



Phylogenomic Analysis of “Red” Genes from Two Divergent Species of the “Green” Secondary Phototrophs, the Chlorarachniophytes, Suggests Multiple Horizontal Gene Transfers from the Red Lineage before the Divergence of Extant Chlorarachniophytes

Yi Yang¹, Motomichi Matsuzaki², Fumio Takahashi^{3,4}, Lei Qu⁵, Hisayoshi Nozaki^{1*}

1 Department of Biological Sciences, Graduate School of Science, University of Tokyo, Bunkyo, Tokyo, Japan, **2** Department of Biomedical Chemistry, Graduate School of Medicine, University of Tokyo, Bunkyo, Tokyo, Japan, **3** College of Life Sciences, Ritsumeikan University, Kusatsu, Shiga, Japan, **4** JST, PRESTO, Kawaguchi, Saitama, Japan, **5** School of Computer Science, Fudan University, Shanghai, P. R. China

Abstract

The plastids of chlorarachniophytes were derived from an ancestral green alga via secondary endosymbiosis. Thus, genes from the “green” lineage via secondary endosymbiotic gene transfer (EGT) are expected in the nuclear genomes of the Chlorarachniophyta. However, several recent studies have revealed the presence of “red” genes in their nuclear genomes. To elucidate the origin of such “red” genes in chlorarachniophyte nuclear genomes, we carried out exhaustive single-gene phylogenetic analyses, including two operational taxonomic units (OTUs) that represent two divergent sister lineages of the Chlorarachniophyta, *Amorphochlora amoebiformis* (= *Lotharella amoebiformis*; based on RNA sequences newly determined here) and *Bigeloviella natans* (based on the published genome sequence). We identified 10 genes of cyanobacterial origin, phylogenetic analysis of which showed the chlorarachniophytes to branch with the red lineage (red algae and/or red algal secondary or tertiary plastid-containing eukaryotes). Of the 10 genes, 7 demonstrated robust monophyly of the two chlorarachniophyte OTUs. Thus, the common ancestor of the extant chlorarachniophytes likely experienced multiple horizontal gene transfers from the red lineage. Because 4 of the 10 genes are obviously photosynthesis- and/or plastid-related, and almost all of the eukaryotic OTUs in the 10 trees possess plastids, such red genes most likely originated directly from photosynthetic eukaryotes. This situation could be explained by a possible cryptic endosymbiosis of a red algal plastid before the secondary endosymbiosis of the green algal plastid, or a long-term feeding on a single (or multiple closely related) red algal plastid-containing eukaryote(s) after the green secondary endosymbiosis.

Citation: Yang Y, Matsuzaki M, Takahashi F, Qu L, Nozaki H (2014) Phylogenomic Analysis of “Red” Genes from Two Divergent Species of the “Green” Secondary Phototrophs, the Chlorarachniophytes, Suggests Multiple Horizontal Gene Transfers from the Red Lineage before the Divergence of Extant Chlorarachniophytes. PLOS ONE 9(6): e101158. doi:10.1371/journal.pone.0101158

Editor: Senjie Lin, University of Connecticut, United States of America

Received: February 22, 2014; **Accepted:** June 3, 2014; **Published:** June 27, 2014

Copyright: © 2014 Yang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by Grants-in-Aid for Challenging Exploratory Research (number 24657045 to HN) and Scientific Research (A) (number 24247042 to HN) from MEXT/JSPS KAKENHI (<http://www.jsps.go.jp/english/e-grants/index.html>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: nozaki@bs.s.u-tokyo.ac.jp

Introduction

Approximately one billion years ago, an ancient cyanobacterium was engulfed by a heterotrophic eukaryote. A primary plastid is assumed to have evolved directly from the uptake of this cyanobacterium [1], [2], [3]. Thus acquiring photosynthetic activity, the eukaryotic host then diversified into the extant members of the supergroup Archaeplastida, which consists of the Chloroplastida (green algae and land plants), Rhodophyceae (red algae), and Glaucophyta [4]. Recent studies suggest that primary endosymbiosis might have also occurred in another lineage: the cercozoan amoeba *Paulinella chromatophora* [5]. The red and green algal ancestors were then involved in secondary endosymbioses, becoming complex secondary plastids. All of the eukaryotes with red algal-derived secondary and tertiary plastids (hereafter, “CASH lineage” [6]; cryptophytes, alveolates, stramenopiles

[Heterokontophyta], and haptophytes) were assigned to the supergroup Chromalveolata in 2005 [4]. However, the number of red secondary endosymbioses has been debated; successive endosymbioses of red algal plastids in the CASH lineage, as a model, is generally growing in popularity [6], [7], [8], [9]. In contrast, green secondary plastids are found in two separate lineages: chlorarachniophytes (supergroup Rhizaria) and euglenophytes (supergroup Excavata) [9]. Recent studies suggest that the green secondary plastids in these lineages may originate from two independent endosymbioses of green algae [10], [11]. Multigene phylogenetic analyses of various eukaryotic lineages has demonstrated that stramenopiles, alveolates, and Rhizaria form a monophyletic group [8], [9], [12], [13], [14] that was very recently reclassified as the supergroup “SAR” (Stramenopiles, Alveolates, and Rhizaria) [15].

During the endosymbiotic process of engulfing, reducing, and integrating plastid ancestors, endosymbiotic gene transfer (EGT) occurred and a bulk of the endosymbiont genes were transferred to the host genomes to remain and function as host nuclear genes; these genes may be relics of past endosymbiotic events that can now be traced back [16], [17]. EGT is a special case of horizontal gene transfer (HGT).

The Chlorarachniophyta is a relatively small algal division inhabiting marine environments. They are rhizarian amoeboflagellates that possess green plastids; their amoeboid single cells often have branching cytoplasmic extensions called reticulopodia. The fact that this algal group has four layers of plastid membranes and a nucleomorph (or endosymbiotically derived eukaryotic nucleus) indicates that the plastids were derived from secondary endosymbiosis of a green algal ancestor [18], [19], [20], [21]. Phylogenetic analyses and pigment composition also point to the green algal origin of the chlorarachniophyte plastids [21], [22], [23], [24]. Thus, genes of green algal ancestry (i.e., genes of the “green lineage”) are expected to reside in the nuclear genomes of the chlorarachniophytes, having been transferred from the nucleomorph to the secondary host nucleus via EGT. However, several recent studies have shown presence of “red” genes of cyanobacterial origin in their nuclear genomes [25], [26], [27], [28].

