, unsolved problems remain because of only one chlorarachniophyte and one euglenophyte OTUs analyzed. In the first chapter of this thesis, I managed to sequence five novel *PRK* genes, including two chlorarachniophytes (*Gymnochlora stellata* and *Chlorarachnion reptans*). By constructing phylogenetic tree, the datum suggests that *PRK* gene was transferred at the ancestral stage.

Curtis et al. (2012) studied *B. natans* and identified 45 red algal-type genes in the nuclear genome. However, the precise origins of these red genes were not resolved, in part because their datasets also typically included only one OTU (*B. natans*) from the Chlorarachniophyta. Genes from more remote chlorarachniophytes are needed to determine whether these red genes originate from the common ancestor or the recent lineage of the chlorarachniophytes. Therefore in the second chapter, in order to expand the diversity of the chlorarachniophyte lineage used in these phylogenetic analyses I chose the chlorarachniophyte species *Amorphochlora amoebiformis* (Ishida et al. 2011) as an additional OTU because *A. amoebiformis* and *B. natans* belong to two sister, basally divergent lineages of the Chlorarachniophyta (Ota et al. 2012). I obtained transcriptome data from *A. amoebiformis* by next generation sequencing and combined them with the *B. natans* nuclear genome data. To extract more "red" genes from the Chlorarachniophyta, moreover, I established another original pipeline and manually checked as much positive outputs as possible. Based on this pipeline and the red genes extracted by Curtis et al. (2012), a total of 11 "red" genes of cyanobacterial origin were found from the chlorarachniophyte lineage, which meet a more stringent criteria mentioned in chapter 3. To explain the presence of 'red' genes in the chlorarachniophyte nuclear genomes, a possible cryptic secondary or tertiary endosymbiosis of a red alga before the secondary endosymbiosis of the current "green" secondary plastids is proposed.
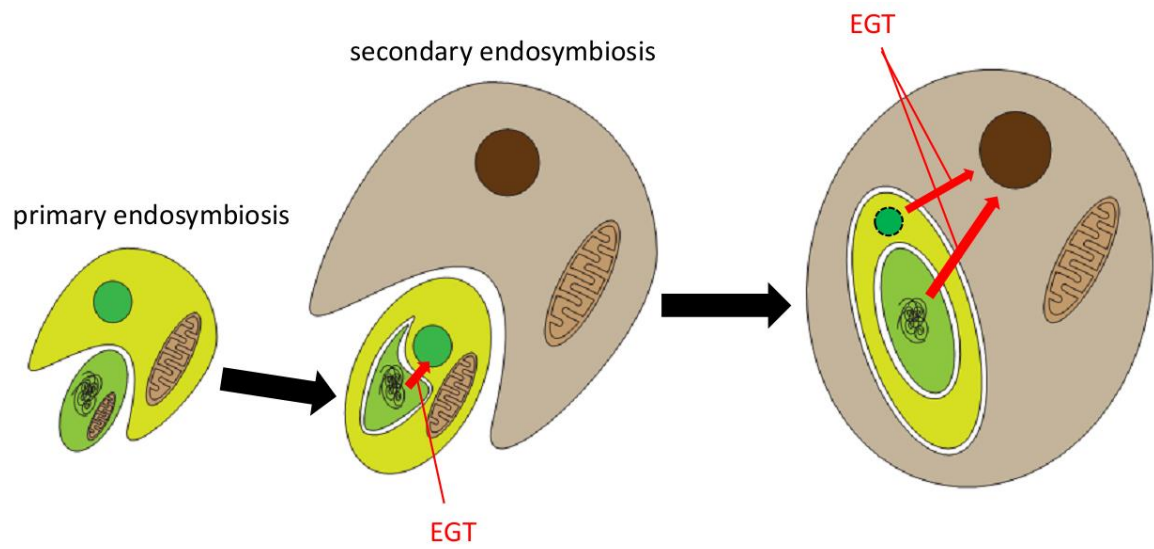
Figures

Figure 1 . **A schematic diagram of primary endosymbiosis and secondary endosymbiosis, red arrows indicate the transferring directions of endosymbiotic gene transfers (EGTs).**

Chapter 2

An extended phylogenetic analysis reveals non-green phosphoribulokinase genes from "green" secondary photosynthetic eukaryotes

Introduction

In primary phototrophs, many of the nuclear-encoded genes encoding CC enzymes are EGT-derived (Martin et al. 1997; Matsuzaki et al. 2004; Reyes-Prieto et al. 2007). PRK (EC 2.7.1.19) is one of those CC enzymes, catalyzing ATP and D-ribulose 5-phosphate into ADP and D-ribulose 1,5-bisphosphate (Petersen et al. 2004). Phylogenetic analysis suggested that PRK sequences are divided into two distantly related classes, Class I and Class II, which share approximately 20% amino acid (aa) identity (Tabita 1994; Brandes et al.1996). Proteobacterial Class I enzymes are octamers, whereas Class II enzymes from cyanobacteria and eukaryotic phototrophs function as tetramers and dimers, respectively (Harrison et al. 1998). Although some CC genes are affiliated with non-cyanobacterial prokaryotic homologs (Matsuzaki et al. 2004; Reyes-Prieto et al. 2007), the Class II *PRK* genes of the photosynthetic eukaryotes form a robust monophyletic group with cyanobacterial homologs, suggesting no gene replacement after the primary endosymbiosis (Figure 2). In addition, *PRK* genes are relatively conserved among the CC genes (Martin et al. 1997). Thus, *PRK* may be an ideal gene to trace the historical events of endosymbioses of the plastid.

Phylogenetic analysis of *PRK* by Petersen et al. (2004) shed light on the unusual origins of the genes from two lineages of "green" secondary phototrophs, Chlorarachniophyta and Euglenophyta. *B. natans* (Chlorarachniophyta) has a "red alga-like" *PRK* gene while *E. gracilis* (Euglenophyta) has a "chromalveolate-like" PRK. Obviously, unsolved problems remain because of the limited sampling of the OTUs analyzed (Matsuzaki et al. 2004; Reyes-Prieto et al. 2007). A recent study on PRK phylogeny including several additional OTUs from green algae and dinoflagellates, in addition to taxon-specific insertion/deletion in the alignment, demonstrated strong affiliation between chromalveolates and *Euglena* likewise between *Bigellowiella* and red algae (Rumphoa et al. 2009; Minge et al. 2010). However, each of the two green secondary phototrophic phyla included only a single OTU, and the taxon samplings were also limited in the Glaucophyta and Chlorophyta (one of the two major clades of green plants) (Matsuzaki et al. 2004; Reyes-Prieto et al. 2007; Rumphoa et al. 2009; Minge et al. 2010).

The present study was undertaken to deduce the origins of "non-green" PRK genes from Euglenophyta and Chlorarachniophyta and to elucidate more natural phylogenetic relationships of PRK from the major algal groups, employing a wider taxon sampling from various photosynthetic eukaryotes. I determined five new *PRK* genes from one euglenophyte, two chlorarachniophytes, and two glaucophytes and obtained several other *PRK* genes from the available genome and EST data to date. My extensive phylogenetic analyses of *PRK* genes, combined with the phylogenetic analysis of *SBP*, demonstrated ancient origins of the "non-green" genes from the two algal groups harboring "green" secondary plastids (Euglenophyta and Chlorarachniophyta), allowing us to comparably discuss the possible origins of CC genes via HGT or EGT from a cryptic endosymbiont of "non-green" algal origin.

Materials and Methods

Strains and culture

The glaucophytes *Gloeochaete wittrockiana* SAG 46.84 and *Glaucocystis nostochinearum* SAG 16.98 were cultured in AF-6 medium (Kato 1982) that was modified according to Kasai et al. (2009). *Eutreptiella gymnastica* NIES-381 and *G.stellata* CCMP 2057 were cultured in L1 medium (Guillard et al. 1993) in which the natural seawater was replaced with Daigo's artificial seawater SP (Nihon Pharmaceutical Co. Ltd., Tokyo, Japan). The cultures were grown at 20 ℃ with a 14 h:10 h light:dark (L:D) cycle. *C.reptans* NIES-624 was grown as described previously (Takahashi et al. 2007).

RNA extraction and cDNA library construction

Cells of *E. gymnastica*, *G. wittrockiana*, and *G. nostochinearum* were crushed using ceramic beads and a Mixer Mill MM 300 (Qiagen, Hilden, Germany), and RNAs were subsequently extracted using the SV total RNA isolation system (Promega, Madison, WI, USA). Cells of *G. stellata* and *C. reptans* were disrupted and homogenized using brushes (Nozaki et al. 1997), and the RNA extraction was performed using the RNeasy Midi Kit (Qiagen). Reverse transcription (RT)-polymerase chain reaction (PCR) for all five RNA samples was performed using the Capfishing full-length cDNA Kit (Seegene, Seoul, Korea). The cDNAs were used as templates for PRK gene isolation.

Cloning and sequencing of *PRK* genes

For amplification of Class II PRK genes from cDNA, I designed degenerate primers based on conserved amino acid (aa) sequences of the published PRK protein sequences (Table 1). Nested PCR amplifications using these degenerate primers were carried out using the recombinant TaqTM DNA polymerase (Takara Bio, Shiga, Japan). PCR was performed with 35 cycles at 95 ℃ for 2 min, 46 ℃ for 2 min, and 66 ℃ for 3 min, followed by 72 ℃ for 15 min using the Takara PCR Thermal Cycler (Takara Bio). First PCR products were amplified by PRK UF-1 and PRK UR-5, and the second were amplified by PRK UF-2 and PRK UR-4 (Table 1). Approximately 240 bp of PCR products were subsequently cloned into a plasmid vector (pCR®4-TOPO®) using a TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA, USA) for sequencing. Plasmids from positive clones were then sequenced using the BigDye™ Terminator v3.1 Cycle Sequencing Kit and the ABI PRISM 3100 genetic analyzer (Applied Biosystems, Foster City, CA, USA). Nucleotide sequences were determined based on at least three clones sharing the same sequence for each. Specific primers (Table1) were designed using the partial sequences of PRK genes obtained from the cloned PCR products. A 3'-rapid amplification of cDNA ends (3'-RACE) was carried out using these specific primers, and the PCR products were sequenced by the direct sequencing method.

Phylogenetic methods

Most Class II PRK sequences were retrieved from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) and Joint Genome Institute (JGI) (http://www.jgi.doe.gov/). In this analysis, besides five new PRK sequences, one brown alga (*Ectocarpus siliculosus*), seven chlorophytes, and several other available sequences were added to OTUs used previously (Petersen et al. 2006; Minge et al. 2010). Sequences of *PRK* genes from *Closterium peracerosum-strigosum-littorale* complex (*Closterium spl* complex) and *Chara braunii* were obtained from unpublished assembled expressed sequence tag (EST) data (Nishiyama pers. comm.). The aa sequences of PRK 42 eukaryotic ingroup and 14 cyanobacteria outgroup OTUs (including five genes sequenced in this study; Table 2) were aligned using SeaView (Gouy et al. 2010), and ambiguous sites were removed from the alignment to produce a data matrix of 327 aa from 56 OTUs (available from TreeBase: http://www.treebase.org/treebase/index.html). All of the *PRK* nucleotide sequences used in the present study cover the entire 327 aa except for EST database-retrieved sequences from the streptophyte *Artemisia annua* (230 aa), *Beta vulgaris* (262 aa), and the glaucophyte *Cyanophora paradoxa* (153 aa).

Bayesian inference (BI) was conducted using MrBayes (ver. 3.1.2; Huelsenbeck et al. 2001) with the WAG+I+Γ4 model. BI consisted of two parallel runs with each of four Markov chain Monte Carlo (MCMC) incrementally heated chains and 1,000,000 generations, with sampling every 100 generations. The first 25% of the generations was discarded as burn-in, and the remaining trees were used to calculate a 50% majority-rule consensus tree and determine the posterior probabilities (PP) of the individual branches. The average standard deviation of split frequencies of the two MCMC iteration runs was below 0.01 for each analysis, indicating convergence. In addition, 1000 replicates of bootstrap analyses using the maximum likelihood (ML) method were performed using both RAxML (ver. 7.0.3; Stamatakis 2006) and PhyML (ver.3.0; Guindon et al. 2003) with the WAG+I+Γ4 model. Maximum parsimony (MP) analysis was also run with PAUP 4.0b10 (Swofford 2002) with the nearest-neighbor-interchange search method to produce bootstrap values based on 1000 replicates.

In addition, I carried out two approximate unbiased tests (AU test) (Shimodaira 2002) to examine the phylogenetic positions of the two monophyletic groups of euglenophytes and chlorarachinophytes. I used two series of the phylogenetic trees of PRK sequences, where topologies of all the OTUs excluding either of the euglenophytes or chlorarachniophytes were fixed, and the alignment (327 aa) as input data. All possible topologies were generated by re-grafting the branch of euglenophytes or chlorarachniophytes using the in-house ruby script. The pools of topologies were analyzed with the AU test using the site-wise log-likelihood values were calculated using PhyML (with WAG model+F+I+Γ4) and used for AU

test conducted by Consel (ver. 0.1k; Shimodaira et al. 2010).

Analyses of *SBP* genes were also carried out based on 275 aa from 37 OTUs representing a wide range of eukaryotic organisms (including two chlorarachniophyte sequences) (Table 3) using the same phylogenetic methods as for the present *PRK* genes described above.

Programs for BI, ML and AU test were executed on a supercomputer (Human Genome Center, University of Tokyo, Japan).

Results

PRK phylogeny

   As shown in Figure 3, PRK sequences from each of the two eukaryote phyla, Euglenophyta and Chlorarachniophyta, with the green secondary plastids was resolved as a monophyletic group with very high support values (1.00 PP in BI and 100% bootstrap values [BV] by the three other methods). Two euglenophyte PRK sequences and those from stramenopiles, haptophytes, and cryptophytes formed a large clade supported by relatively high support values (1.00 PP and 93–96% BV), whereby the Euglenophyta represented a derived position. Three chlorarachniophyte and red algal homologs were resolved as a monophyletic group supported by 1.00 PP and 50-70% only in ML analyses. The monophyly of green plant (land plants and green algae) homologs was moderately supported (with 1.00 PP and 62–86% BV), and the green plant and chromalveolate sequences (including euglenophyte OTUs) formed a large monophyletic group with high support values (1.00 PP and 98–100% BV). Three OTUs of glaucophytes were resolved as a monophyletic group with 1.00 PP and 61-82% BV, and constituted a basal eukaryotic group with red algal and chlorarachniophyte homologs. However, phylogenetic relationships within this basal group were not well resolved.

Based on the tree topology (Figure 3) and the patterns of insertion/deletion sequences (Figure 4), PRK proteins were subdivided into five groups, cyanobacteria, stramenopiles plus euglenophytes, cryptophytes plus haptophytes, green plants, and the basal eukaryotic group (glauocophytes, red algae, and chlorarachniophytes). The results of AU tests (Figure 5) showed that the tree topologies, in which the euglenophyte PRK clade was nested within or sister to the chlorophyte clade and at the basal positions of green plants, were rejected at the 5% level (Figure 5A). Although the AU test did not reject the positioning of the chlorarachinophyte sequences at most of the basal branches of the tree and at distal branches of green plant homologs, the topologies where the chlorarachniophyte OTUs are positioned within the chromalveolate clade were rejected at the 5% level (Figure 5B).

SBP phylogeny

   The phylogenetic results of SBP proteins are shown in Figure 6. As in the previous study (Teich et al. 2007), the SBP proteins were subdivided into two groups, one of which is composed of possible plastid-targeted proteins from green plants, Euglena, four red algal sequences, Cyanophora, and two chlorarachniophytes (*Bigelowiella* and *Gymnochlora*). The two chlorarachiophytes and two red algae (*Porphyra* and *Chondrus*) formed a moderate monophyletic group, with 0.94 PP in BI and 61–69% BV in three other phylogenetic methods. Euglena was positioned within one of the chlorophyte lineages that constituted a robust monophyletic group (with 1.00 PP in BI and 65–81% BV in three other

phylogenetic methods).

Eukaryotic *SBP* is a nuclear-encoded gene of bacterial ancestry (Martin et al. 1996). Teich et al. argued that *SBP* genes found in phototrophic eukaryotes were likely to have originated from a single recruitment of plastid-targeted enzyme in photosynthetic eukaryotes after primary endosymbiosis and a further distribution to algae with secondary plastids via EGT (Teich et al. 2007). My results are consistent with this scenario and I postulate that euglenophyte and chlorarachinophyte *SBP* genes were transferred from a green alga and a red alga, respectively.

Discussion

Origin of the "non-green" genes from Chlorarachniophyta and its implication with secondary "green" plastids

Earlier phylogenetic analyses of the plastid-encoded genes and the nuclear-encoded plastid-targeted PsbO proteins demonstrated that the "green" secondary plastids of Chlorarachniophyta and Euglenophyta were of distinct origins, but the sequences from these two groups and green plants formed a robust monophyletic group as a whole (Rogers et al. 2007; Takahashi et al. 2007). On the other hand, EST data of *B.natans* showed the composition of nucleus genome is a mixture of genes derived from various sources (Archibald et al. 2003). Consistent with previous studies (Petersen et al. 2006; Minge et al. 2010), my phylogenetic analysis of PRK proteins (Figure 3) and comparison of insertion/deletion sequences in the alignment (Figure 4) suggested that the three chlorarachniophyte *PRK* genes likely originated from red algae rather than the "green" lineage. Several lines of research on the phylogeny of the Chlorarachniophyta (Ishida et al. 1999; Archibald et al. 2004; Chantangsi et al. 2010; Parfrey et al. 2010) indicate that the three chlorarachniophyte genera examined in this study were distributed widely within this phylum, which provide the basis for postulating the phylum-wide sampling of *PRK* genes. Therefore, it is likely that the red alga-like *PRK* genes were transferred from an ancestral red alga before divergence of the Chlorarachniophyta.

Recent nuclear multigene phylogenetic studies of eukaryotes suggested that Rhizaria (including chlorarachniophytes) was a sister group to the clade composed of stramenopiles and alveolates (Hackett et al. 2007; Burki et al. 2009; Parfrey et al. 2010). However, the PRK protein phylogeny in this study showed that the clade composed of red algae and chlorarachniophytes is robustly separated from the chromalveolates (Figure 3). In addition, the AU test rejected the tree topology in which the chlorarachiophyte PRK genes were associated with stramenopiles (Figure 5B), and the unique insertion/deletion sequences supported the affiliation between chlorarachniophytes and red algae (Figure 4). Taking all these results together, it is unlikely that the chlorarachiophyte PRK genes were derived from the host component of Rhizaria.