In 2003 Archibald et al. [25] showed that eight genes of the chlorarachniophyte *Bigelowiella natans* were derived from red algae or red secondary plastid-bearing algae. More recently, Curtis et al. [28] identified 45 red algal-type genes in the nuclear genome sequence of *B. natans*. However, Curtis et al. [28] discussed that their results have to be treated with caution, and that the testing of hypotheses of possible biological explanations for the diversity of algal nuclear genes seen in *B. natans*, such as the relative contributions of EGT versus HGT, cannot currently be carried out without careful consideration of taxon sampling and methodological artifacts. The precise origins of these red genes were not resolved, in part because their datasets typically included only one operational taxonomic unit (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 we 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 [32]. We obtained transcriptome data from *A. amoebiformis* by next-generation sequencing and combined it with the *B. natans* nuclear genome data. To extract more “red” genes from the Chlorarachniophyta, we established another original pipeline and manually checked as many positive outputs as possible. Based on this pipeline and the red genes extracted by Curtis et al. [28], a total of 10 “red” genes of cyanobacterial origin were found from the chlorarachniophyte lineage.

Materials and Methods

Strain and culture conditions

Amorphochlora amoebiformis CCMP2058 (designated as *Lotharella amoebiformis*) was obtained from NCMA (Provasoli-Guillard National Center for Marine Algae and Microbiota; <https://ncma.bigelow.org/>) and cultured in L1 medium [33] 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°C with a 14-h light: 10-h dark cycle. For RNA extraction, 4 L of culture, grown in two flasks containing 2 L each, was cultivated for a period of 57 days.

RNA extraction

Cells of *A. amoebiformis* CCMP2058 were ruptured manually using quartzzone sand in liquid nitrogen for 10 min, and RNA was subsequently extracted using the SV Total RNA Isolation System (Promega, Madison, WI, USA). The quantity of total RNA was measured with a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA), until the quantity reached 300 µg with a concentration of 6 µg/µl. The extracted total RNA was then sent to Takara Bio Inc. (Otsu, Shiga, Japan) for further processing, including poly(A) purification and GS FLX+ (Roche Applied Science, Mannheim, Germany) 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 Inc., containing all 197,073 single reads, were assembled using Trinity (<http://trinityrnaseq.sourceforge.net/>) [34] on a 2 × quad-core Xeon E5-2650 (2.00 GHz, Sandy Bridge-EP) platform (Intel Corporation, Santa Clara, CA, USA). 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 (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). The BLASTP was carried out against the National Center for Biotechnology Information database (NCBI; <http://www.ncbi.nlm.nih.gov/>) and local databases that were retrieved from NCBI (expressed sequence tag [EST] data), the DOE Joint Genome Institute (JGI; <http://www.jgi.doe.gov/>), several unpublished datasets, and our *A. amoebiformis* sequences that were prepared as described above (Table S1). Multiple sequence alignments were generated using Muscle (v3.7 by Robert C. Edgar, <http://www.drive5.com/muscle/>) [35], [36]. We limited the local databases used in the first round (group A in Table S2), as several EST databases resorted to stretching the length of alignment gaps when the quality of their sequences was low. Meanwhile, we used a trimming script to exclude mainly 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 preserved 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) [37] phylogenetic analyses were carried out with the WAG+Γ4 model as a fast filter (which ignored bootstrap values) 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 BLAST database (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 repeated with 1000 replications of bootstrap analysis. Analyses based on PhyloBayes 3.3 (<http://www.atgc-montpellier.fr/phylobayes/>) [38], [39], [40] were also carried out with the WAG+ Γ 4 model, and the “good run” option was used.

Almost all of the alveolate OTUs were automatically removed during our gene-mining process via BLAST due to their divergent or long-branched sequences. However, alveolates belong to the SAR supergroup, along with chlorarachniophytes and stramenopiles [6], [12], [13], [14], [15]. 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 [41] 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, original local databases were constructed for the first- and second-round filters and they were added to the NCBI database (see above). Curtis et al.’s supplementary data [28] have been compared and added to evaluate and cross check our results.

In addition, we carried out approximate unbiased tests (AU test) [29] to examine the phylogenetic position of the chlorarachniophyte OTUs, except for one tree in which two chlorarachniophyte OTUs showed separate phylogenetic positions (see Results). We used nine series of the phylogenetic trees by RaxML (without alveolate OTUs), where topologies of all the OTUs, excluding chlorarachniophytes, were fixed. The alignment was used as input data. All possible topologies were generated by re-grafting the branch of chlorarachniophytes, using the in-house ruby script. The pools of topologies were analyzed with the AU test, using the site-wise log-likelihood values calculated with PhyML (ver. 3.0 [30]) (with the WAG model+F+I+ Γ 4). The AU test was conducted using Consel (ver. 0.2 [31]).

Results

Gene mining

To elucidate the contribution of genes of the red lineage to genome mosaicism in Chlorarachniophyta, we searched the proteome of *B. natans* for proteins showing red algal affiliations. First, we used the 21,708 predicted proteins in *B. natans* available from NCBI as queries to conduct a thorough phylogenomic search. Out of the 21,708 queries, 3,436 proteins had more than 10 OTUs with which to construct phylogenetic 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 stramenopiles/alveolates, red algae, or Chloroplastida. Furthermore, only a small portion of these (259 candidates) had more than two cyanobacterial OTUs in their phylogenetic trees, which is the minimum requirement for a cyanobacterial origin in eukaryote OTUs. Approximately half (129) of the 259 eukaryotic genes of possible cyanobacterial origin showed affinities of the chlorarachniophyte OTUs with Chloroplastida. The remainder might include genes that show a “red lineage” affinity for the chlorarachniophyte homolog. After further refinement of our sampling sequence pool and the sequence pool from the supplementary data of Curtis et al. [27], we selected 6 and 7 cyanobacteria-type hits, respectively, which ultimately resulted in a total of 10 genes that were likely representative of almost all eukaryote OTUs originating from primary plastid EGT and chlorarachniophytes positioned in the red lineage (within, or sister to, the red algae and/or CASH lineages) (Figure S1, File S1).

These 10 cyanobacterial gene trees (without alveolate OTUs) were supported by bootstrap values (BV) $\geq 75\%$ and posterior probability (PP) ≥ 0.95 for the affiliation of the chlorarachniophytes with the red lineage (Figures 1–3, Figures S2–S8 in File S2, Table S2). Our in-house ruby scripts for gene mining and the amino acid alignments and tree topologies of the 10 genes are available from GitHub (<https://github.com/djmyabbay>).

Among the 10 genes, genes encoding an ATP binding cassette transporter (ABC), an mRNA binding protein (RNABP), and a geranylgeranyl reductase (GGR), were identified by both pipelines (the present study and Curtis et al. [28]). Three other queries, including one putative membrane protein (PMP), one hypothetical protein (HP), and a phosphoribulokinase (PRK), were found only by our mining pipeline (Figure S1).

Single-gene phylogenetic analyses

Seven of the ten trees showed robust monophyly of *B. natans* and *A. amoebiformis* (representing the two divergent sister lineages of the Chlorarachniophyta) (with BV $\geq 97\%$ and PP ≥ 0.99). The GGR tree exhibited moderate monophyly of the chlorarachniophytes (with 71–73% BV and 0.99 PP), whereas in the other two trees, one showed separation between *B. natans* and *A. amoebiformis* and the other lacked an *A. amoebiformis* sequence (Table S2). Four of the eight trees with chlorarachniophyte monophyly showed affinity of the chlorarachniophytes with the CASH lineage (CASH type), although the support values were weak to moderate (with 54–76% BV and 0.96–0.99 PP), even in analyses without alveolate OTUs. Chlorarachniophyte PRK genes indicated origins directly from a red algal ancestor, as reported previously [27] (Red type). In the remaining three trees showing chlorarachniophyte monophyly, the phylogenetic position of the chlorarachniophytes within the red lineage was ambiguous (Ambiguous type).

CASH-type trees. Four genes of CASH lineage encoded plastid-targeted proteins that were directly or indirectly related to photosynthesis or plastid functions: GGR, RNABP, plastid division protein FtsZ (PDP, filamenting temperature-sensitive mutant Z), and photosystem II stability assembly (PS2SAF). FtsZ is a prokaryotic homologue of the eukaryotic protein tubulin and can be considered a functional housekeeping gene of plastids (plastid division) [42], [43], [44]. Figure 1 shows the robust monophyly of the *B. natans* and *A. amoebiformis* FtsZ proteins (with 99% BV and 0.99 PP). Chlorarachniophytes, stramenopiles, and the haptophyte *Emiliania* formed a clade (with 54% BV and 0.99 PP), to which the clade composed of three species of cryptophytes (red algal nucleomorphs) and the red alga *Cyanidioschyzon* was basal. In contrast, 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 it has been suggested to be required for chlorophyll *a* binding [45], [46]. Phylogenetic analysis of PS2SAF revealed a robust chlorarachniophyte clade (with 100% BV and 1.00 PP) (Figure 2). As with 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 two genes, GGR and RNABP, showed affinity of the chlorarachniophytes with the CASH lineage (with BV $\geq 60\%$ and PP ≥ 0.96), the addition of alveolate OTUs lowered the branch support (below 50% BV and 0.95 PP) in the RNABP tree (Figures S3, S5 in File S2).

Red-type trees. In the tree of PRK sequences, the chlorarachniophytes, including *A. amoebiformis* sequences, formed a robust clade that was closely related to the red algae [27] (Figure S8 in