My phylogenetic analysis (Figure 6) showed that at least two chlorarachniophyte sequences were nested within the red algae-derived *SBP* clade, which suggested a single HGT from an ancestral red alga to the ancestor of chlorarachniophytes. A previous study proposed a hypothesis that the plastid-targeted SBP proteins of non-cyanobacterial origin was introduced and replaced the original cyanobacterial counterpart in the common ancestor of primary phototrophs, i.e., green plants, glaucophytes and red algae (Reyes-Prieto et al. 2007). As is the case with PRK, it

is likely that the "red" *SBP* enzymes functionally replaced the original "green" *SBP* in the chlorarachniophyte plastids, providing evidence that the proteome of secondary plastids is an evolutionary mosaic as seen in primary plastids (Reyes-Prieto et al. 2007).

Archibald et al. (2003) suggested that eight *B. natans* genes encoding plastid enzymes were derived from red algae or secondary algae harboring red algal plastid, and that these "red" lineage genes in *B. natans* were acquired via HGT through the feeding of red algal prey organisms by mixotrophic host chlorarachiophytes. Thus, one possible explanation for the origin of the multiple red algal-derived genes in the chlorarachniophyte nuclear genomes is that these genes were transferred from red algal prey organisms via HGT. Furthermore, the fact that the essential functions of CC enzymes in the plastid metabolism lead to an idea that the red algal prey might have had a close interaction with the host and provided CC enzymes which accordingly enhanced photosynthetic performance. Alternatively, the prey organisms might have been captured by and retained in an ancestral (and probably non-photosynthetic) chlorarachniophyte as an endosymbiont, which was then replaced by green algal endosymbiont giving rise to the extant secondary plastid in Chlorarachniophyta (Figure. 7).

As consistent with a previous study (Minge et al. 2010), my phylogenetic analysis and comparison of insertion/deletion sequences in the alignment (Figure 4 and Figure 8) suggested that the dinoflagellate *PRK* genes were red algal-derived, while the basal branches of the clade including dinoflagellates and red algae was not well resolved, possibly due to the long-branched dinoflagellate genes. An apparent long-branch artifact in the PhyML analysis resulted in a different phylogenetic pattern, where the dinoflagellate PRK sequences were associated with the chromalveolate counterparts with a moderate support value (data not shown) as reported by Rumpho et al. (2009). If dinoflagellates and chlorarachniophytes possess the PRK genes of different origins, it is likely that the red algal *PRK* genes in an ancestral chlorarachniophyte were acquired via HGT or EGT, independently of secondary endosymbiosis which gave rise to the red plastids in an common ancestor of Chromalveolata, followed by loss of the "red" *PRK* genes in stramenopiles (Figure 9A). Meanwhile, if the dinoflagellate genes share the same red algal origin with the chlorarachniophyte counterparts, the source of the "red" *PRK* genes might be attributed to a hypothesized secondary endosymbiosis in the common ancestor of SAR (Figure 9B) (Keeling et al. 2010). Assuming that the latter interpretation is correct, multiple losses of "red" secondary plastid should be considered in most non-photosynthetic Rhizaria and basally branching stramenopiles (Bodył et al. 2009). Considering the multiple losses scenario (Figure 9B) requires a number of 'evolutionary steps' (Bodył et al. 2009), more parsimonious seems to be the former interpretation: EGT or HGT of CC enzyme genes from a red alga in the common ancestor of Chlorarachniophyta, independently of secondary endosymbiosis giving rise to alveolate plastids (Figure

9A). Further investigation of PRK genes in Rhizaria and improved phylogeny of Rhizaria (Chantangsi et al. 2010) will clarify the complex history of the CC enzyme gene families.

"Non-green" origins of the PRK genes from Euglenophyta

The PRK sequences from the secondary phototrophic group Euglenophyta also showed the "non-green" affiliation in this analyses (Figure 3), despite the well-established notion that the euglenophyte plastids originated from green plants (Rogers et al. 2007; Takahashi et al. 2007).   My results of AU test also did not support that the euglenophyte *PRK* genes originated from the prasinophyte-like secondary endosymbiont which gave rise to the secondary plastid in Euglenophyta or a basal lineage of the green plants (Figure 5A). Earlier phylogenetic research on nuclear genes (Mullner et al. 2010; Leander 2004; Linton et al. 2010) suggested that Euglena and Eutreptiella are representative genera of two major monophyletic groups in Euglenophyta. My phylogenetic tree and comparison of insertion/deletion characters in the alignment demonstrated that *PRK* genes of these two genera are both chromalveolate-like (Figures 3, 4), suggesting that the HGT of chromalveolate *PRK* genes might have taken place before the divergence of the extant members of Euglenophyta.

A recent study of putative chromalveolate-derived genes in *E.gracilis* and *P. trichophorum* (phagotrophic euglenid), based on the single gene-based phylogenetic analysis using EST data, proposed a testable hypothesis on an ancient EGT from a chromalveolate ancestor to the common ancestor of Euglenida (including both phototrophic and heterotrophic euglenoids) (Maruyama et al. 2011). The present study demonstrated the presence of "non-green" *PRK* genes in Chlorarachniophyta and Euglenophyta, possibly originating from nongreen phototrophs via HGT (or EGT). One possible explanation for the presence of "non-green" genes in euglenids is that the ancestor might have been a photosynthetic eukaryote harboring chromalveolate-derived plastids, given that *PRK* genes are present only in extant photosynthetic organisms (Figure 3) and loss of photosynthesis seems to be associated with the loss of *PRK* genes soon after (Maruyama et al. 2011).

Origin of "green" *PRK* genes in chromalveolates

My tree's topology robustly resolved that green plants constitute a monophyletic group adjacent to chromalveolates as a sister group (Figure 3). Given the phylogenetic analyses of eukaryotes using slowly evolving nuclear genes suggesting that chromalveolates are sister to green plants (Nozaki et al. 2007, 2009), the sister relationship of PRK genes between green plants and chromalveolates may have resulted from their host cell phylogeny. This implies

that chromalveolates might once have been photosynthetic algae harboring primary plastids which shared the same origin with green plants' counterparts (Nozaki et al. 2007, 2009; Matsuzaki et al. 2008). Under this view, after the divergence between chromalveolates and green plants, the *PRK* genes within some chromalveolate lineages might have been retained in the host nuclei even after the original "green" plastids were replaced by the extant "red" plastids via secondary endosymbiosis of red alga (Matsuzaki et al. 2008).

Alternatively, Moustafa et al. (2009) argued that a chromalveolate ancestor might once have harbored a green algal endosymbiont. Besides the PRK gene, an expanded list of green-related genes has been reported in chromalveolates (Frommolt et al. 2008; Huang et al. 2008), which is consistent with the hypothesis on an EGT event from a green alga, possibly a prasinophyte ancestor, in the ancestor of chromalveolates (Moustafa et al. 2009). Richer taxon sampling will provide insights into the complex history of gene transfer events between chromalveolates and green plants as well as other eukaryotic lineages.

Tables

Table 1. Degenerate primers designed for Class II PRK genes.

Name	Direction/positions [a]	Sequence of nucleotide acids (5'→3')
PRK UF-1	forward/120-138	GGI BTI CGI CGI GAY WSI GG
PRK UF-2	forward/144-165	GGI AAR WSI ACI TTY HTI MG
PRK UR-4	reverse/411-393	DB RTG RTT RTA DAT IGG YTT
PRK UR-5	reverse/483-468	WD IGG RTG IAR ICC YTC

[a] Corresponding to nucleotide sequence of coding region of *PRK* from *Chlamynomonas reinhardtii* (M36123)

Table.2. List of PRK genes analyzed in this study

Taxon		Scientific name/strain designation	NCBI accession no. or origin of data
Excavata	Euglenozoa	*Euglena gracilis*	AAX13964
		Eutreptiella gymnastica / NIES-381	AB643659[3]
Chromalveolata	Heterokontophyta	*Vaucheria litorea*	AAK21910
		Ectocarpus siliculosus	CBN76689
		Odontella sinensis	CAA69902
		Phaeodactylum tricornutum	ACI65926
		Thalassiosira pseudonana	EED92818
	Cryptophyta	*Guillardia theta*	AAX13960
	Haptophyta	*Prymnesium parvum*	AAX13966
		Emiliania huxleyi	GE162398;GE182408; GE166409
		Pavlova lutheri	AAX13959
	Dinophyta	*Alexandrium catenella*	EX456942;EX456054
		Lingulodinium polyedrum	AAX13961
		Pyrocystis lunula	AAX13962
		Amphidinium carterae	CF065476;CF065984
Archaeplastida	Viridiplantae	*Micromonas sp.*	ACO69442
		Micromonas pusilla	EEH58022
		Ostreococcus lucimarinus	ABO95553
		Ostreococcus tauri	CAL51688
		Chlamydomonas reinhardtii	AAA33090
		Volvox carteri	EFJ49815
		Chlorella variabilis	EFN55598
		Chlorella sp.	JGI[1]
		Arabidopsis thaliana	AAG50797
		Arteminis annua	JGI[1]
		Oryza sativa 1	BAF09739
		Oryza sativa 2	EAY95389
		Zea mays	ACG42204
		Triticum aestivum	CAB56544
		Populus trichocarpa	EEE78434
		Pisum sativum	CAA72118
		Spinacia oleracea	CAA30499
		Beta vulgaris	AAD55057;BQ582838; DV501631
		Selaginella moellendorffii	EFJ11946
		Physcomitrella patens subsp. Patens	EDQ79298
	Glaucophyta	*Cyanophora paradoxa*	EC664008;EC654359
		Glaucocystis nostochinearum / SAG 16.98	AB643660[3]
		Gloeochaete wittrockiana /SAG46.84	AB643661[3]

	Rhodophyta	*Galdieria sulphuraria*	CAC80070
		Chondrus crispus	AAX13965
		Cyanidioschyzon merolae	CMF117C[2]
Rhizaria	Chlorarachniophyta	*Bigelowiella natans*	AAP79209
		Gymnochlora stellata / CCMP2057	AB643662[3]
		Chlorarachnion reptans / NIES-624	AB643663[3]
Cyanobacteria		*Cyanothece sp.*	ACB50346
		Microcystis aeruginosa	BAG03667
		Synechococcus sp.	ABD03803
		Microcoleus chthonoplastes	EDX77839
		Synechocystis sp.	AAA27293
		Thermosynechococcus vulcanus	BAC11759
		Acaryochloris marina	ABW29884
		Nostoc sp.	BAB75822
		Anabaena variabilis	ABA20405
		Nostoc punctiforme	ACC81289
		Nodularia spumigena	EAW45547
		Synechococcus elongatus	BAD78757
		Trichodesmium erythraeum	ABG52717
		Lyngbya sp.	EAW3376

[1] Sequence obtained from http://genome.jgi-psf.org/

[2] Sequence obtained from Cyanidioschyzon merolae Genome Project

[3] Sequenced in this study

Table.3.   List of   SBP genes analyzed in this study

Taxon		Scientific name/strain designation	NCBI accession no. or origin of data
Excavata	Euglenozoa	*Euglena gracilis*	DQ508154
		Trypanosoma brucei	CAC70746
		Trypanosoma cruzi	XP_812624
Chromalveolata	Alveolata	*Neospora caninum*	CF939128; CF967456
		Toxoplasma gondii	CV653900; AA012450
		Tetrahymena thermophila	DY677322
		Paramecium tetraurelia	CT743108
	Heterokontophyta	*Phaeodactylum tricornutum*	CU732165
		Thalassiosira pseudonana	JGI[1]
	Cryptophyta	*Guillardia theta*	DQ508151
	Haptophyta	*Prymnesium parvum*	DQ508152
		Emiliania huxleyi	CX778868
	Dinophyta	*Lingulodinium polyedrum*	DQ508153
Archaeplastida	Viridiplantae	*Micromonas sp.*	ACO64947
		Ostreococcus tauri	CAL53197
			BP086378; AV641895;
		Chlamydomonas reinhardtii	BP095751; BG858946
		Volvox carteri	FD837335;FD813925
		Chlamydomonas sp.	BAA94305
		Arabidopsis thaliana	AAK96860
		Oryza sativa	AAO22558
		Zea mays	AY105080
		Triticum aestivum	P46285
		Marchantia polymorpha	DQ508155
		Pinus taeda	CO171047
		Spinacia oleracea	O20252
		Physcomitrella patens subsp. Patens	FC442624
	Glaucophyta	*Cyanophora paradoxa*	TBestDB: CPL00000215[4]
	Rhodophyta	*Galdieria sulphuraria*	GDB[3]
		Chondrus crispus	DQ508156
		*Cyanidioschyzon merolae*1	CMI196C[2]
		*Cyanidioschyzon merolae*2	CMT362C[2]
		Porphyra yezoensis	DQ508157
Rhizaria	Chlorarachniophyta	*Bigelowiella natans*	AAP79184
		Gymnochlora stellata	ACF24550
Fungi		*Gibberella zeae*	XP_383303
		Magnaporthe grisea	XP_367798
		Neurospora crassa	EAA27816

[1] Sequence obtained from http://genome.jgi-psf.org/
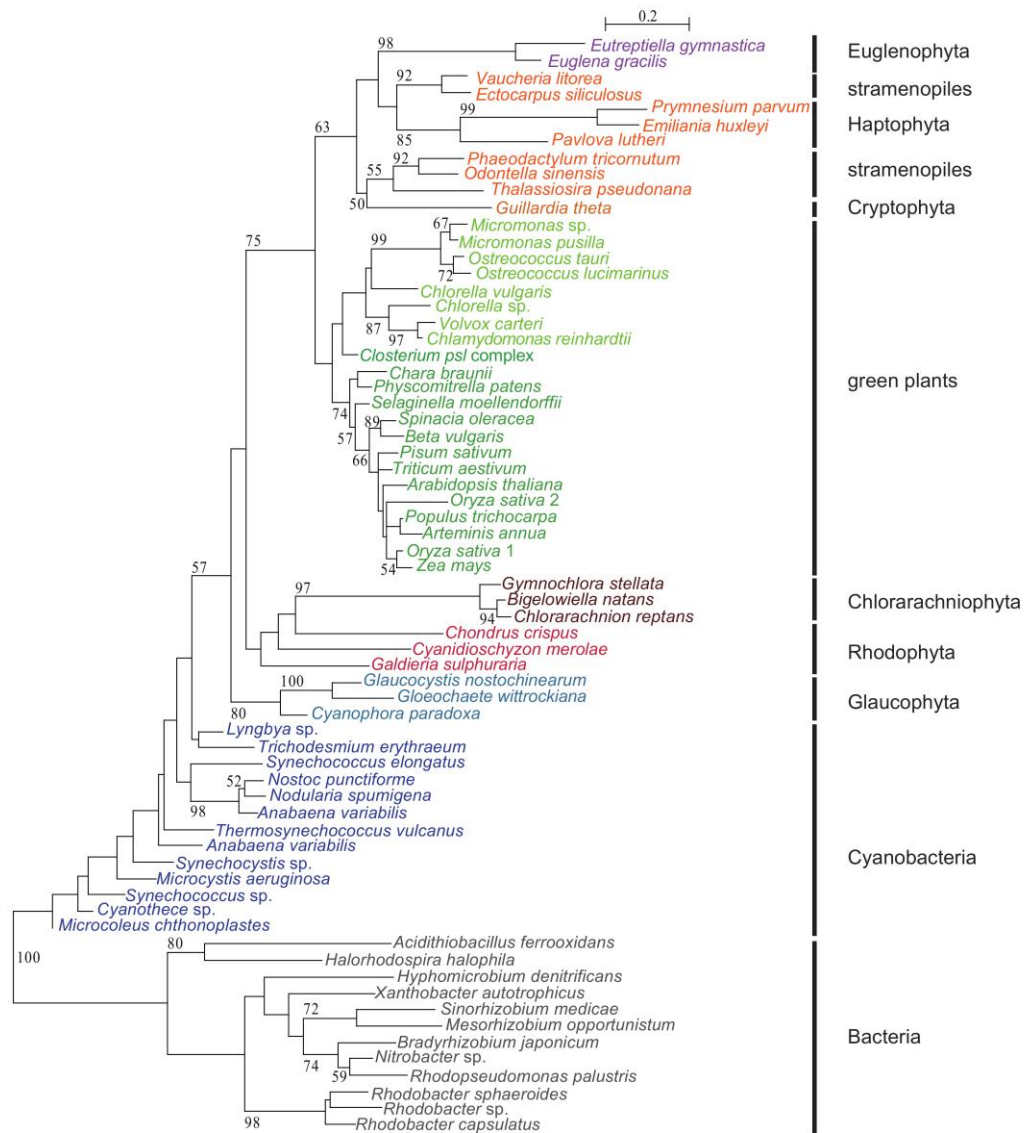
Figures

Figure 2. Phylogeny of 12 OTUs of PRK (Class I) and 56 OTUs of PRK (Class II) using RaxML. The tree was inferred using the Bayesian method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values) with RaxML.
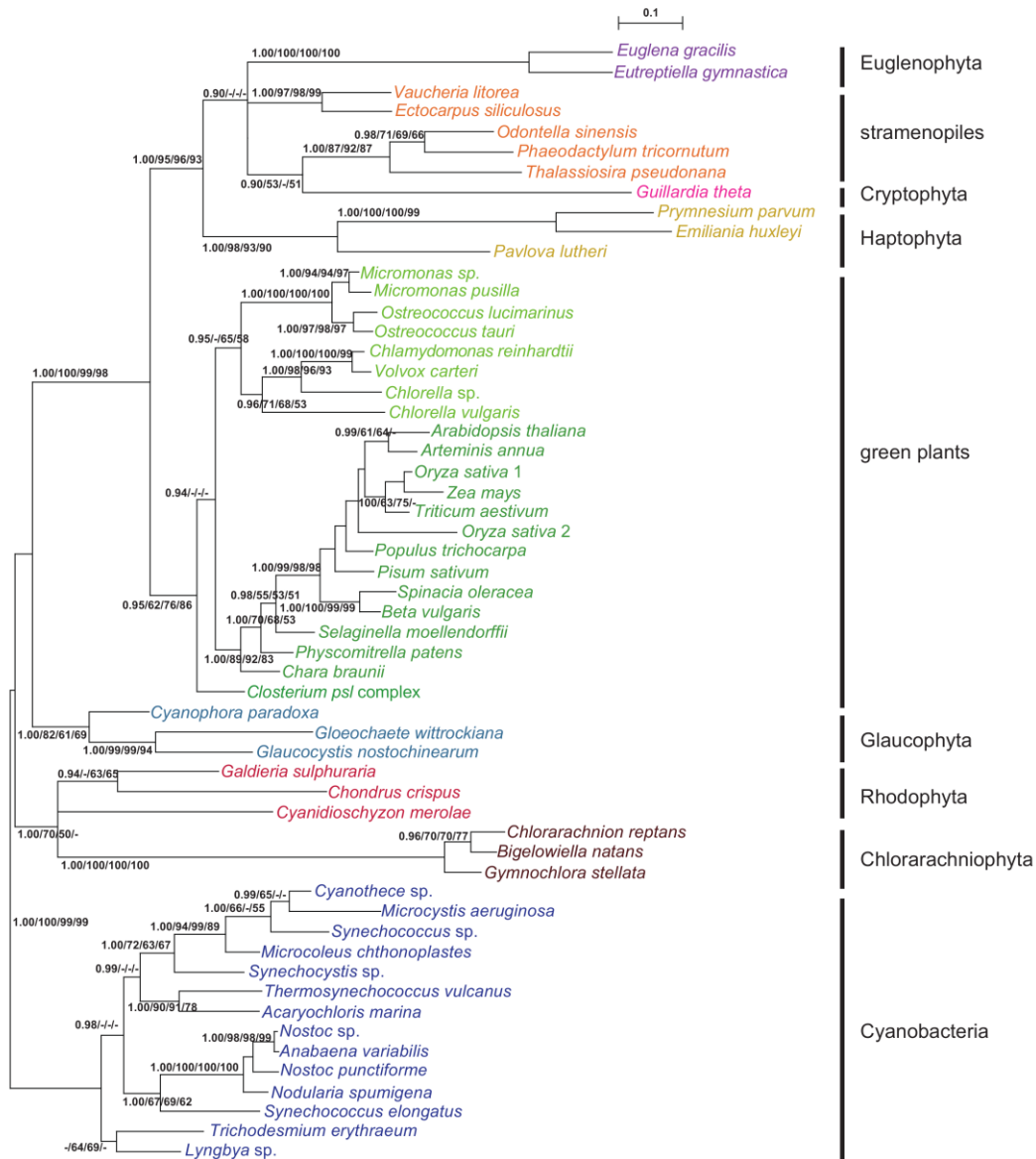
Figure 3. Phylogeny of phosphoribulokinase proteins from 56 operational taxonomic units of photosynthetic organisms. The tree was inferred using the Bayesian method with the WAG+I+gamma model. Numbers at branches represent support values (≥0.9 posterior probability or ≥50% bootstrap values) from Bayes/RaxML/PhyML/MP.
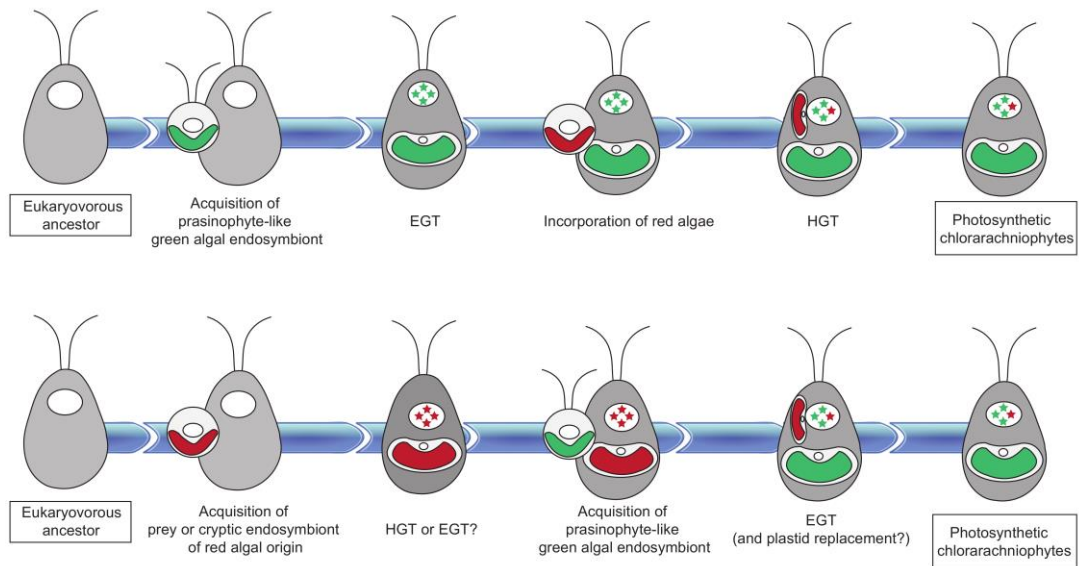
Figure 4. Taxon-specific gaps in alignment of phosphoribulokinase amino acid sequences from 56 operational taxonomic units (Figure 1) plus 4 dinoflagellates. Numbers at the top represent amino acid positions of the *Chlamydomonas reinhardtii* PRK protein (AAA33090).

Figure 5A. Results of the AU tests for assessing the phylogenetic position of euglenophyte clades into a 54-taxon BI tree. The posterior probabilities (≥0.9) of the individual branches are shown. The probability for each topology is indicated with the branch thickness.

Figure 5B. Results of the AU tests for assessing the phylogenetic position of chlorarachniophyte clades into a 53-taxon BI tree. The posterior probabilities (≥0.9) of the individual branches are shown. The probability for each topology is indicated with the branch thickness.

Figure 6. Phylogeny of sedoheptulose-1,7-bisphosphatase proteins from 35 operational taxonomic units of eukaryotes. The tree was inferred using the Bayesian method with the WAG+I+gamma model. Numbers at branches represent support values (≥0.9 posterior probability or ≥50% bootstrap values) using Bayes/RaxML/PhyML/MP.
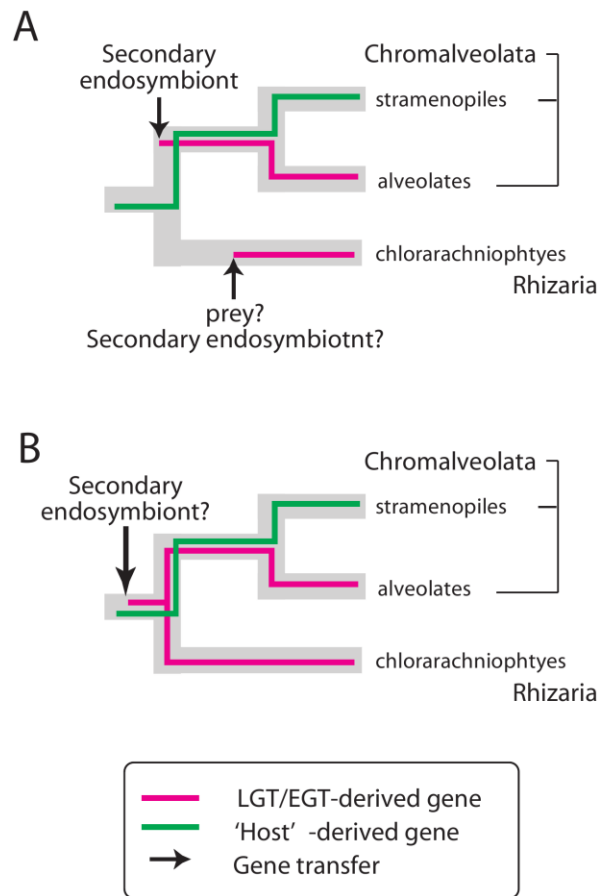
.

Figure 7. Two possible scenarios for the complex history of secondary endosymbiosis in the common ancestor of the Chlorarachniophyta.

.

Figure 8. Phylogeny of PRK proteins from 60 operational taxonomic units including two OTUs from dinophytes. The tree was inferred using the Bayesian method with the WAG+I+gamma model. Numbers at branches represent support values (≥0.9 posterior probability or ≥50% bootstrap values) using Bayes/RaxML/PhyML/MP.

.

Figure 9. Hypotheses on evolutionary history of "non-green" CC enzymes in Chlorarachniophyta.

.

Chapter 3

Phylogenomic analysis of "red " genes in the "green" secondary phototrophs chlorarachniophytes suggests a possible cryptic endosymbiosis of a "red" plastid

Introduction

Archibald et al. (2003) showed that eight genes of the chlorarachniophyte *B.natans* were derived from red algae or red secondary plastid-bearing algae. More recently, Curtis et al. (2012) identified 45 red algal-type genes in the nuclear genome sequence of *B. natans*. However, the precise origins of these red genes were not resolved, in part because their data sets typically included only one OTU (*B. natans*) from the Chlorarachniophyta. Genes from more remote chlorarachniophytes are needed to determine whether these red genes originate from the common ancestor or the recent lineage of the chlorarachniophytes

To expand the diversity of the chlorarachniophyte lineage used in these phylogenetic analyses I chose the chlorarachniophyte species *Amorphochlora amoebiformis* as an additional OTU because *A. amoebiformis* and *B. natans* belong to two sister, basally divergent lineages of the Chlorarachniophyta (Ota et al. 2012). Here, I obtained transcriptome data from *A. amoebiformis* by next generation sequencing and combined them with the *B. natans* nuclear genome data. To extract more "red" genes from the Chlorarachniophyta, I established another original pipeline and manually checked as much positive outputs as possible. Based on this pipeline and rechecking the red genes extracted by Curtis et al. (2012), a total of 11 "red" genes of cyanobacterial origin were found from the chlorarachniophyte lineage. To explain the presence of 'red' genes in the chlorarachniophyte nuclear genomes, a possible cryptic secondary or tertiary endosymbiosis of a red alga before the secondary endosymbiosis of the current "green" plastids is proposed.

Materials and Methods

Strain and culture

  A.amoebiformis CCMP2058 (designated as *L.amoeboformis*) was sent from NCMA (Provasoli-Guillard National Center for Marine Algae and Microbiota; https://ncma.bigelow.org/) and cultured in L1 medium (Hallegraeff et al. 2003) in which the natural seawater was replaced with Daigo's artificial seawater SP (Nihon Pharmaceutical Co. Ltd., Tokyo). The cultures were grown at 20 ℃ with a 14 h:10 h light:dark (L:D) cycle. Four litre medium cultivated in two flasks of two litres for a period of 57 days was brought to RNA extraction.

RNA extraction

  Cells of A. amoebiformis were broken open manually using quartzone sand in liquid nitrogen for 10 min, and RNAs were subsequently extracted using the SV total RNA isolation system (Promega). Quantity of total RNA was measured by NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and Qubit® 2.0 Fluorometer (Life Technologies), until the quantity reached 300 μg with a concentration of 6 μg/μl. The extracted total RNA was then sent to TAKARA BIO for further processing steps, including poly(A) purification and GS FLX+ analysis
(http://catalog.takara-bio.co.jp/jutaku/basic_info.asp?unitid=U100005162).

Transcriptome data assembly

The GS FLX+ output fasta data provided by TAKARA BIO, containing all 197,073 single reads, were assembled using Trinity (http://trinityrnaseq.sourceforge.net/) (Grabherr et al. 2011) on a 2x8Core Xeon E5-2650 2.00GHz SandyBridge-EP platform. The resulting 11,669 mRNA-derived contigs were translated in both directions to form 23,338 amino acid sequences (with the longest coding sequences among the three frames in each direction), which were subsequently formatted for analysis using local BLASTP.

Phylogenetic methods

The predicted 21,708 amino acid sequences available from the *B. natans* nuclear genome data
[http://genome.jgi-psf.org/pages/dynamicOrganismDownload.jsf?organism=rhizaria] were used as queries for BLASTP. The BLASTP was carried out against the NCBI and local databases that were retrieved from NCBI and Joint Genome Institute (JGI; http://www.jgi.doe.gov/), several unpublished data and my A.amoebiformis sequences prepared as described above (Table 4). Multiple sequence alignments were generated using Muscle (v3.7 by Robert C. Edgar, http://www.drive5.com/muscle)(Edgar 2004a and 2004b). I limited local databases used in the first round (group A in Table 5) as some expressed sequence

tags (EST) databases result in stretching length of gaps in alignments when quality of their sequences was bad. Meanwhile I used a trimming script to mainly exclude sequences with more than 15% gaps in each alignment. When the fasta data output contained less than four sequences, the trimming process was redone using an alternative trimming option which preserve those with less than 70% gaps, considering the "gap stretch effect" of a rough local BLAST database. Redundant OTUs with the same specific name were also excluded automatically. 'First-round' RaxML (7.2.7) (Stamatakis 2006) phylogenetic analyses were carried out with WAG+Γ4 model as a fast filter (which ignored bootstrap values) in order to remove trees containing less than three cyanobacterial OTUs and those with eukaryote genes that did appear to show plastid EGT (with basally positioned cyanobacterial OTUs). Results that passed the first-round filter were checked manually for tree topology supporting the cyanobacterial origin of eukaryote genes. All possible plastid EGT queries were searched in BLASTP once more with an extensive local database of BLAST (group B in Table S2). According to alignments and tree topologies, long branched OTUs were excluded manually. As the final outputs of second-round phylogenetic analyses, all RaxML analyses were redone with 1000 replications of bootstrap analysis. Analyses based on PhyloBayes 3.3 (http://www.atgc-montpellier.fr/phylobayes/) (Lartillot et al. 2004, 2006, 2007) were also carried out with WAG+ Γ4 model, and the"good run"option were used.

Almost all the alveolate OTUs were automatically removed during my gene mining process via BLAST due to their divergent or long-branched sequences. However, alveolates belong to SAR with chlorarachniophyes and stramenopiles (Burki et al. 2007; Frommolt et al. 2008; Hampl et al. 2009; Adl et al. 2012; Archibald 2012). Thus, analyses were also carried out using the single-gene data matrix with additional alveolate OTUs for comparison ("B" series of figures if present).

Recently, Deschamps and Moreira (2012) pointed out problems for automated massive phylogenomic analyses based on the restriction of available genomic data that are unevenly distributed among the tree of eukaryotes. Therefore, results that passed the first-round filter were re-analyzed with the local databases originally constructed (see above), and Curtis et al. (2012)'s supplementary data have been compared and added to my results to evaluate and cross-check my results.

Results and Discussion

Gene mining

To elucidate the contribution of genes of the red lineage to genome mosaicism in Chlorarachniophyta, I searched the proteome of *B. natans* for proteins showing red algal affiliations. First, I used the 21,708 predicted proteins in *B. natans* available from NCBI as queries to conduct a thorough phylogenomic search. 3,436 proteins out of the 21,708 queries had more than 10 OTUs with which to construct phylogentic trees and their affinities were examined in an automated fashion using a ruby script. Less than half (1,551) of the 3,436 proteins were categorized as showing a chlorarachniophyte phylogenetic affiliation with stramnopiles/alveolates, red algae or Chloroplastida. Nevertheless, only a small portion of them (259 candidates) had more than two cyanobacterium OTUs in their phylogenetic trees, which is a minimum requirement for a cyanobacterial origin in eukaryote OTUs. Approximately half (129) of the 259 eukaryotic genes of possible cyanobacterial origin show affinities of the chlorarachniophyte OTUs with Chloroplastida. The rest might include genes that show a "red lineage" affinity for the chlorarachniophyte homolog. After further refinement of my sampling sequence pool and that from the supplementary data of Curtis et al. (2012), I selected 6 and 8 cyanobacteria-type hits, respectively, resulting in a total of 11 genes that likely represent almost all eukaryote OTUs originating from the plastid primary EGT and chlorarachniophytes positioned in the red lineage (within or sister to red algae and/or CR group) (Figure 10). These 11 cyanobacterial gene trees showed ≥69% bootstrap values (BV) and/or ≥0.97 posterior probability (PP) for the affiliation of the chlorarachniophytes with the red lineage (Table 5).

Among the 11 genes, genes encoding ATP binding cassette transporter (ABC), mRNA binding protein (RNABP), and geranylgeranyl reductase (GGR), were identified by both pipelines (the present study and Curtis et al. 2012). Three other queries, one putative membrane protein, one unidentified putative membrane protein, and PRK, were found only from my mining pipeline (Figure 10).

Single-gene phylogenetic analyses

Nine of the 11 trees show robust monophyly of *B. natans* and *A. amoeboformis* (representing the two divergent sister lineages of the Chlorarachniophyta) (≥95% BV and/or ≥0.98 PP) whereas in the other two trees one showed separation between *B. natans* and *A. amoeboformis* and the other lacked *A. amoeboformis* sequence (Table 5). Five of the nine with chlorarachiophyte monophyly showed weak or moderate affinity of the chlorarachniophytes with CR group (CR type) (with ≥60% BV or ≥0.97 PP). Chlorarachniophyte PRK showed their origins directly from a red algal ancestor as reported previously in Chapter 2 (Red type). In the remaining three trees showing chlorarachniophyte monophyly, the phylogenetic position of the chlorarachniophytes within the red lineage was ambiguous (Unclassified type).

CR type trees.

Five genes of CR type encode plastid-targeted proteins that are directly or indirectly related to photosynthesis or plastid functions: GGR, RNABP, plastid division protein FtsZ (PDP, filamenting temperature-sensitive mutant Z), 6-phosphogluconate dehydrogenase (GND), and photosystem II stability assembly (PS2SAF). FtsZ is a prokaryotic homologue of the eukaryotic protein tubulin and can be considered a functional house-keeping gene of plastids (plastid division) (Löwe et al. 1998; Bi et al. 1991; Strepp et al. 1998). Figure 11 shows the robust monophyly of the *B. natans* and *A. amoeboformis* FtsZ proteins (with 93% BV and 0.99 PP). Chlorarachniophytes and stramenopiles plus the haptophyte *Emiliania* showed a modest clade (with 54% BV and 0.99 PP), to which the clade composed of three species of cyptophytes and the red alga Cyanidioschyzon were basal. On the other hand, PS2SAF is one of the four major multi-subunit protein complexes of the thylakoid membrane of oxygenic photosynthetic organisms. PS2SAF is essential for photosystem II (PSII) biogenesis and required for assembly of an early intermediate in PSII assembly that includes D2 (psbD) and cytochrome b559, and has been suggested to be required for chlorophyll a binding (Peltier et al 2002; Meurer et al. 1998). Phylogenetic results of PS2SAF (Figure 12) showed a robust chlorarachniophyte clade (with 100% BV and 1.00 PP). As in FtsZ, the chlorarachniophytes, stramenopiles and the haptophyte *Emiliania* constituted a clade (with 60-68% BV and 0.96-0.98 PP) from which two red algae and two cryptophytes were separated. Although the remaining three genes, GGR, RNABP and GND, showed affinity of chlorarachiophytes with the CR group with ≥60% BV and/or ≥0.98 PP, addition of alveolate OTUs lowered the support value below 50% BV and 0.95 PP in RNABP tree (Figures 13-15).

Red type trees.

As in Chapter 2, the tree topology of *PRK* genes obtained herein may indicate that they were transferred directly from a red algal ancestor to the common ancestor of the extant chlorarachniophytes because the chlorarachniophytes including A. amoebiformis sequences formed a robust clade that is closely related to the red algae (Figure 15). However, OTUs from secondary/tertiary eukaryotes with red algal plastids (CR group) were separated from the red algae, suggesting that the PRK of these eukaryotes might have experienced a gene replacement after the typical secondary EGT from the red algal ancestor. A similar separation between CR group and the lineage composed of red algae and chlorarachniophytes exists in the plastid-targeted SBP tree   (Figure 16).