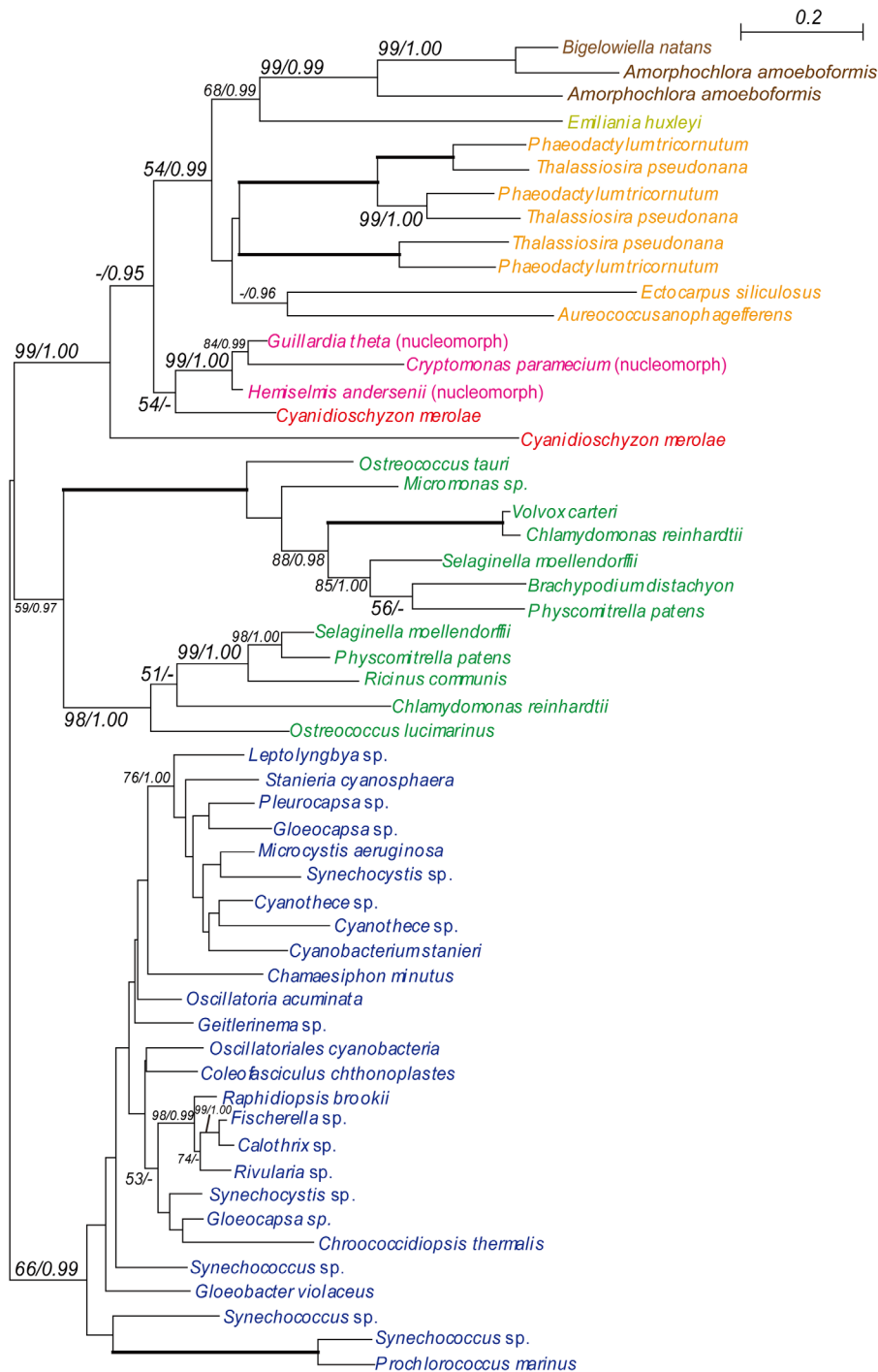


Figure 1. Phylogeny of PDP (FtsZ) showing chlorarachniophyte proteins closely related to red algal plastid-containing eukaryote homologues. The tree was inferred using the RaxML method with the WAG+I-gamma model. Numbers at branches represent support values (bootstrap values $\geq 50\%$ or posterior probability ≥ 0.95) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values of 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. doi:10.1371/journal.pone.0101158.g001

File S2). As in our previous study [27], this tree topology may indicate that the PRK genes were transferred directly from a red algal ancestor to the common ancestor of the extant chlorarachniophytes. However, OTUs from secondary/tertiary eukaryotes with red algal plastids (CASH lineage) were separated from the red algae. This suggests that the PRK genes of CASH lineage

might have experienced a gene replacement after the typical secondary/tertiary EGT from the red algal plastid-containing eukaryotes [9]. A similar separation between the CASH lineage and the lineage composed of red algae and chlorarachniophytes was present in the plastid-targeted sedoheptulose-bisphosphatase (SBP) tree [27] (Figure S9 in File S2).

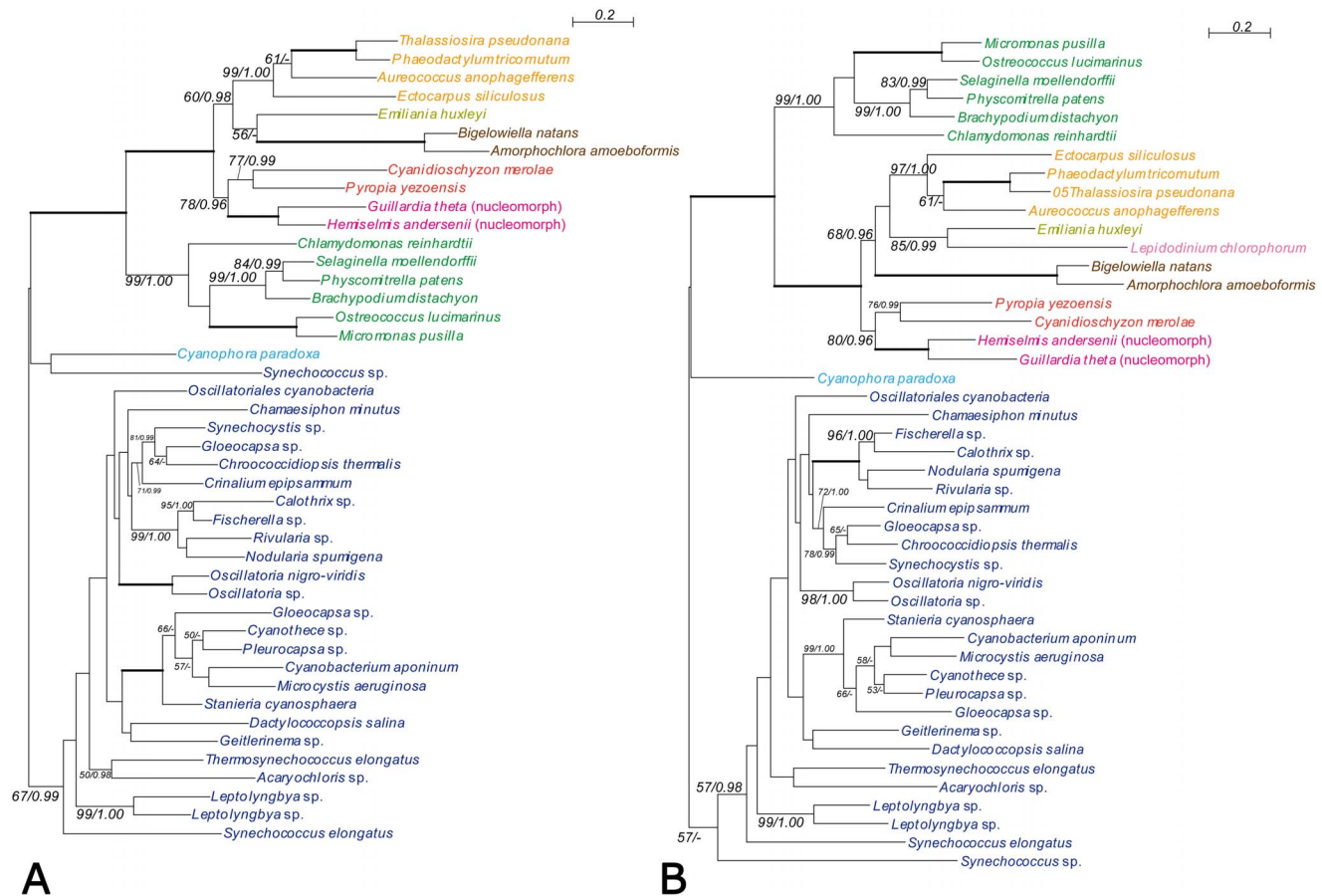


Figure 2. Phylogeny of PS2SAF showing chlorarachniophyte proteins closely related to algal plastid-containing eukaryote homologues. The trees were inferred using the RaxML method with the WAG+Gamma model. Numbers at branches represent support values (bootstrap values $\geq 50\%$ or posterior probability ≥ 0.95) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values of 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. (A) Lacking alveolate OTUs. (B) Containing alveolate OTUs.

doi:10.1371/journal.pone.0101158.g002

Ambiguous-type trees. Due to the limited number of available sequences, the phylogenetic position of the chlorarachniophyte clade within 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 [47], [48], [49]. Chlorarachniophyte PGK gene sequences were robustly monophyletic (with 97–99% BV and 0.99 PP). Chlorarachniophytes, red algae, and red algal secondary/tertiary algae formed a clade with 79% BV and 1.00 PP (Figure 3A). However, for all three genes, no statistical support was obtained regarding the phylogenetic position of the chlorarachniophytes within the red lineage (Figure 3, Figures S4, S7 in File S2).