Unclassified trees.

Due to the limited number of available sequences, phylogenetic position of the

chlorarachiophyte clade inside the red lineage was ambiguous in the three genes: ribosomal protein rps22 (RPS22), hypothetical protein Y (HP) and phosphoglycerate kinase (PGK). For example, PGK is present in all living organisms as one of the two ATP-generating enzymes in glycolysis. In the gluconeogenic pathway, PGK catalyzes the reversible transfer of a phosphate group from 1,3-bisphosphoglycerate to ADP producing 3-phosphoglycerate and ATP(Dhar et al. 2010; Blake 1997; Pielak et al. 2010). Chlorarachniophyte PGK genes are robustly monophyletic (with 97-99% BV and 0.99 PP). Chlorarachniophytes and red algae plus red algal secondary/tertiary algae were found to form a clade with 69-79% BV and 0.99-1.00 PP (Figure 17). However, no statistical support values were obtained regarding the phylogenetic position of the chlorarachniophytes within the red lineage in all of the three genes (Figure 17-19). .

Ambiguous trees.

  Two protein trees were found to lack resolution of monophyly of chlorarachniophytes. The ABC protein contained only a single chlorarachniophyte OTU (*B. natans*), and a putative membrane protein (PMP) showed a phylogenetic split between *B. natans* and *A. amoeboformis*. Perhaps gene duplications and possible replacements might have occurred to result in such a split in chlorarachniophyte PMP genes as represented by the three separate lineages of diatoms (each including a *Thalassiosira* OTU) (Figure 20,21).

Based on the extensive analysis of single-gene trees using OTUs from two evolutionarily distinct chlorarachiophytes, I identified eleven genes of cyanobacterial origin that showed chlorarachiophytes to have an affiliation with the red algae and/or CR group. Among the 11 single-gene trees, nine demonstrated robust monophyly of the two chlorarachniophyte OTUs. Thus, multiple horizontal gene transfers from the red lineage must have occurred before the divergence of the extant chlorarachniophytes, indicating a possible cryptic endosymbiosis from the red lineage. Alternatively, all of such EGT-like genes might have resulted from multiple horizontal gene transfers from long-time feeding of a single or closely related plastid-containing eukaryotes.

Five of the nine gene trees (GGR, RNABP, PDP, GND and PS2SAF) showed weak to moderate statistical support for the affiliation of the chlorarachniophyte homologues with those of the CR group (Figures 11-15). Although chlorarachniophyte PRK genes apparently showed their origins directly from a red algal ancestor (Figure 22) as reported previously in Chapter 1, the genes from CR group did not belong to the red algal lineage. Therefore, PRK genes of the CR group might have experienced the gene replacement after the ancient, typical secondary EGT scenario from the red algal ancestor. Thus, the apparent affiliation PRK homologues between chlorarachniophytes and red algae (Figure 22) might have resulted from the remove of PRK of CR group from the red lineage by such a

gene replacement. A similar situation may be considered in the red algal *SBP* homologues of chlorarachniophytes (Figure 16). Taken together, cryptic endosymbiosis of a secondary or tertiary plastid of the red lineage might have occurred in or before the common ancestor of extant Chlorarachniophyta.

Among the 11 genes, four (GGR, PS2SAF, GND and PRK) are obviously photosynthesis-related (Bailey-Serres et al. 1992; Meurer et al. 1998;Tanaka et al. 1999;; Peltier et al. 2002; Petersen et al. 2006) and five (ABC, PDP, rps22, RNABP and PGK) are plastid-related (L öwe et al. 1998; Bi et al. 1991; Strepp et al. 1998; Dhar et al. 2010; Blake 1997; Pielak et al. 2010; Jasinski et al. 2003; M øller et al. 2001; Li et al.1995; Yang et al. 1996). Furthermore, all eukaryotic OTUs in the 11 trees possess plastids except several stramenopile and excavate taxa in PGK, RNABP and GND trees (Figure 14,15,17). Thus, those 11 genes most likely directly originate from the pre-existing photosynthetic eukaryotes with secondary or tertiary red plastids. Thus, before the secondary endosymbiosis that gave rise to the green algal plastid that currently exists, the ancestor of chlorarachniophytes might have harbored a red-algal plastid of CR group. Subsequently, the secondary endosymbiosis of the green alga might have discarded such pre-existing yellow or brown plastids in the ancestor of the chlorarachniophytes.

Tables

Table 4 Sequence data used in two sets of local databases

localDB of species used	group A	group B	saurce
Calliarthron tuberculosum	Y	Y	NCBI
Cyanidioschizon merolae	Y	Y	Matsuzaki et al. 2004
Ectocarpus siliculosus	Y	Y	JGI
Pyropia yezoensis	Y	Y	NCBI
Aureococcus anophagefferens	N	Y	JGI
Cyanophora paradoxa	N	Y	NCBI
Euglena gracilis	N	Y	Maruyama et al. 2011
Emiliania huxleyi	N	Y	JGI
Galdieria sulphuraria	N	Y	Galdieria sulphuraria Genome Project at MSU*
Amorphochlora amoeboformis	N	Y	unpublished
Peranema trichophorum	N	Y	Maruyama et al. 2011
Phytophthora capsici	N	Y	NCBI
Porphyridium cruentum	N	Y	NCBI
Schizochytrium aggregatum	N	Y	NCBI
Schizochytrium limacinum	N	Y	NCBI

* http://genomics.msu.edu/galdieria/

Table 5. List of 11 chlorarachniophyte red-derived genes of cyanobacterial origin resolved in this study.

gene	abbreviation	function	Statiscal support values for chlorarachniophyte affiliation with red lineage (CR group and/or red algae) (RAxML BV/PP). Parentheses in right side represent resuluts of phylogeny with alveolate OTUs.	Statiscal support values for chlorarachniophytes monophyly (RAxML BV/PP). Parentheses in right side represent resuluts of phylogeny with alveolate OTUs.	Statiscal support values for chlorarachnio-phyte affiliation inside the red linegae (RAxML BV/PP). Parentheses in right side represent resuluts of phylogeny with alveolate OTUs	Classification of gene based on chlorarachnio-phyte affiliation within red lineage	DDBJ/NCBI/EMBL accession number of chlorarachniophyte sequence
Geranylgeranyl reductase	GGR	synthesis phytol for the synthesis of tocopherol and chlorophyll	(96/1.00)|(61/0.96)	(73/0.99)|(71/0.99)	(76/0.99)|(77/0.99)	CR type	XX 000000
mRNA binding protein	RNABP	helping process chloroplast precursor mRNAs	(69/-)|(62/-)	(99/1.00)|(99/1.00)	(66/0.96)|(-/-)	CR type	XX 000000

Name	Abbr.	Function				Type	Accession
plastid division protein Ftsz	PDP	prokaryotic homologue to the eukaryotic protein tubulin	(99/1.00)	(99/0.99)	(54/0.99)	CR type	XX 000000
6-phosphogluconate dehydrogenase	GND	stimulates the decarboxylation of 6-phosphor-gluconate to ribulose-5-phosphate and CO_2 and generates a molecule of NADPH	(60/0.99)|(71/1.00) with CR	(100/1.00)|(100/1.00)	(60/0.99)|(71/1.00)	CR type	XX 000000
photosystem II stability assembly factor	PS2SAF	required for assembly of an early intermediate in PSII assembly	(100/1.00)|(100/1.00)	(100/1.00)|(100/1.00)	(60/0.98)|(68/0.96)	CR type	XX 000000
ATP binding cassette transporter	ABC	transports cation against a concentration gradient	(100/1.00)|(50/-)	No sufficient OTU	unresolved	unclassified	XX 000000

putative membrane protein	PMP	unknown	(76/0.96)|(77/0.97)	separated	unresolved	unclassified	XX 000000
phosphoribulokinase	PRK	ATP + D-ribulose 5-phosphate to ADP + D-ribulose 1,5-bisphosphate	(79/0.99)(52/-) with red algae	(100/1.00)|(100/1.00)	separated	red type	XX 000000
ribosomal protein rps22	RPS22	encodes nuclear-coded chloroplast ribosomal protein	(83/0.98)	(100/1.00)	unresolved	ambiguous type	XX 0000000
hypothetical protein	HP	unknown	(79/-)	(100/1.00)	unresolved	ambiguous type	XX 0000000
phosphoglycerate kinase	PGK	ATP-generating enzymes in glycolysis	(79/1.00)|(69/0.99)	(99/0.99)|(97/1.00)	unresolved	ambiguous type	XX 000000

Figures

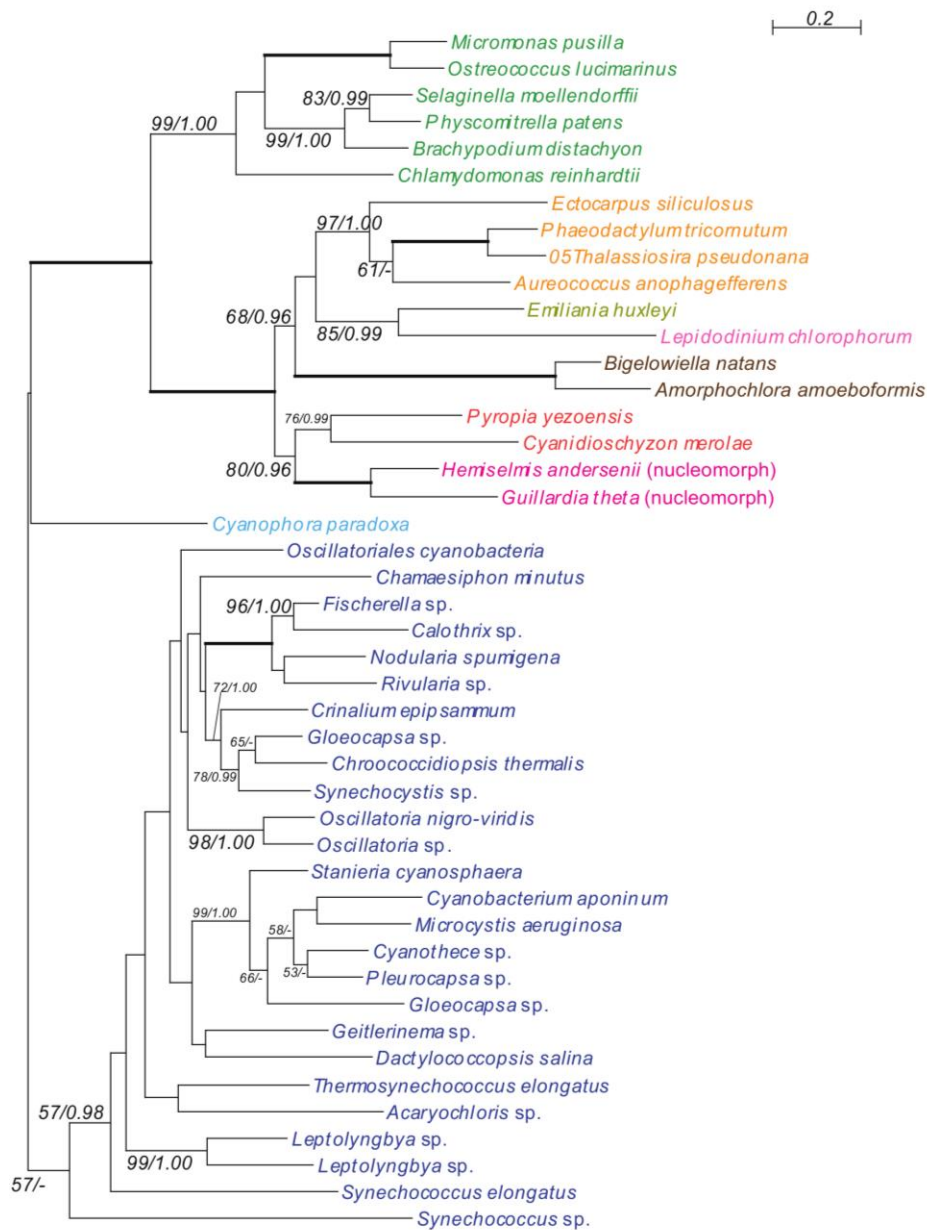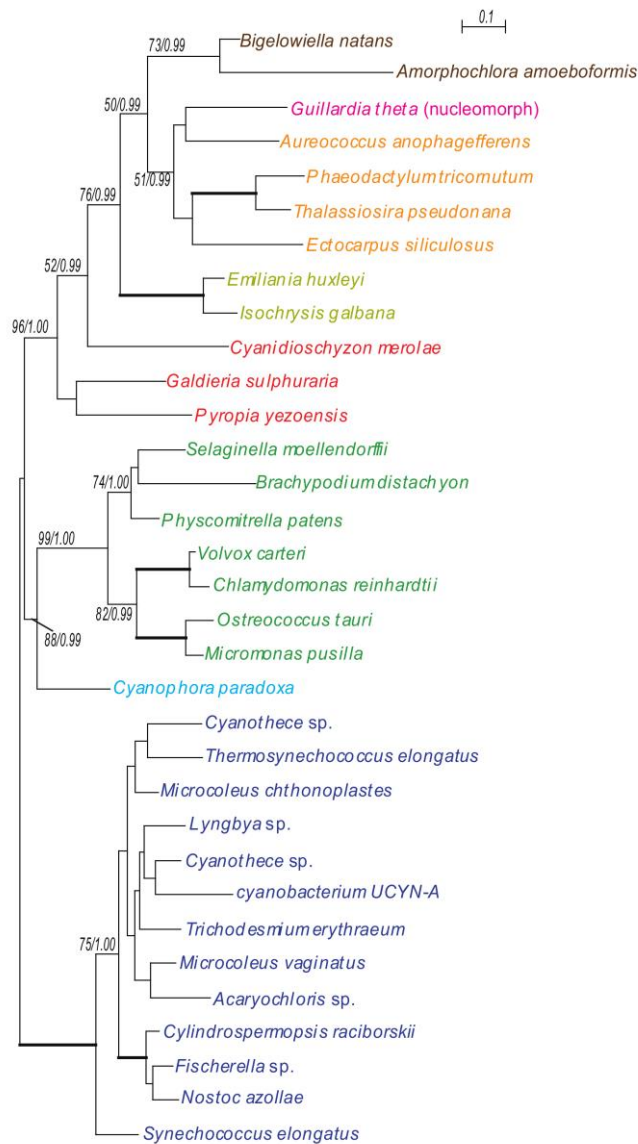**Figure 10. Outputs overlapping status of two pipelines**

Figure 11. Phylogeny of PDP (FtsZ) showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The tree was inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta.

Figure 12A. Phylogeny of PS2SAF showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.

The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure lacks alveolate OTUs.
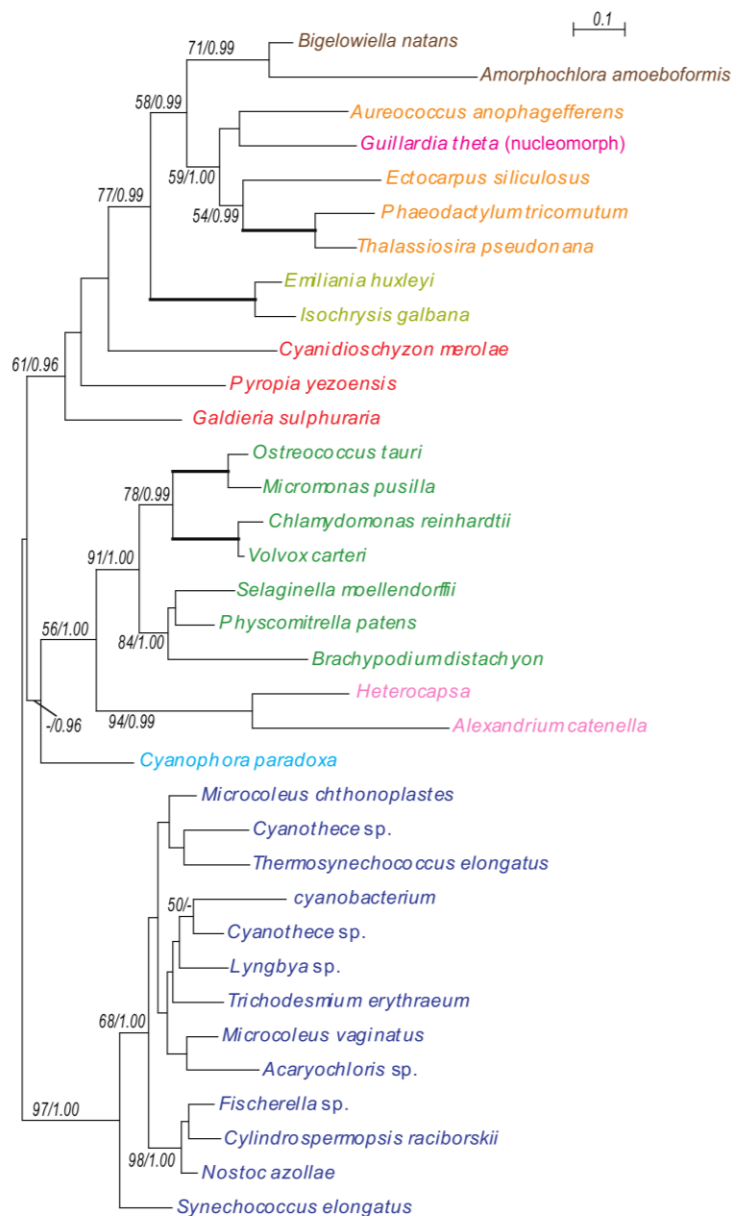
Figure 12B. Phylogeny of PS2SAF showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure Contains alveolate OTUs.
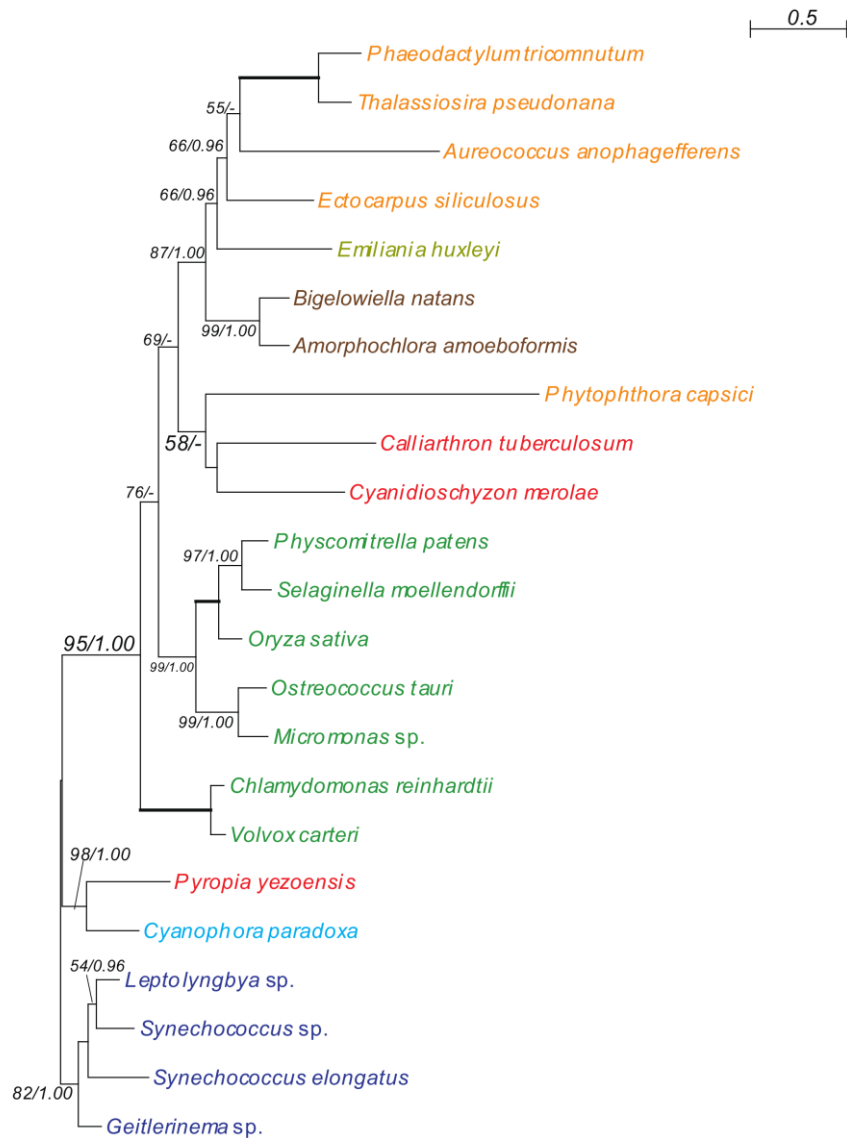
Figure 13A. Phylogeny of GGR showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure lacks alveolate OTUs.
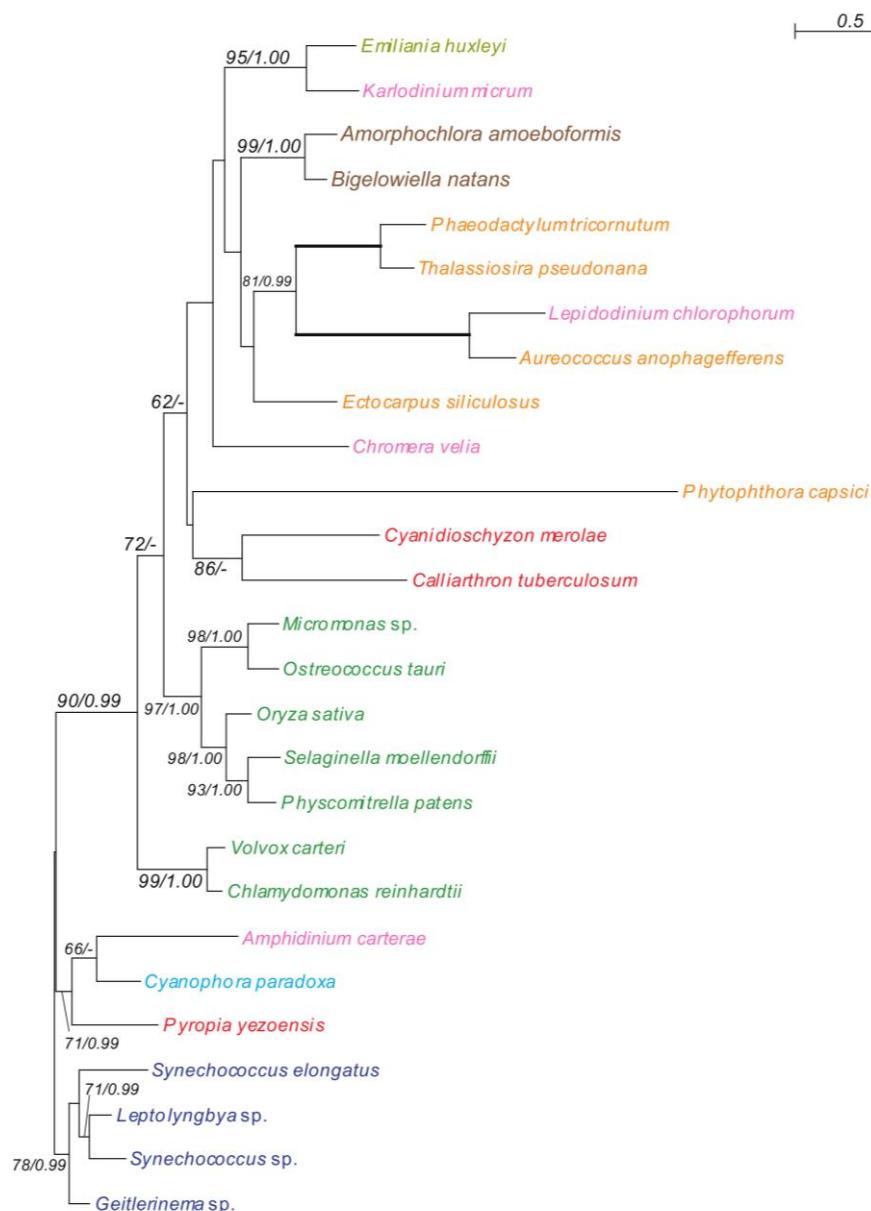
Figure 13B. Phylogeny of GGR showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.

The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure Contains alveolate OTUs.
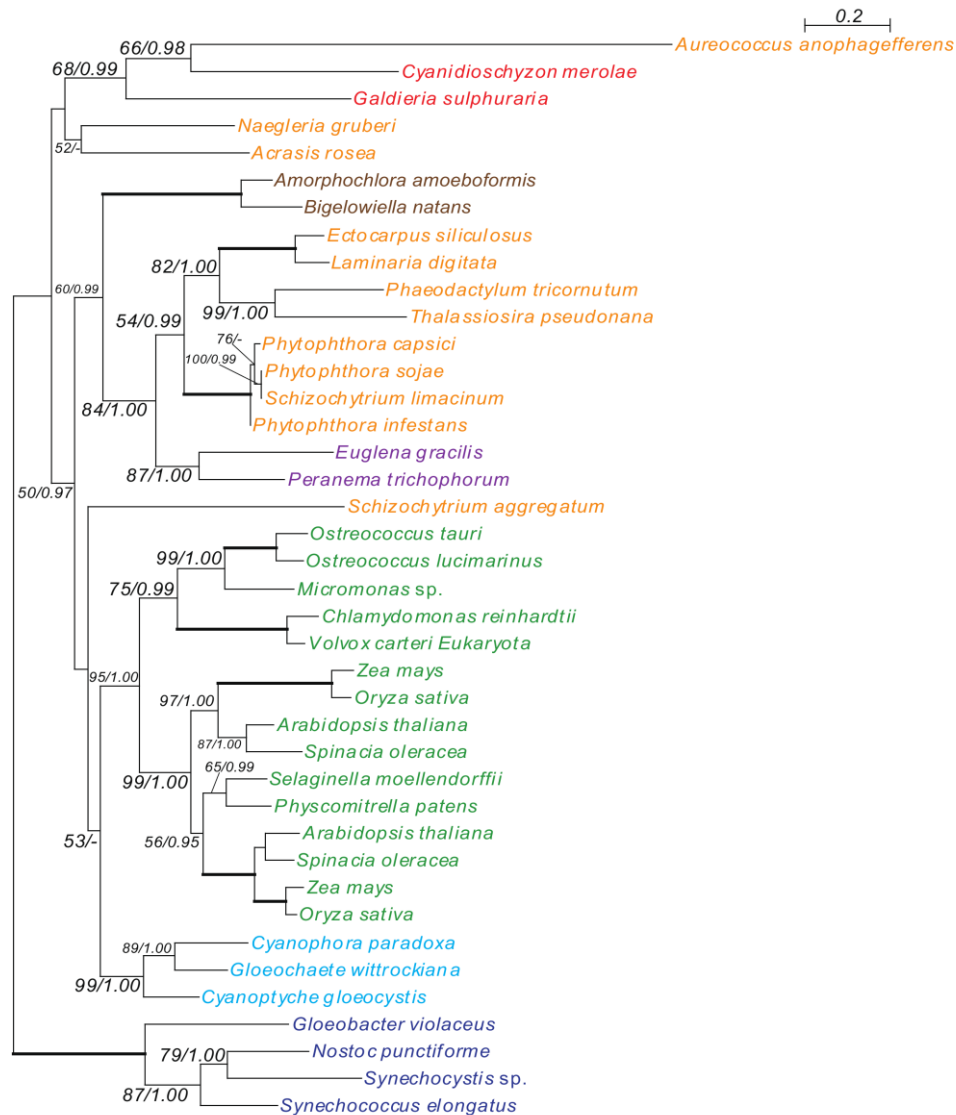
Figure 14A. Phylogeny of RNABP showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure lacks alveolate OTUs.

Figure 14B. Phylogeny of RNABP showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.

The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure Contains alveolate OTUs.
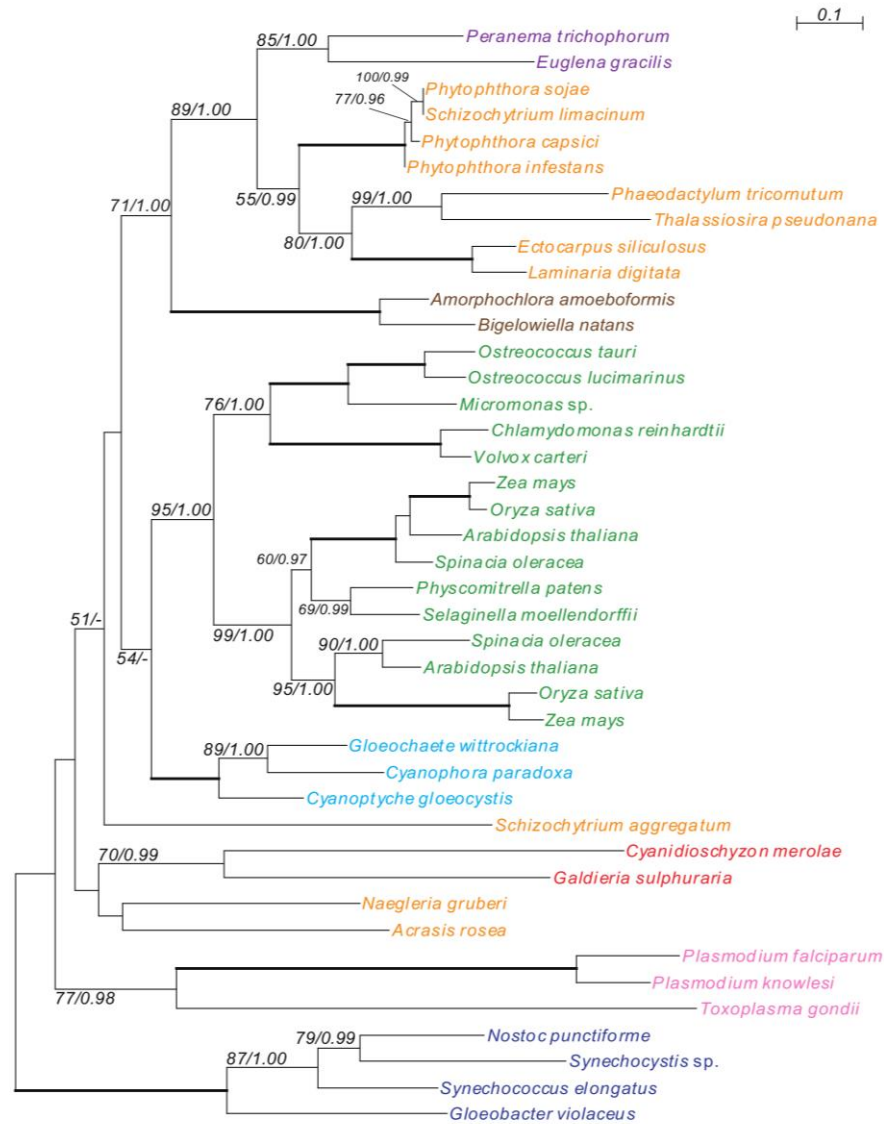
Figure 15A. Phylogeny of GND showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure lacks alveolate OTUs.
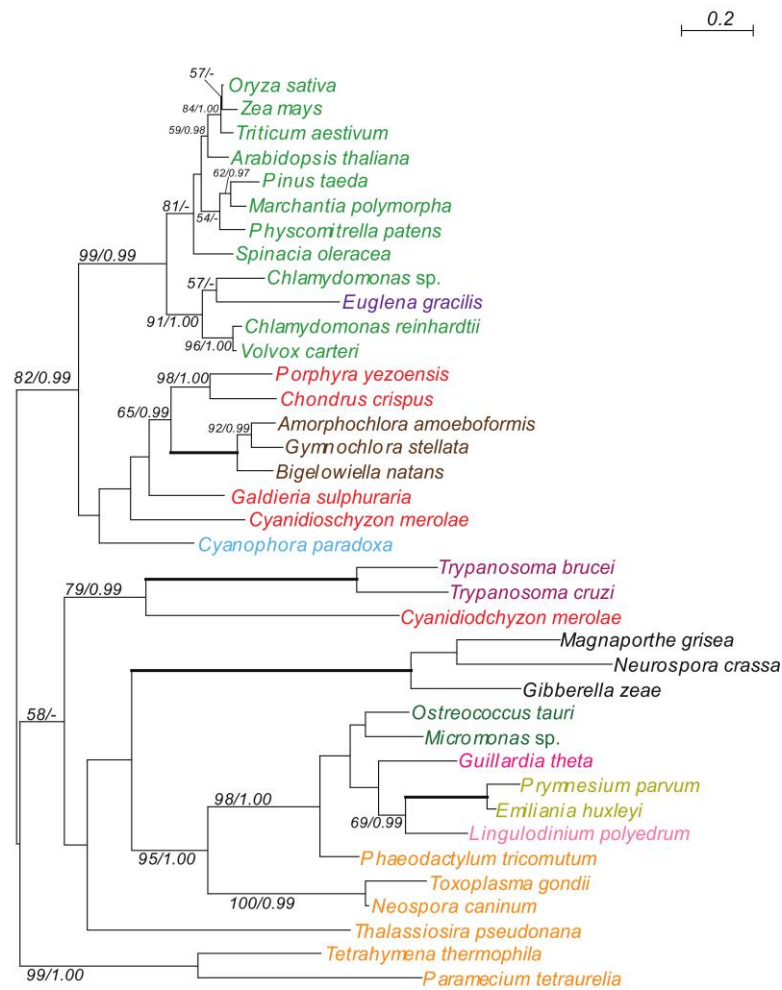
Figure 15B. Phylogeny of GND showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure Contains alveolate OTUs.
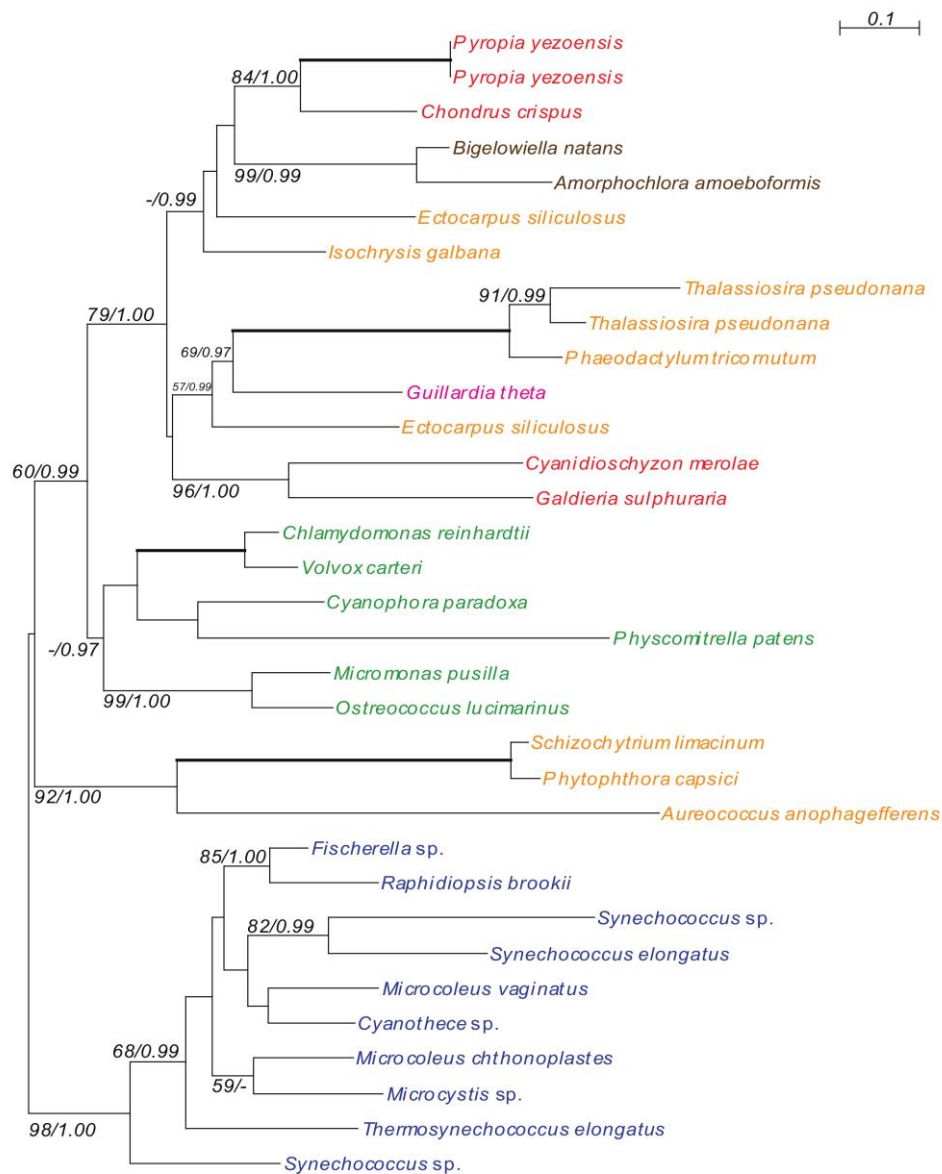
Figure 16. Phylogeny of SBP showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta.
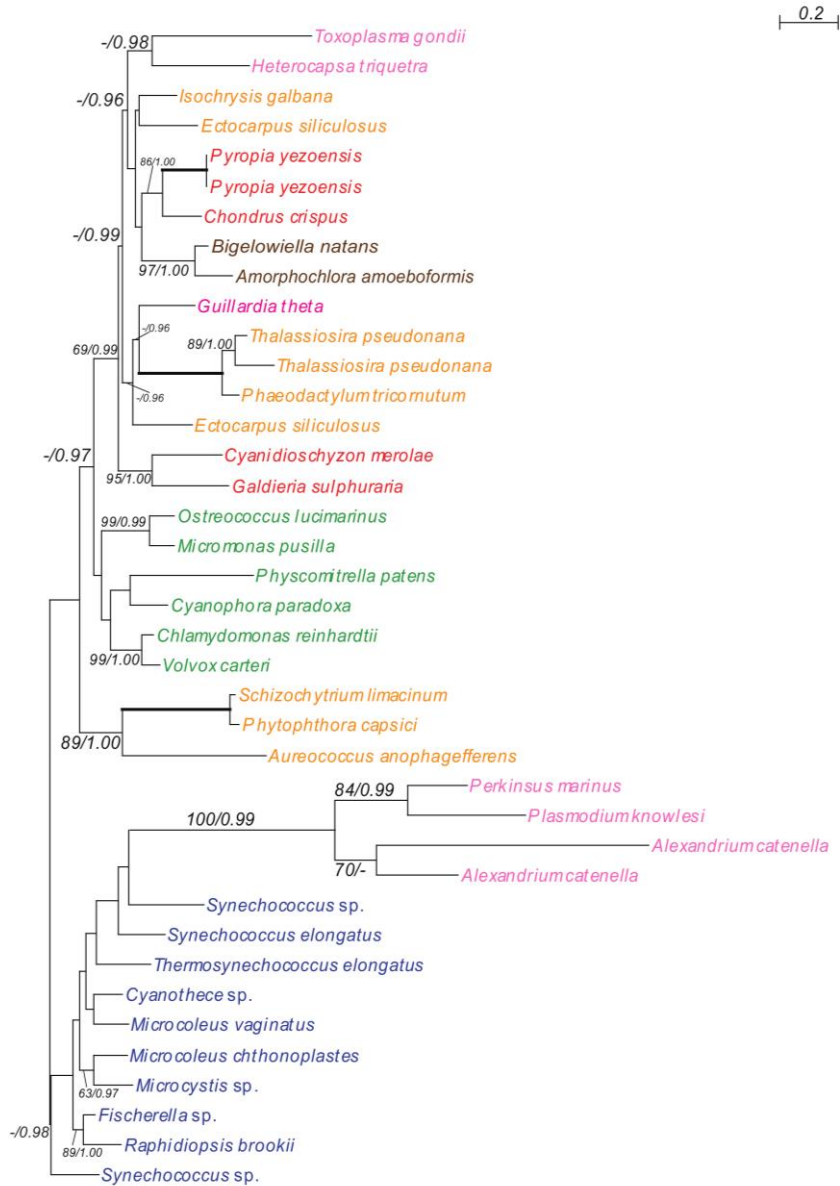
Figure 17A. Phylogenies of PGK showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure lacks alveolate OTUs.
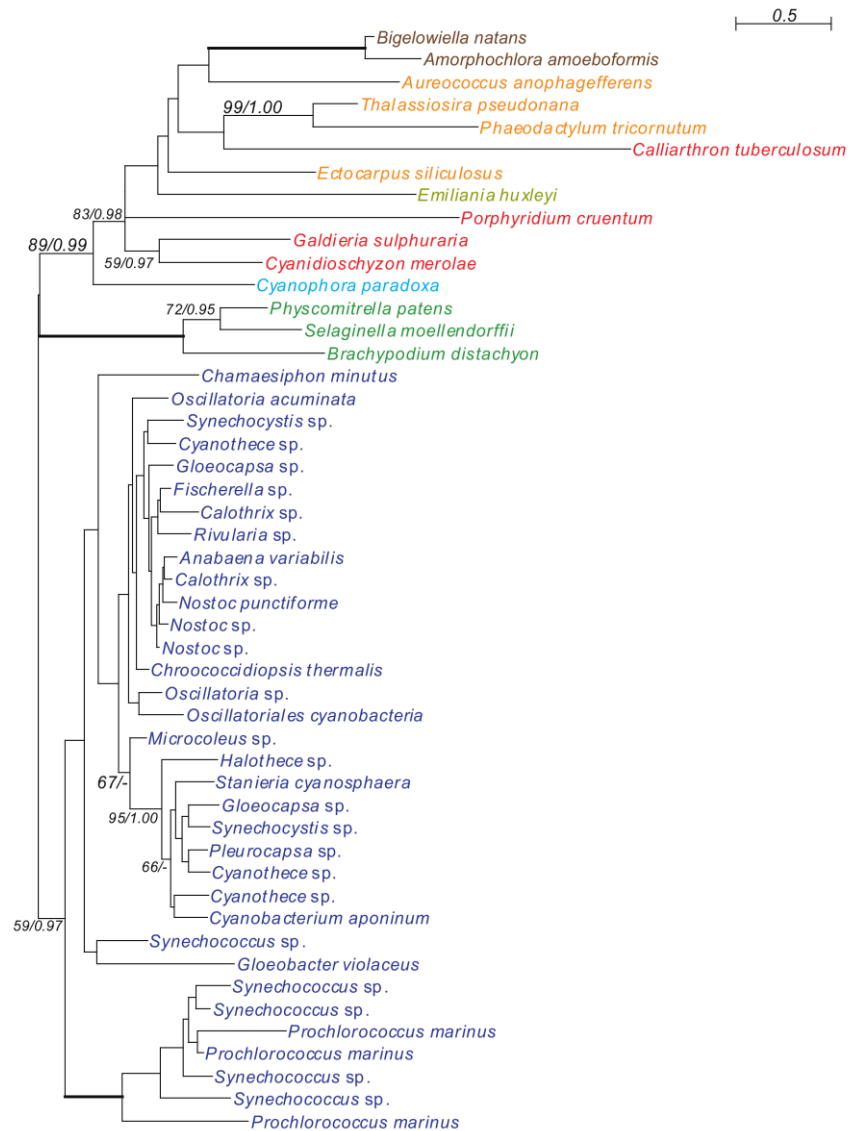
Figure 17B. Phylogenies of PGK showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues. The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure contains alveolate OTUs.
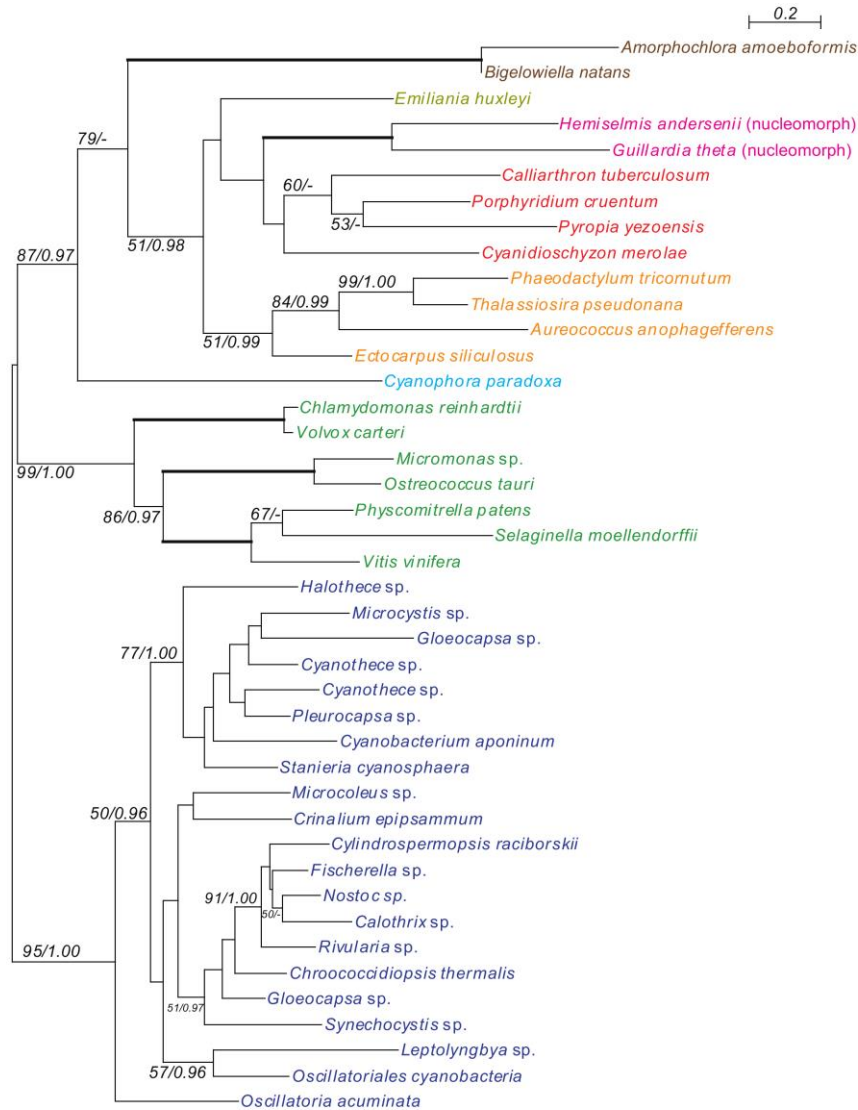
Figure 18. Phylogeny of RPS22 showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta.