Other trees. Two protein trees did not demonstrate monophyly of the chlorarachniophytes. The ABC protein contained only a single chlorarachniophyte OTU (*B. natans*), and the putative membrane protein (PMP) demonstrated a phylogenetic split between *B. natans* and *A. amoebiformis*. Gene duplications or replacements may have resulted in the split in chlorarachniophyte PMP genes portrayed by the three separate lineages of diatoms (each including a *Thalassiosira* OTU) (Figures S2, S6 in File S2).

AU tests

Based on our AU tests of nine trees (Figures S10–S18 in File S3), only RPS22 rejected the possibility that the chlorarachniophyte lineage is positioned within the Chloroplastida (excluding the land plants) with $p \leq 0.05$ (Figure S15 in File S3). However, the possibility that the chlorarachniophyte lineage is positioned in the basal lineage of the UTC group (origin of the secondary plastids of the chlorarachniophytes [9], [10]; including *Chlamydomonas*, *Volvox*, and *Chlorella*) was rejected by PS2SAF, RNABP, and PRK. For the other trees, our tests did not reject the possibilities, possibly due to the limited information of the single-gene trees.

Discussion

Based on the extensive analysis of single-gene trees using OTUs from two evolutionarily distinct chlorarachniophytes, we identified 10 genes of cyanobacterial origin that supported an affiliation of chlorarachniophytes with the red algae and/or CASH lineage. For four of the nine trees, our AU test rejected the possibility that the chlorarachniophyte genes originated from secondary endosymbiosis of the extant chlorarachniophyte green plastids (Figures S10–S18 in File S3). Therefore, although these genes are of cyanobacterial origin, they do not likely originate from the green

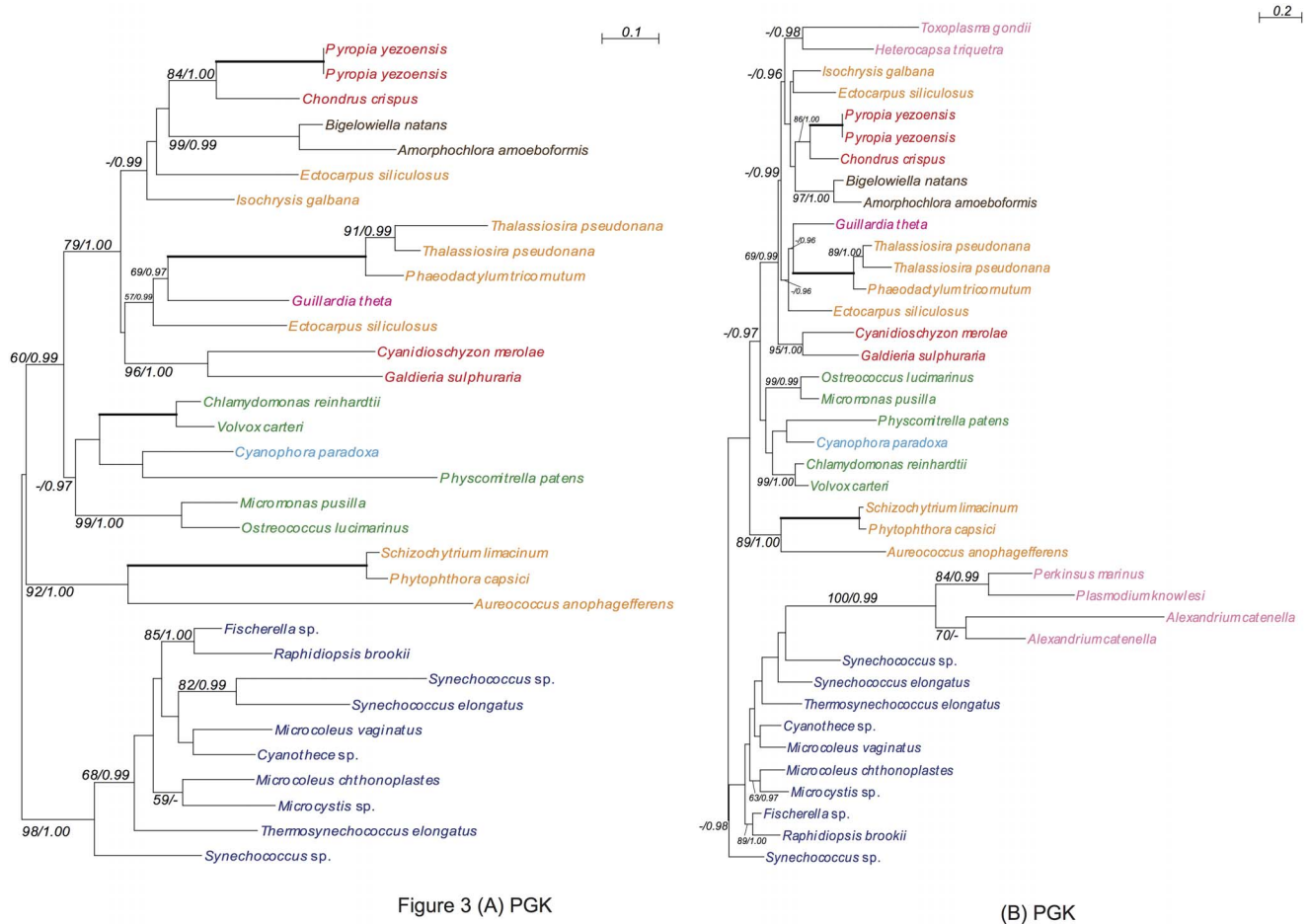


Figure 3. Phylogenies of PGK showing chlorarachniophyte proteins closely related to red algal plastid-containing eukaryote homologues. The trees were inferred using the RaxML method with the WAG+I+gamma model. Numbers at branches represent support values (bootstrap values $\geq 50\%$ or posterior probability ≥ 0.95) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values of 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlorarachniophyta. (A) Lacking alveolate OTUs. (B) Containing alveolate OTUs.
doi:10.1371/journal.pone.0101158.g003

algal ancestor of the secondary plastids in the extant chlorarachniophytes; they most likely originate from the red lineage. Of the 10 single-gene trees, seven demonstrated robust monophyly of the two chlorarachniophyte OTUs, and the GGR tree showed moderate monophyly of the chlorarachniophytes (Table S2). Because the two chlorarachniophyte OTUs represent two sister (basally divergent) lineages of the Chlorarachniophyta [32], multiple HGTs from the red lineage likely occurred in the common ancestor of the extant chlorarachniophytes.