Figure 19. Phylogeny of HP showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta.
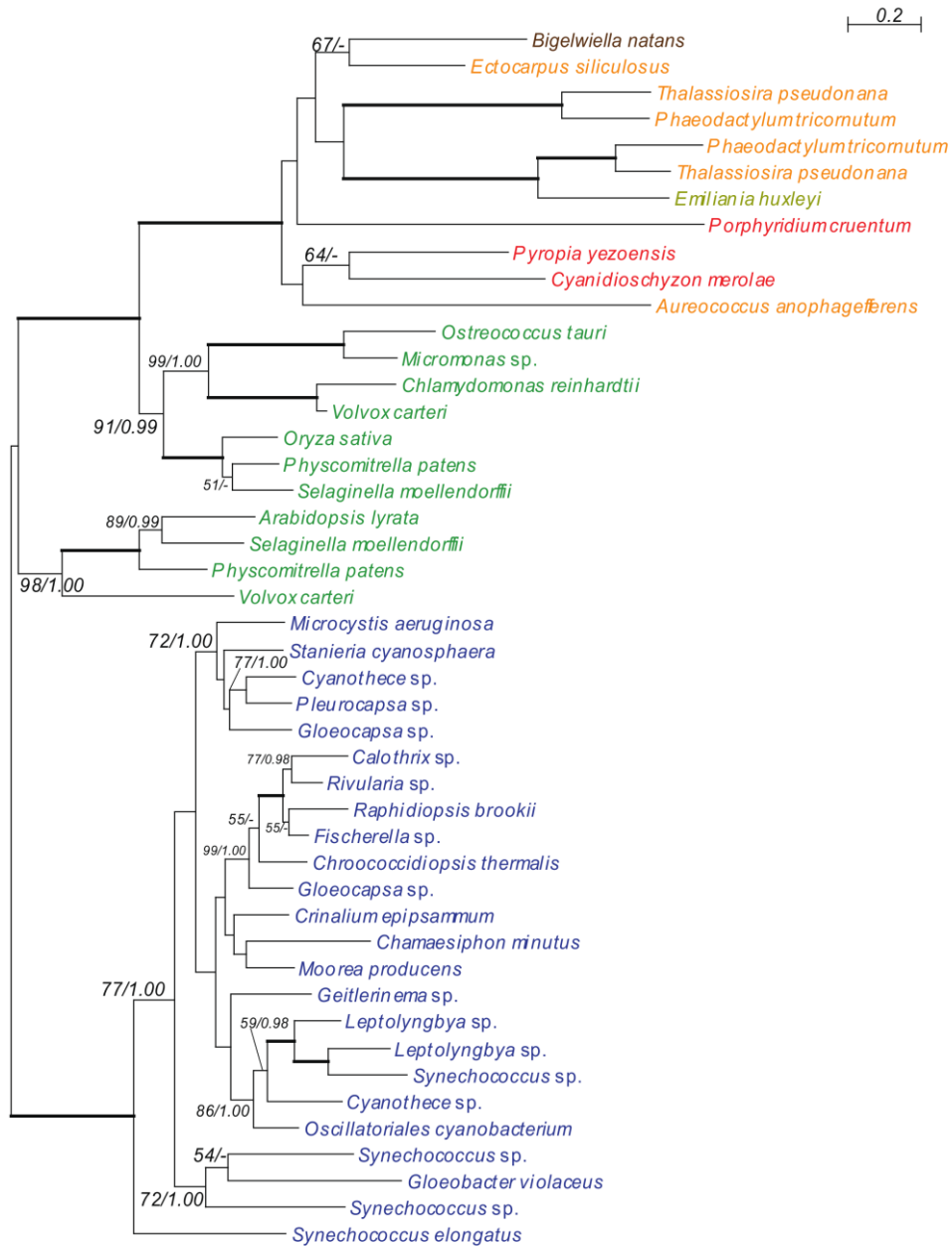
Figure 20A. Phylogeny of ABC showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure lacks alveolate OTUs.
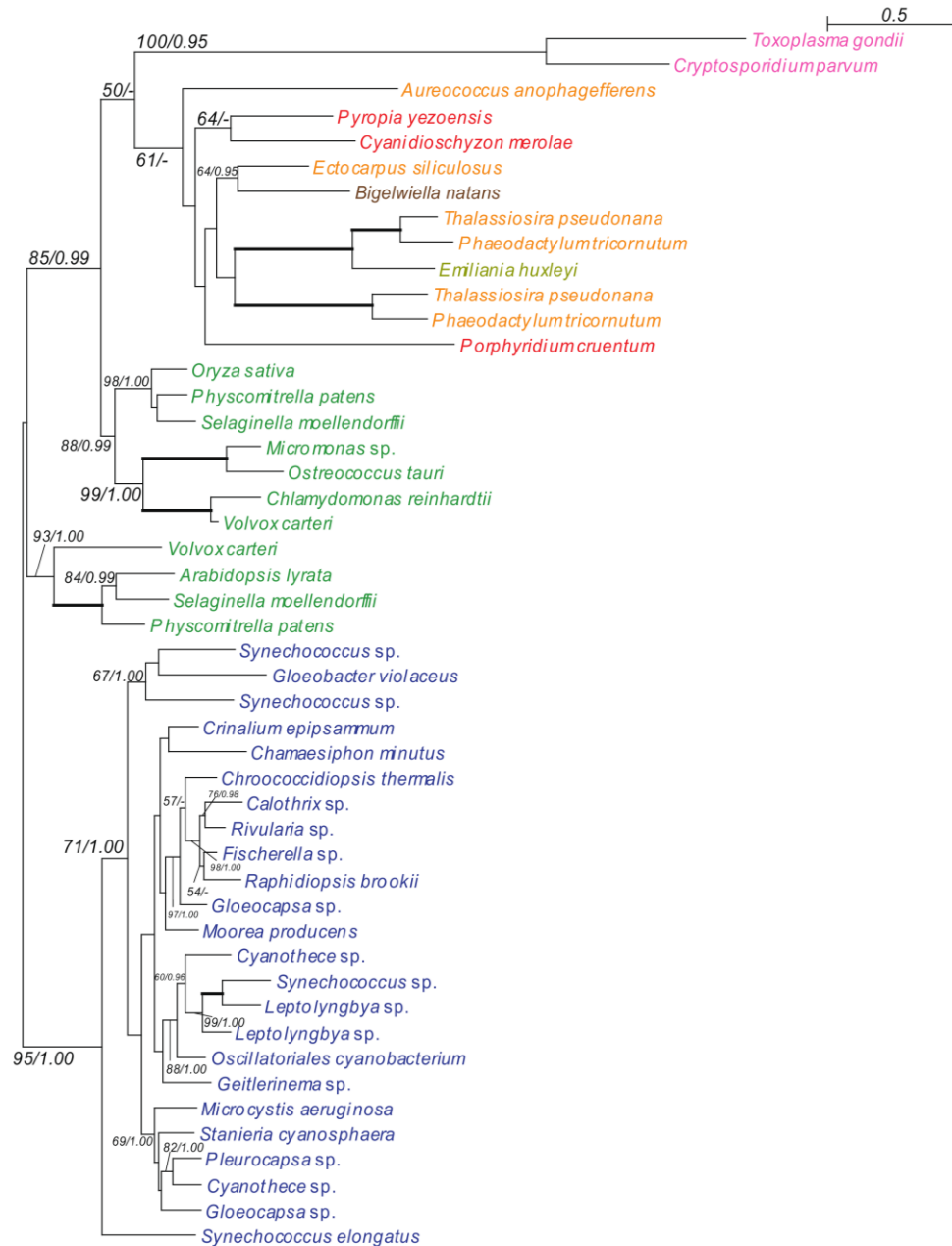
Figure 20B. Phylogeny of ABC showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure contains alveolate OTUs.
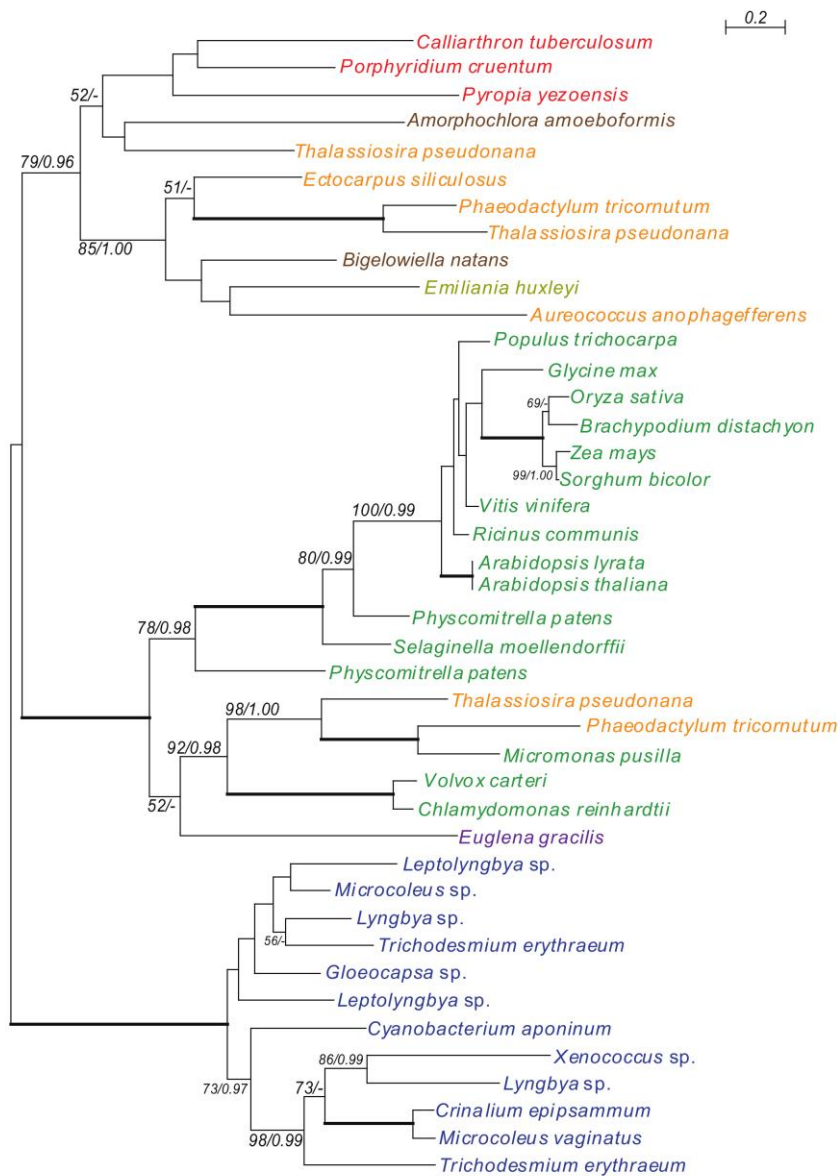
Figure 21. Phylogeny of PMP showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure lacks alveolate OTUs.
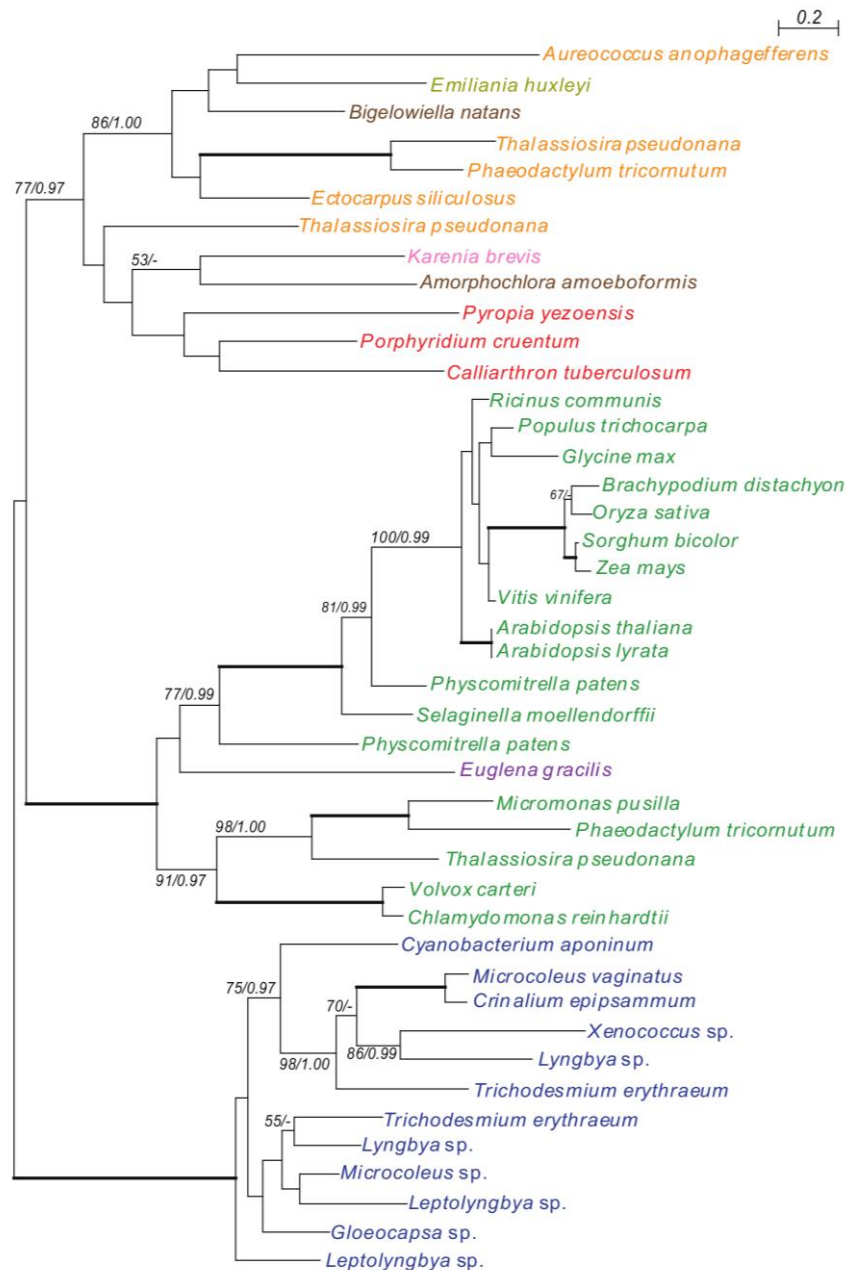
Figure 21B. Phylogeny of PMP showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure contains alveolate OTUs.
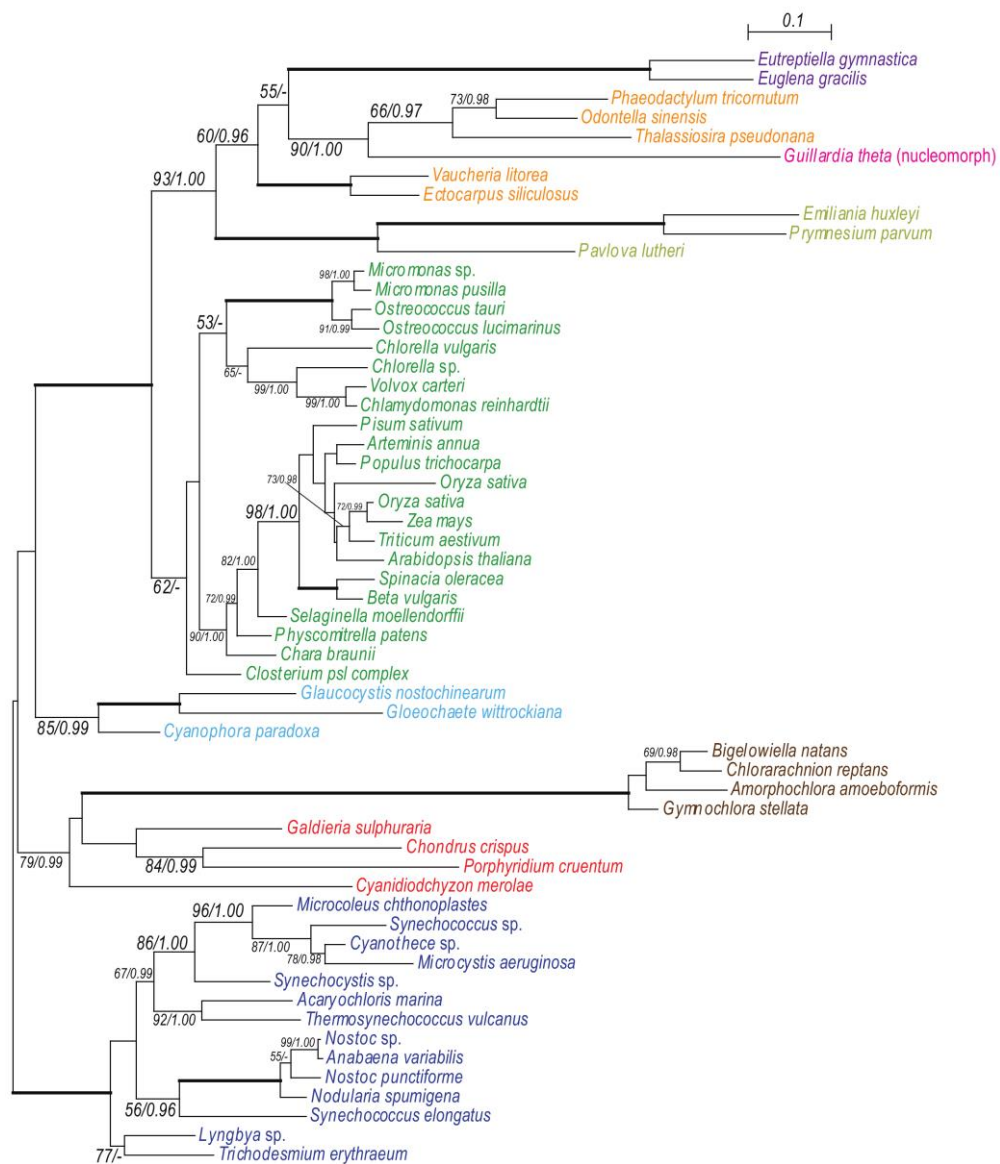
Figure 22A. Phylogeny of PRK with additional chlorarachniophyte OTUs showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure lacks alveolate OTUs.
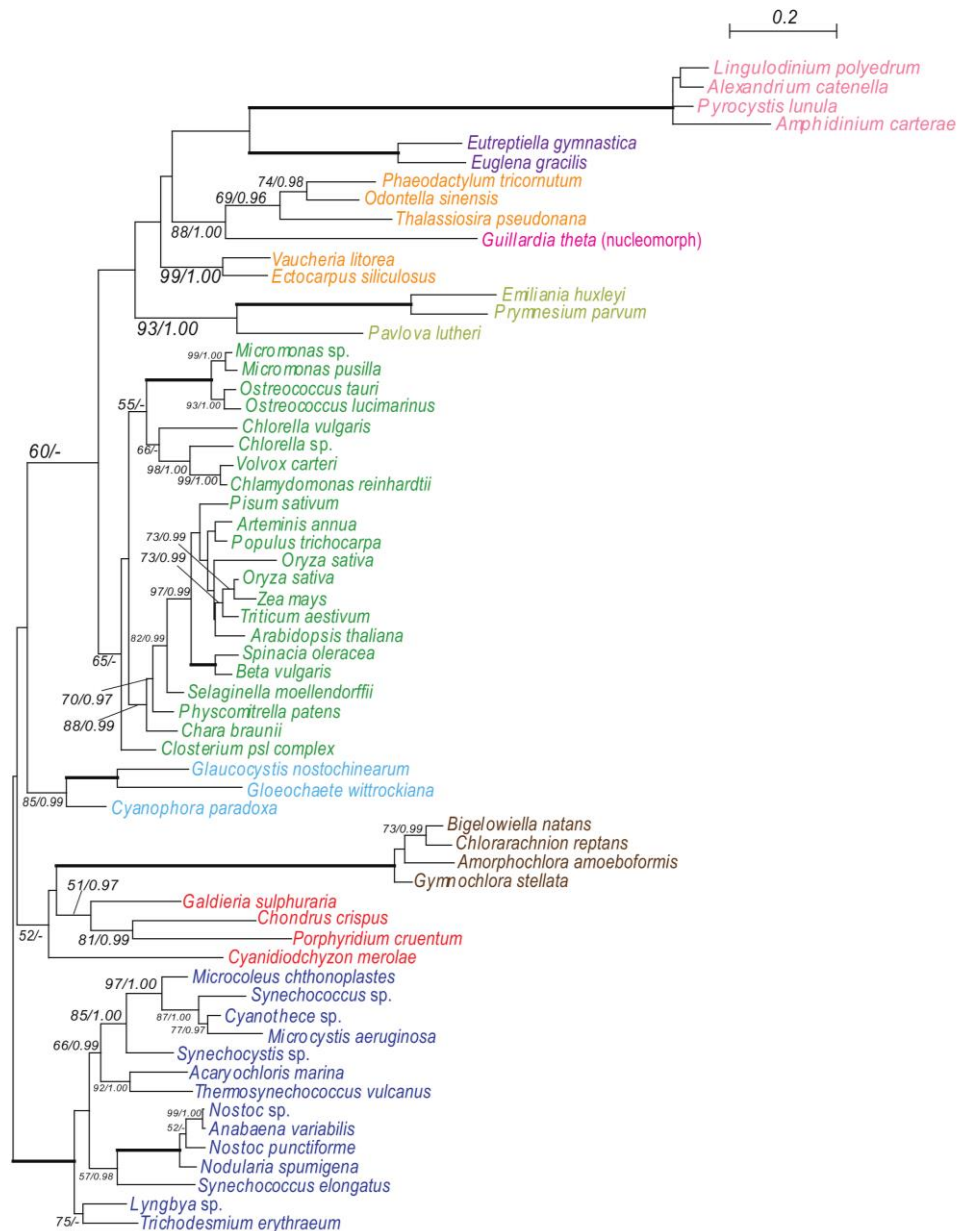
Figure 22B. Phylogeny of PRK with additional chlorarachniophyte OTUs showing chlorarachiophyte proteins closely related to red secondary/tertiary eukaryote homologues.
The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (≥50% bootstrap values or ≥0.95 posterior probability) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values are 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. This figure contains alveolate OTUs.