Four of the nine gene trees (CASH type trees; GGR, RNABP, PDP and PS2SAF) showed weak to moderate statistical support for the affiliation of the chlorarachniophyte homologues with those of the CASH lineage (Figures 1, 2, Figures S3, S5 in File S2), whereas three other gene trees were ambiguous (Ambiguous type; RPS22, HP and PGK). Although the chlorarachniophyte PRK genes apparently originated directly from a red algal ancestor (Figure S8 in File S2) as reported previously [27], the genes from the CASH lineage did not belong to the red algal lineage. Therefore, PRK genes of the CASH lineage may have experienced gene replacement after the typical secondary EGT scenario from the red algal ancestor. Thus, the apparent affiliation of chlorarachniophyte PRK homologues with those from red algae

(Figure S8 in File S2) [27] may be the result of the removal of CASH-lineage PRK genes from the red lineage by such a gene replacement. A similar situation is possible with the red algal and chlorarachniophyte SBP homologues (Figure S9 in File S2) [27]. Alternatively, the ancestor of the chlorarachniophyte PRK and/or SBP genes may have experienced a chance HGT directly from the red algae lineage. Therefore, although the statistical support for affiliation between chlorarachniophyte OTUs and the CASH lineage was not robust in the four CASH type trees, the CASH lineage may be the most probable origin of the red genes that are believed to have invaded the nuclear genome of the common ancestor of extant Chlorarachniophyta via HGT.

Among the 10 genes, 4 (PDP, GGR, PS2SAF and PRK) were obviously photosynthesis- and/or plastid-related [27], [41], [42], [43], [44], [45], [46], [50]. Furthermore, all eukaryotic OTUs in the ten trees possessed plastids, except for several plastid-lacking stramenopile and excavate taxa in the PGK and RNABP trees (Figure 3, Figure S5 in File S2). Thus, those 10 genes most likely originated directly from photosynthetic eukaryotes. As discussed above, the multiple HGTs likely took place in the common ancestor of the extant chlorarachniophytes, most likely from a red algal plastid-containing eukaryote. Multiple HGTs directly from

photosynthetic eukaryotes may be explained by three alternative hypotheses: 1) A cryptic endosymbiosis of a red algal plastid-containing eukaryote in the common ancestor of the extant chlorarachniophytes, 2) Multiple HGTs from a long-term feeding on a single (or multiple closely related) red algal plastid-containing eukaryote(s), or 3) Vertical gene transmission following HGTs from the red alga to the common ancestor of SAR.

Under the first hypothesis, the ancestor of the chlorarachniophytes might have harbored a red-algal plastid prior to the secondary endosymbiosis that gave rise to the green algal plastid that currently exists. Subsequently, the secondary endosymbiosis of the green alga might have resulted in the discarding of such pre-existing red-algal plastids by the ancestor of the chlorarachniophytes. Thus, the origin of the green secondary plastids of extant chlorarachniophytes would represent a replacement of the pre-existing red algal-plastid. This scenario considers that the pre-existing red algal-plastid might have originated from secondary, tertiary, or quaternary endosymbiosis of the cryptic red plastid from the ancestral red lineage [6]. During the proposed cryptic endosymbiosis, EGT could have occurred, resulting in nuclear-encoded, plastid-targeted genes of red lineage. Then, the majority of such red lineage genes could have been replaced by green genes via the secondary EGT of the green plastid of the extant chlorarachniophytes. However, the rest of the red lineage genes in the host nucleus might have been retained, not affected by the green EGT gene replacement, and we now identify them as red lineage genes in the chlorarachniophyte nuclear genome.

The second hypothesis considers that the multiple HGTs from the red lineage might have occurred after the common ancestor of the extant chlorarachniophytes established its green secondary plastid. The ancestor with a green secondary plastid could have experienced a long period in which it fed on a single (or multiple closely related) red algal plastid-containing eukaryote(s). During the long-term feeding, the ancestors of the red genes resolved in the chlorarachniophyte nuclear genome (Table S2) could have replaced pre-existing green genes via multiple HGTs. This evolutionary scenario requires an explanation as to why the HGTs from the red lineage are concentrated in the common ancestors of the chlorarachniophytes. Probably, the secondary endosymbiosis that gave rise to green plastid was based on active feeding on the eukaryotic cells of the common ancestor of the extant chlorarachniophytes, and such phagotrophy could have been active and prominent immediately after the secondary endosymbiosis. Recently, a very interesting, eukaryote-eating euglenophyte was discovered [51]. This new genus, *Rapaza*, has typical green secondary plastids, but actively engulfs a prasino-phyte green alga with its feeding apparatus. Since *Rapaza* is positioned most basally within the euglenophyte lineage [51], it may exhibit a strong phagotrophic activity retained from the ancestral stage of the euglenophytes. However, no such a basal chlorarachniophyte has yet been discovered [32].

The third hypothesis is based on the sister relationship between Rhizaria and stramenopiles-alveolates (SA) that harbor the red secondary or tertiary plastids [15], and on the assumption that the common ancestor of SAR might have harbored secondary or tertiary red plastids. Based on the third hypothesis, the red genes found in the chlorarachniophytes and the red secondary or tertiary plastids in the extant SA might have been vertically transmitted from the common ancestor of SAR. This scenario may be supported by the present suggestion that the CASH lineage (including SA) might be the most probable origin of the red genes in the nuclear genome of the Chlorarachniophyta (see above). Because the red genes of chlorarachniophyte nuclear genomes identified in the present study (Figures 1–3, Figures S2–S8 in

File S2, Table S2) most likely originated directly from photosynthetic eukaryotes (see above), the common ancestor of the extant chlorarachniophytes might have retained the red plastid vertically transmitted from the common ancestor of SAR. The red plastid in the common ancestor of the chlorarachniophytes might have been then replaced by the green plastid of the extant Chlorarachniophyta. The red genes found in the chlorarachniophyte nuclear genomes might be relics of host genomic contents after the secondary endosymbiosis of the green plastid. However, provided extant lineages in Rhizaria lack plastids except for *Paulinella chromatophora* [5] and chlorarachniophytes that are positioned distally within the rhizarian lineage [52], there must have been multiple losses of red plastids during evolution from the common ancestor of SAR in order to explain the presence of various plastid-lacking basal lineages and the derived photosynthetic lineage Chlorarachniophyta within Rhizaria. Thus, parsimony principles seem to disagree with the multiple losses of red plastids, and vertical gene transmission following HGTs from the red alga to the common ancestor of SAR may not likely explain the presence of “red” nuclear genes in chlorarachniophytes.

Conclusion

The ancestor of extant chlorarachniophytes most likely experienced multiple HGTs from the red lineage prior to or soon after the secondary endosymbiosis that gave rise to the green plastid of the chlorarachniophytes. A recent study of the plastid-possessing euglenid *Euglena gracilis* and plastid-lacking euglenid *Peranema trichophorum* suggested multiple HGTs from the red lineage in the common ancestor of the euglenids [53]. Likewise, multiple gene transfers from the green lineage were recently suggested in the ancestor of the CASH lineage such as diatoms with red algal plastids [54], [55]. Thus, the actual evolutionary histories of both of the green and red plastid-containing secondary/tertiary phototropic lineages are likely more complex than previously thought.

Supporting Information

Figure S1 Red-derived genes of cyanobacterial origin resolved by two pipelines.
(PDF)

Table S1 Sequence data used in construction of local-databases for BLASTP in this study. For groups A and B, see Materials and Methods in the main text.
(XLS)

Table S2 List of the 10 chlorarachniophyte red-derived genes of cyanobacterial origin resolved in this study.
(XLS)

File S1 Amino acid alignments of ten genes extracted based on the present gene mining and SBP (Figures 1–3, Figures S2–S9 in File S2).
(PDF)

File S2 Phylogenetic trees of ABC, GGR, RPS22, RNABP, PMP, HP, PRK, and SBP sequences (Figures S2–S9). The trees were inferred using the RaxML method with the WAG+H+gamma model. Numbers at branches represent support values (bootstrap values $\geq 50\%$ or posterior probability ≥ 0.95) from RaxML/PhyloBayes. Thick branches represent RaxML and PhyloBayes support values of 100% and 1.00, respectively. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; light pink-Alveolata; orange-

stramenopiles; brown-Chlorarachniophyta; purple-Euglenophyta; black-Fungi; violet-Kinetoplastida. (A) Lacking alveolate OTUs. (B) Containing alveolate OTUs. (PDF)

File S3 Results of the AU tests for assessing placement of chlorarachniophyte lineage in nine trees: PDP, PS2SAF, PGK, ABC, GGR, RPS22, RNABP, HP, and PRK (Figures S10–S18). The trees were inferred using the RaxML method with the WAG+I+gamma model. Branches that were rejected by AU test ($p \leq 0.05$) for placement of chlorarachniophyte lineage are indicated with dashed lines. Colors of taxa: dark blue-Cyanobacteria; navy blue-Glaucophyta; green-Chloroplastida; red-Rhodophyceae; pink-Cryptophyta; yellow-Haptophyta; baby pink-Alveolata; orange-stramenopiles; brown-Chlor-

arachniophyta; purple-Euglenophyta; black-Fungi; violet-Kinetoplastida. (PDF)

Acknowledgments

We thank Dr. John D. Archibald for critical reading and correction of the early version of the manuscript and Dr. Shinichiro Maruyama for his kind discussion on the pipeline. The super-computing resource was provided by Human Genome Center, Institute of Medical Science, University of Tokyo.

Author Contributions

Conceived and designed the experiments: YY HN. Performed the experiments: YY FT. Analyzed the data: YY MM LQ. Contributed reagents/materials/analysis tools: YY MM LQ. Wrote the paper: YY HN.