Chapter 4

General Discussion

Given that the origin of the green secondary plastids of extant chlorarachiophytes would represent a replacement of the pre-existing CR group plastid (see above), two alternate hypotheses can be proposed regarding the pre-existing CR group plastid or origin of the possible EGT genes from CR group in the Chlorarachniophyta (Figure 23).

The first scenario considers that the pre-existing CR group plastid might have originated from tertiary or quaternary endosymbiosis of the cryptic red plastid from the ancestral lineage of the CR group ("EGT scenario," Figure 23A)(Archibald 2012). During such cryptic endosymbiosis, EGT might have occurred, resulting in nuclear-encoded, plastid-targeted genes of CR group. Then, majority of such CR group EGT genes might have been replaced by the green genes via secondary EGT of the green plastid of the extant chlorarachniophytes. On the other hand, the rest of the CR group genes in the host nucleus might have been retained in the host nucleus, not affected by the green EGT gene replacement, and I now identify them as red lineage genes in the chlorarachiophyte nuclear genome.
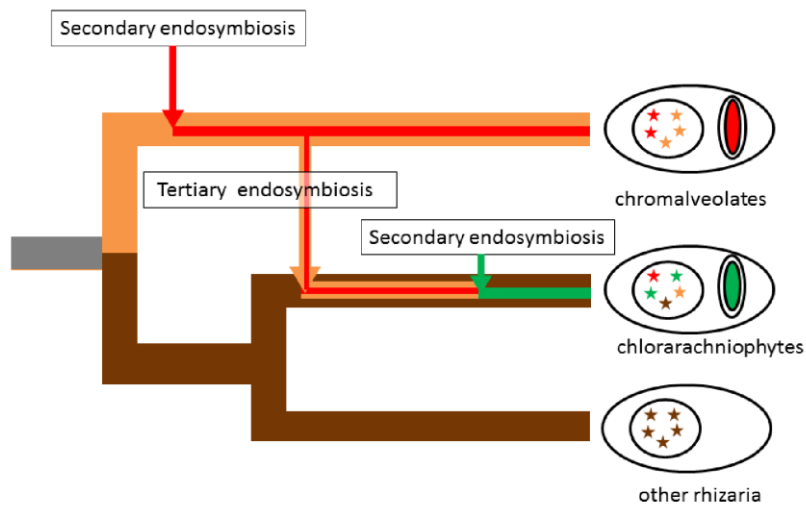
The alternate hypothesis posits that the current CR group EGT genes in the host chlorarachiophyte genomes might originate directly, via vertical transmission, from the ancient secondary endosymbiosis that might have occurred in the common ancestor of SAR ("host relics scenario," Figure 23B) (Archibald 2012). In this hypothesis, there is no need to consider additional EGT from alveolates/stramenopiles. Thus, the common ancestor of the extant chlorarachniophytes might have retained the red secondary plastid originating from the ancient secondary endosymbiosis in the common ancestor of SAR. The red secondary plastid might have been then replaced by the green plastid of the extant Chlorarachniophyta. The "red" genes of the chlorarachniophye nuclear genomes identified in the present study might be ancient relics of host genomic contents. However, provided extant lineages in Rhizaria lack plastids except for chlorarachniophytes (Parfrey et al. 2010), there must have been multiple losses of plastids within Rhizaria. Thus, parsimony principles seem to favor the former EGT scenario.

All things considered, unlike what I expected previously in Chapter 1, the ancestor of extant chlorarachniophytes may not have experienced the secondary endosymbiosis of a red alga. Instead, prior to acquisition of the green alga type plastid, there was once an ancestral plastid of CR group (or SAR (Archibald 2012; Burki et al. 2007; Hampl et al. 2009; Frommolt et al. 2008)) in chlorarachniophytes. As far as my results show, multiple transfers of cyanobacterial genes of CR group were transferred to the host genome of the ancestral chlorarachniophyte, which then diversified to a lineage with functional green plastids.
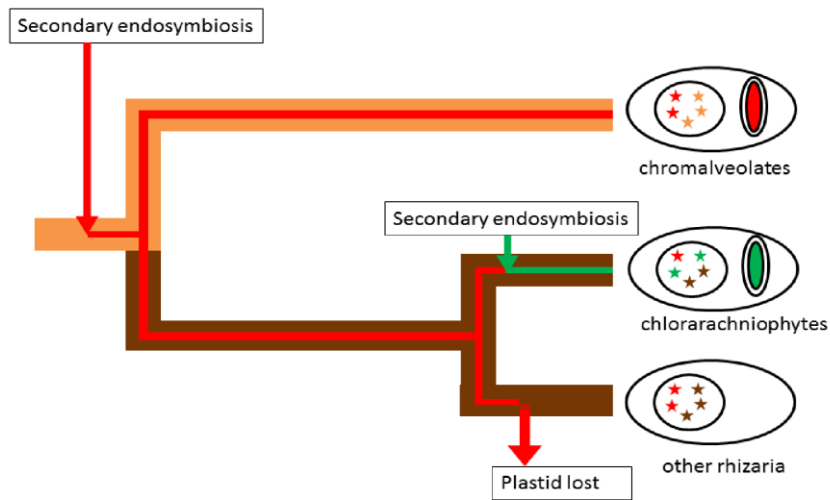
A recent study of the plastid-possessing euglenid *E.gracilis* and plastid-lacking

euglenid *P.trichophorum* suggested a possible cryptic tertiary endosymbiont of a red plastid in the common ancestor of euglenids, and such a tertiary plastid might have been replaced by the ancestor of the extant green secondary plastid of the euglenids during the secondary endosymbiosis (Maruyama et al. 2011). The present study suggested a similar plastid replacement in the origin of the green secondary plastids in the Chlorarachniophyta. In addition, the secondary red plastids of stramenopiles such as diatoms might have evolved via replacement of the pre-existing green plastid (Nozaki et al. 2009; Moustafa et al. 2009). Thus, secondary plastid endosymbiosis seems unlikely in the plastid lacking, heterotrophic host, and might have been supported by the pre-existing plastid-targeting system in the host nucleus. However, Pillet and Pawlowski (Pillet et al. 2013) failed to show any nuclear-encoded genes having a function in photosynthetic activity or plastid maintenance in a transcriptome of the foraminiferan *Elphidium margaritaceum* (Rhizaria) in which acquired plastids remain active inside the foraminiferan cell for several months (known as kleptoplastidy). This suggests no pre-existing gene systems are needed for establishment of the long-term maintenance of photosynthetic organisms or plastids in the heterotrophic host. Thus, kleptoplastidy may not always represent an incipient stage of plastid endosymbiosis.

Figures

Figure 23. Schematic diagram of two alternating scenarios of origin of "red" genes in the chlorarachniophyte nuclear genomes.
Both scenarios are based on the hypothesis that the extant green secondary plastids of Chlorarachniophyta originate from plastid replacement of the pre-existing red secondary or tertiary plastids.

Acknowledgements

References

Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A, Hoppenrath M, Lara E, Le Gall L, Lynn DH, McManus H, Mitchell EA, Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S, Shadwick RS, Schoch CL, Smirnov A, Spiegel FW: **The revised classification of eukaryotes**. *J Eukaryot Microbiol* 2012, 59: 429-493.

Adl SM, Simpson AGB, Farmer MA, Andersen RA, 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 MFJR: **The new higher level classification of eukaryotes with emphasis on the taxonomy of protists**. *J Eukaryot Microbiol* 2005, 52:399–451.

Archibald JM, Keeling PJ: **Actin and ubiquitin protein sequences support a cercozoa/floraminiferan ancestry for the plasmodiophorid plant pathogens**. *J Eukaryot Microbiol* 2004, 51: 113-118.

Archibald JM: **The evolution of algae by secondary and tertiary endosymbiosis. Genomic insights into the biology of algae. Piganeau**, *G. (ed.) Elsevier Ltd: Academic Press* 2012, 87-118.

Archibald JM, Lane CE: **Going, going, not quite gone: Nucleomorphs as a case study in nuclear genome reduction**. J Hered 2009, 582-590.

Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ: **Lateral gene transfer and the evolution of plastid-targeted proteins in the seconda plastid-containing alga *Bigelowiella natans***. *PNAS* 2003, 100:7678-7683.

Bailey-Serres J, Nguyen MT: **Purification and characterization of cytosolic 6-phosphogluconate dehydrogenase isozymes from maize**. *Plant Physiol* 1992, 100:1580-1583.

Bi EF, Lutkenhaus J: **FtsZ ring structure associated with division in *Escherichia coli***. *Nature* 1991, 354: 161–164.

Blair JE, Shah P, Hedges SB: **Evolutionary sequence analysis of complete eukaryote genomes**. *BMC Bioinformatics* 2005, 6:53.

Blake C: **Phosphotransfer hinges in PGK**. *Nature* 1997, 385:204 – 205.

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

Brandes HK, Hartmann FC, Lu T-Y, Larimer FW: **Efficient expression of the gene for spinach phosphoribulokinase in Picia pastoris and utilization of the recombinant enzyme to explore the role of regulatory cysteinyl residue by site-directed mutagenesis.** *J Biol Chem* 1996, 271: 6490–6496.

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: **Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, Telonemia and Centroheliozoa, are related to photosynthetic chromalveolates.** *Genome. Bio Evol* 2009: 231-238.

Burki F, Shalchian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI, Jakobsen KS, Pawlowski J: **Phylogenomics reshuffles the eukaryotic supergroups.** *PLoS ONE* 2007,2: e790.

Cavalier-Smith T: Principles of protein and lipid targeting in secondary symbiogenesis: **Euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree.** *J Euk Microbiol* 1999, 46:347-366.

Chantangsi C, Hoppenrath M, Leander BS: **Evolutionary relationships among marine cercozoans as inferred from combined SSU and LSU rDNA sequences and polyubiquitin insertions.** *Mol Phylogenet Evol* 2010, 57: 518-527.

Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, Hopkins JF, Kuo A, Rensing SA, Schmutz J, Symeonidi A, Elias M, Eveleigh RJ, Herman EK, Klute MJ, Nakayama T, Oborn k M, Reyes-Prieto A, Armbrust EV, Aves SJ, Beiko RG, Coutinho P, Dacks JB, Durnford DG, Fast NM, Green BR, Grisdale CJ, Hempel F, Henrissat B, H öppner MP, Ishida K, Kim E, Kořený L, Kroth PG, Liu Y, Malik SB, Maier UG, McRose D, Mock T, Neilson JA, Onodera NT, Poole AM, Pritham EJ, Richards TA, Rocap G, Roy SW, Sarai C, Schaack S, Shirato S, Slamovits CH, Spencer DF, Suzuki S, Worden AZ, Zauner S, Barry K, Bell C, Bharti AK, Crow JA, Grimwood J, Kramer R, Lindquist E, Lucas S, Salamov A, McFadden GI, Lane CE, Keeling PJ, Gray MW, Grigoriev IV, Archibald JM: **Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs.** *Nature* 2012, 492(7427):59-65.

Deschamps P, Moreira D: **Reevaluating the green contribution to diatom genomes**. *Genome Biol Evol* 2012, 4: 795–800.

Dhar A, Samiotakis A, Ebbinghaus S, Nienhaus L, Homouz D, Gruebele M, Cheung MS: **Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding**. *Proc Natl Acad Sci USA* 2010, 107:17586 – 17591.

Edgar RC: MUSCLE: **a multiple sequence alignment method with reduced time and space complexity**. *BMC Bioinformatics* 2004b, 5:113.

Edgar RC: MUSCLE: **multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004a,32: 1792-1797.

Frommolt R, Werner S, Paulsen H, Goss R, Wilhelm C, Zauner S, Maier UG, Grossman AR, Bhattacharya D, Lohr M: **Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis**. *Mol Biol Evol* 2008, 25: 2653–2667.

Gould SB, Waller RF, McFadden GI: **Plastid evolution**. *Annu Rev Plant Biol* 2008, 59: 491–517.

Gouy M, Guindon S, Gascuel O: SeaView version 4 : **a multiplatform graphical user interface for sequence alignment and phylogenetic tree building.** *Mol Biol Evol* 2010, 27: 221-224.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A: **Full-length transcriptome assembly from RNA-seq data without a reference genome.** *Nature Biotechnolgy* 2011 29: 644-652.

Guillard RRL, Hargraves PE: **Stichochrysis immobilis is a diatom, not a chrysophyte.** *Phycologia* 1993, 32: 234-236.

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

Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rummele SE, Bhattacharya D: **Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates.** *Mol Biol Evol* 2007, 24: 1702–1713.

Hallegraeff GM, Anderson DM, Cembella AD, Enevoldsen HO (eds): **Manual on Harmful Marine Microalgae. 2nd edn. Place de Fontenoy, Paris:** *the United Nations Educational, Scientific and Cultural Organization*, 2003.

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AG, Roger AJ: **Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups"**. *Proc Natl Acad Sci USA* 2009, 106: 3859–3864.

Harrison DH, Runquist JA, Holub A, Miziorko HM: **The crystal structure of phosphoribulokinase from *Rhodobacter sphaeroides* reveals a fold similar to that of adenylate kinase**. *Biochemistry* 1998, 37: 5074-85.

Hedges SB, Blair JE, Venturi ML, Shoe JL: **A molecular timescale of eukaryote evolution and the rise of complex multicellular life**. *BMC Evol Biol* 2004, 4:2.

Huang J, Gogarten JP: **Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids?**. *Genome Biol* 2008, 8: R99.

Huelsenbeck JP, Ronquist F: MRBAYES: **Bayesian inference of phylogeny**. *Bioinformatics* 2001, 17: 754-755.

Ishida K, Green BR, Cavalier-Smith T: **Diversification of a chimaeric algal group, the chlorarachniophytes: phylogeny of nuclear and nucleomorph small-subunit rRNA genes**. *Mol Bio Evol* 1999, 16: 321-331.

Jasinski M, Ducos E, Martinoia E, Boutry M: **The ATP-binding cassette transporters: structure, function, and gene family comparison between rice and *Arabidopsis***. *Plant Physiol.* 2003, 131:1169-1177.

Johnson CH, Kruft V, Subramanian AR: **Identification of a plastid-specific ribosomal protein in the 30 S subunit of chloroplast ribosomes and isolation of the cDNA clone encoding its cytoplasmic precursor.** *J Biol Chem.* 1990, 265:12790-12795.

Kasai F, Kawachi M, Erata M, Mori F, Yumoto K, Sato M, Ishimoto M: **NIES-Collection List of Straind, 8th Edition**. *Jap J Phycol* 2009, 57(1)Suppl. : 1-350.

Kato S: **Laboratory culture and morphology of *Colacium vesiculosum Ehrb*. (Euglenophyceae).** *Jap J Phycol* 1982, 30: 63-67.

Keeling PJ: **The endosymbiotic origin, diversification and fate of plastids.** *Philos Biol Sci* 2010, 365: 729-748.

Keeling, P.J, Archibald JM: **Organelle evolution: what's in a name?** *Current Biology* 2008, 18: 345–347.

Ishida K, Yabuki A, Ota S: **Research note: *Amorphochlora amoebiformis* gen. et comb. nov. (Chlorarachniophyceae)** *Phycol Res* 2011, 59:52-53.

Lartillot N, Brinkmann H, Philippe H: **Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model.** *BMC Evol*

Biol 2007, 8:7.

Lartillot N, Philippe H: **A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement Process.** *Mol Biol Evol* 2004, 21: 1095-1109.

Lartillot N, Philippe H: **Computing Bayes factors using thermodynamic integration**. *Systematic Biology* 2006, 55:195-207.

Leander BS: **Did trypanosomatid parasites have photosynthetic ancestors?** 2004

Li YF, Zhou DX, Clabault G, Bisanz-Seyer C, Mache R: **Cis-acting elements and expression pattern of the spinach rps22 gene coding for a plastid-specific ribosomal protein**. Plant Mol Biol. 1995, 28:595-604.

Linton EW, Ishikawa AK, Kim JI, Shin W, Bennett MS, Kwiatowski J, Zakrys B, Tremer RE: **Reconstructing euglenoid evolutionary relationships using three genes: nuclear SSU and LSU, and chloroplast SSU rDNA sequences and the description of euglenaria gen. nov. (Euglenophyta)**. *Protist* 2010, 161: 603-619.

L öwe J, Amos LA: **Crystal structure of the bacterial cell-division protein FtsZ.** *Nature* 1998, 391: 203–206.

Martin W, Mustafa AZ, Henze K, Schnarrenberger C: **Higher-plant chloroplast and cytosolic fructose-1, 6-bisphosphatase isoenzymes: origins via duplication rather than prokaryote-eukaryote divergence.** *Plant Mol Biol* 1996, 32: 485-491.

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D: **Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus.** *Proc Natl Acad Sci USA* 2002, 99:12246–12251.

Martin W, Schnarrenberger C : **The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis.** *Curr Genet* 1997, 32: 1-18.

Maruyama S, Suzaki T, Weber AP, Archibald JM, Nozaki H: **Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids.** *BMC Evol Biol* 2011, 11:105.

Matsuzaki M, Kuroiwa H, Kuroiwa T, Kita K, Nozaki, H: **A cryptic algal group unveiled: a plastid biosynthesis pathway in the oyster parasite *Perkinsus marinus*.** *Mol Biol Evol* 2008, 25: 1167-1179.

Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyagishima SY, Mori T, Nishida K, Yagisawa F, Nishida K, Yoshida Y, Nishimura Y, Nakao S, Kobayashi T, Momoyama Y, Higashiyama T, Minoda A, Sano M, Nomoto H, Oishi K, Hayashi H, Ohta F, Nishizaka S, Haga S, Miura S, Morishita T, Kabeya Y, Terasawa K, Suzuki Y, Ishii Y, Asakawa S, Takano H, Ohta N, Kuroiwa H, Tanaka K, Shimizu N, Sugano S, Sato N, Nozaki H, Ogasawara N, Kohara Y, Kuroiwa T: **Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D.** *Nature* 2004, 428: 653-657.

McFadden GI, van Dooren GG: Evolution: **red algal genome affirms a common origin of all plastids.** *Curr Biol* 2004, 14:R514-516.

Meurer J, Plücken H, Kowallik KV, Westhoff P: **A nuclear-encoded protein of prokaryotic origin is essential for the stability ofphotosystem II in *Arabidopsis thaliana*.** *EMBO J* 1998, 17:5286-5297.

Michael Syvanen, Clarence I. Kado: **Horizontal Gene Transfer Academic Press**. 2002, 405.

Minge MA, Shalchian-Tabrizi K, Tørresen OK, Takishita3 K, Probert I, Inagaki Y, Klaveness D, Jakobsen KS: **A phylogenetic mosaic plastid proteome and unusual plastid-targeting signals in the green-colored dinoflagellate *Lepidodinium chlorophorum*.** *BMC Evol Biol* 2010,10:191.

Møller SG, Kunkel T, Chua NH: **A plastidic ABC protein involved in intercompartmental communication of light signaling.** *Genes Dev*. 2001, 15:90-103.

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

Mullner AN, Angeler DG, Samuel R, Linton EW, Triemer RE: **Phylogenetic analysis of phagotrophic phototrophic and osmotrophic euglenoids by using the nuclear 18S rDNA sequence.** IJSEM 2010, 51: 783-791.

Nowack EC, Grossman AR: **Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*.** *Proc Natl Acad Sci USA* 2012, 109:5340-5345.

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

Nozaki H, Iseki M, Hasegawa M, Misawa K, Nakada T, Sasaki N, Watanabe M: **Phylogeny of primary photosynthetic eukaryotes as deduced from slowly evolving nuclear genes**. *Mol Bio Evol* 2007, 24: 1592-1595.

Nozaki H, Ito M, Watanabe MM, Takano H, Kuroiwa T: **Analysis of morphological species of** *Carteria* **(Volvocales, Chlorophyta) based on rbcL gene sequences**. *J Phycol* 1997, 33: 864-867.

Nozaki H, Maruyama S, Matsuzaki M, Nakada T, Kato S, Misawa K: **Phylogenetic positions of Glaucophyta, green plants (Archaeplastida) and Haptophyta (Chromalveolata) as deduced from slowly evolving nuclear genes**. *Mol Phylog Evol* 2009, 53: 872-880.

Okamoto N, Inouye I: **A secondary symbiosis in progress?** *Science* 2005, 310:287.

Ota S, Vaulot D: *Lotharella reticulosa* **sp. nov.: a highly reticulated network forming chlorarachniophyte from the Mediterranean Sea**. *Protist* 2012, 163: 91-104.

Parfrey LW, Grant J, Tekle YI, Lasek-Nesselquist E, Morrison HG, Sogin ML, Patterson DJ, Katz LA: **Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life.** *Syst Biol* 2010, 59: 518–533.

Peltier JB, Emanuelsson O, Kalume DE, Ytterberg J, Friso G, Rudella A, Liberles DA, Söderberg L, Roepstorff P, von Heijne G, van Wijk KJ: **Central functions of the lumenal and peripheral thylakoid proteome of** *Arabidopsis* **determined by experimentation and genome-wide prediction**. *Plant Cell* 2002, 14:211-236.

Petersen J, Teich R, Brinkmann H, Cerff R: **A "green" phosphoribulokinase in complex algae with red plastids: evidence for a single secondary endosymbiosis leading to haptophytes, cryptophytes, heterokonts, and dinoflagellates**. *J Mol Evol* 2006, 62: 143-157.

Pielak GJ, Miklos AC: **Crowding and function reunite**. *Proc Natl Acad Sci USA* 2010, 107: 17457-17458.

Pillet L, Pawlowski J: **Transcriptome analysis of foraminiferan** *Elphidium margaritaceum* **questions the role of gene transfer in kleptoplastidy**. *Mol Biol*

Evol 2013, 30: 66-69.

Reyes-Prieto A, Bhattacharya D: **Phylogeny of Calvin cycle enzymes supports Plantae monophyly**. *Mol Phylogenet Evol* 2007, 45: 384–391.

Reyes-Prieto A,Weber APM, Bhattacharya D: **The Origin and establishment of the plastid in algae and plants**. *Annu Rev of Genet* 2007, 41.1: 147-168.

Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ: **The complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts**. *Mol Biol Evol* 2007, 24: 54–62.

Rumphoa ME, Pochareddya S, Worfula JM, Summerb EJ, Bhattacharyac D, Pelletreaua KN, Tylerd MS, Leee J, Manhartf JR, Soulea KM: **Molecular characterization of the Calvin cycle enzyme phosphoribulokinase in the stramenopile alga *Vaucheria litorea* and the plastid hosting mollusc *Elysia chlorotica***. *Mol. Plant* 2009, 2: 1384-1396.

Shimodaira H, Hasegawa M: **CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics** 2010, 17: 1246-1247.

Shimodaira H: **An approximately unbiased test of phylogenetic tree selection**. *Syst Biol* 2002, 51: 492–508.

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

Strepp R, Scholz S, Kruse S, Speth V, Reski R: **Plant nuclear gene knockout reveals a role in plastid division for the homolog of the bacterial cell division protein FtsZ, an ancestral tubulin**. *Proc Natl Acad Sci USA* 1998, 95: 4368–4373.

Swofford DL: **PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts** 2002.

Tabita FR: **The biochemistry and molecular regulation of carbon dioxide metabolism in cyanobacteria.** In: *Bryant DA (ed) The molecular biology of cyanobacteria. Kluwer Dordrecht* 1994, 437–467.

Takahashi F, Okabe Y, Nakada T, Sekimoto H, Ito M, Kataka H, Nozaki H: **Origins of the secondary plastids of euglenophyta and chlorarachniophyta as revealed by an analysis of the plastid-targeting, nuclear-encoded gene *psbO*. *J***

phycol 2007, 43: 1302-1309.

Tanaka R, Oster U, Kruse E, Rudiger W, Grimm B: **Reduced activity of geranylgeranyl reductase leads to loss of chlorophyll and tocopherol and to partially geranylgeranylated chlorophyll in transgenic tobacco plants expressing antisense RNA for geranylgeranyl reductase**. *Plant Physiol* 1999, 120:695-704.

Teich R, Zauner S, Baurain D, Brinkmann H, Petersen J: **Origin and distribution of Calvin cycle fructose and sedoheptulose bisphosphatases in Plantae and complex algae: a single secondary origin of complex red plastids and subsequent propagation via tertiary endosymbioses**. *Protist* 2007, 158: 263-276.

Thrash JC, Boyd A, Huggett MJ, Grote J, Carini P, Yoder RJ, Robbertse B, Spatafora JW, Rapp éMS, Giovannoni SJ: **Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade**. *Sci Rep* 2011, 1:13.

Timmis JN, Ayliffe MA, Huang CY, Martin W: **Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes**. *Nat Rev Genet* 2004, 5:123–135.

Yang J, Schuster G, Stern DB: **CSP41, a sequence-specific chloroplast mRNA binding protein, is an endoribonuclease**. *Plant Cell*. 1996, 8:1409-1420.

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