References

- Reyes-Prieto A, Weber AP, Bhattacharya D (2007) The origin and establishment of the plastid in algae and plants. *Annu Rev Genet* 41: 147–168.
- Gould SB, Waller RF, McFadden GI (2008) Plastid evolution. *Annu Rev Plant Biol* 59: 491–517.
- Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D (2004) A molecular timeline for the origin of photosynthetic eukaryotes. *Mol Biol Evol* 21: 809–818.
- Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, et al. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol* 52: 399–451.
- Mackiewicz P, Bodyl A, Gagat P (2012) Protein import into the photosynthetic organelles of *Paulinella chromatophora* and its implications for primary plastid endosymbiosis. *Symbiosis* 58: 99–107.
- Baurain D, Brinkmann H, Petersen J, Rodriguez-Ezpeleta N, Stechmann A, et al. (2012) Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol Biol Evol* 27: 1698–1709.
- Felsner G, Sommer MS, Gruenheit N, Hempel F, Moog D, et al. (2011) ERAD components in organisms with complex red plastids suggest recruitment of a preexisting protein transport pathway for the periplastid membrane. *Genome Biol Evol* 3: 140–150.
- Burki F, Okamoto N, Pombert JF, Keeling PJ (2012) The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc Biol Sci* 279: 2246–2254.
- Archibald JM (2012) The evolution of algae by secondary and tertiary endosymbiosis. Genomic insights into the biology of algae. In: Piganeau G, editor. *Genomic Insight into the Biology of Algae*. 1st Edition (Advances in Botanical Research). Waltham: Academic Press. 87–118.
- Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ (2007) The complete chloroplast genome of the chlorarachniophyte *Bigeloviella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Mol Biol Evol* 24: 54–62.
- Takahashi F, Okabe Y, Nakada T, Sekimoto H, Ito M, et al. (2007) Origins of the secondary plastids of Euglenophyta and Chlorarachniophyta as revealed by an analysis of the plastid-targeting, nuclear-encoded gene *psbO*. *J Phycol* 43: 1302–1309.
- Burki F, Shalchian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI, et al. (2007) Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE* 2: e790.
- Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, et al. (2009) Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic “supergroups”. *Proc Natl Acad Sci USA* 106: 3859–3864.
- Nozaki H, Yang Y, Maruyama S, Suzuki T (2012) A case study for effects of operational taxonomic units from intracellular endoparasites and ciliates on the eukaryotic phylogeny: phylogenetic position of the Haptophyta in analyses of multiple slowly evolving genes. *PLoS ONE* 7: e50827.
- Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, et al. (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol* 59: 429–493.
- Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, et al. (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci USA* 99: 12246–12251.
- Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* 5: 123–135.
- Hibberd DJ, Norris RE (1984) Cytology and ultrastructure of *Chlorarachnion reptans* (Chlorarachniophyta divisio nova, Chlorarachniophyceae classis nova). *J Phycol* 20: 310–330.
- Ludwig M, Gibbs SP (1989) Evidence that the nucleomorphs of *Chlorarachnion reptans* (Chlorarachniophyceae) are vestigial nuclei: morphology, division and DNA-DAPI fluorescence. *J. Phycol* 25: 385–394.
- McFadden GI, Gilson PR, Hofmann CJ, Adcock GJ, Maier UG (1994) Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga. *Proc Natl Acad Sci U S A* 91: 3690–3694.
- McFadden GI, Gilson PR, Hofmann CJ (1997) Division Chlorarachniophyta. In: Bhattacharya D, editor. *Origins of Algae and Their Plastids*. Springer-Verlag Wien New York. 175–185.
- Van de Peer Y, Rensing SA, Maier UG, De Wachter R (1996) Substitution rate calibration of small subunit ribosomal RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. *Proc Natl Acad Sci U S A* 93: 7732–7736.
- Ishida K, Cao Y, Hasegawa M, Okada N, Hara Y (1997) The origin of chlorarachniophyte plastids, as inferred from phylogenetic comparisons of amino acid sequences of EF-Tu. *J Mol Evol* 45: 682–687.
- Sasa T, Takaichi S, Hatakeyama N, Watanabe MM (1992) A novel carotenoid ester, linoxanthin dodecenoate, from *Pyramimonas parkae* (Prasinophyceae) and a chlorarachniophyte alga. *Plant Cell Physiol* 33: 921–925.
- Arichald JM, Rogers MB, Toop M, Ishida K, Keeling PJ (2003) Lateral gene transfer and the evolution of plastid-targeted proteins in the second plastid-containing alga *Bigeloviella natans*. *PNAS* 100: 7678–7683.
- Petersen J, Teich R, Brinkmann H, Cerff R (2006) 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* 62: 143–157.
- Yang Y, Maruyama S, Sekimoto H, Sakayama H, Nozaki H (2011) An extended phylogenetic analysis reveals ancient origin of “non-green” phosphoribulokinase genes from two lineages of “green” secondary photosynthetic eukaryotes: Euglenophyta and Chlorarachniophyta. *BMC Res Notes* 4: 330.
- Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, et al. (2012) Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* 492: 59–65.
- Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51: 492–508.
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696–704.
- Shimodaira H, Hasegawa M (2010) CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246–1247.
- Ota S, Vaulot D (2012) *Lotharella reticulosa* sp. nov.: a highly reticulated network forming chlorarachniophyte from the Mediterranean Sea. *Protist* 163: 91–104.
- Hallegraeff GM, Anderson DM, Cembella AD, Enevoldsen HO (eds) (2003) *Manual on Harmful Marine Microalgae*. 2nd edn. Paris: the United Nations Educational, Scientific and Cultural Organization.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology* 29: 644–652.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797.
- Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
- Stamatakis A (2006) RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.
- Lartillot N, Philippe H (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* 21: 1095–1109.
- Lartillot N, Philippe H (2006) Computing Bayes factors using thermodynamic integration. *Systematic Biology* 55: 195–207.
- Lartillot N, Brinkmann H, Philippe H (2007) Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol* 8: 7.
- Deschamps P, Moreira D (2012) Reevaluating the green contribution to diatom genomes. *Genome Biol Evol* 4: 795–800.

42. Löwe J, Amos LA (1998) Crystal structure of the bacterial cell-division protein FtsZ. *Nature* 391: 203–206.
43. Bi EF, Lutkenhaus J (1991) FtsZ ring structure associated with division in *Escherichia coli*. *Nature* 354: 161–164.
44. Strepp R, Scholz S, Kruse S, Speth V, Reski R (1998) 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* 95: 4368–4373.
45. Peltier JB, Emanuelsson O, Kalume DE, Ytterberg J, Friso G, et al. (2002) Central functions of the lumenal and peripheral thylakoid proteome of *Arabidopsis* determined by experimentation and genome-wide prediction. *Plant Cell* 14: 211–236.
46. Meurer J, Plücker H, Kowallik KV, Westhoff P (1998) A nuclear-encoded protein of prokaryotic origin is essential for the stability of photosystem II in *Arabidopsis thaliana*. *EMBO J* 17: 5286–5297.
47. Dhar A, Samiotakis A, Ebbinghaus S, Nienhaus L, Homouz D, et al. (2010) Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. *Proc Natl Acad Sci USA* 107: 17586–17591.
48. Blake C (1997) Phosphotransfer hinges in PGK. *Nature* 385: 204–205.
49. Pielak GJ, Miklos AC (2010) Crowding and function reunite. *Proc Natl Acad Sci USA* 107: 17457–17458.
50. Tanaka R, Oster U, Kruse E, Rudiger W, Grimm B (1999) 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* 120: 695–704.
51. Yamaguchi A, Yubuki N, Leander BS (2012) Morphostasis in a novel eukaryote illuminates the evolutionary transition from phagotrophy to phototrophy: description of *Rapaza viridis* n. gen. et sp. (Euglenozoa, Euglenida). *BMC Evol Biol* 12: 29.
52. Parfrey LW, Grant J, Tekle YI, Lasek-Nesselquist E, Morrison HG, et al. (2010) Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. *Syst Biol* 59: 518–533.
53. Maruyama S, Suzaki T, Weber AP, Archibald JM, Nozaki H (2011) Eukaryote-to-eukaryote gene transfer gives rise to genome mosaicism in euglenids. *BMC Evol Biol* 11: 105.
54. Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, et al. (2009) Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 324: 1724–1726.
55. Nozaki H, Maruyama S, Matsuzaki M, Nakada T, Kato S, et al. (2009) Phylogenetic positions of Glaucophyta, green plants (Archaeplastida) and Haptophyta (Chromalveolata) as deduced from slowly evolving nuclear genes. *Mol Phylog Evol* 53: 872–880.