

RESEARCH ARTICLE

Open Access

# Extensive horizontal gene transfer, duplication, and loss of chlorophyll synthesis genes in the algae

Heather M Hunsperger, Tejinder Randhawa and Rose Ann Cattolico\*

## Abstract

**Background:** Two non-homologous, isofunctional enzymes catalyze the penultimate step of chlorophyll *a* synthesis in oxygenic photosynthetic organisms such as cyanobacteria, eukaryotic algae and land plants: the light-independent (LIPOR) and light-dependent (POR) protochlorophyllide oxidoreductases. Whereas the distribution of these enzymes in cyanobacteria and land plants is well understood, the presence, loss, duplication, and replacement of these genes have not been surveyed in the polyphyletic and remarkably diverse eukaryotic algal lineages.

**Results:** A phylogenetic reconstruction of the history of the POR enzyme (encoded by the *por* gene in nuclei) in eukaryotic algae reveals replacement and supplementation of ancestral *por* genes in several taxa with horizontally transferred *por* genes from other eukaryotic algae. For example, stramenopiles and haptophytes share *por* gene duplicates of prasinophytic origin, although their plastid ancestry predicts a rhodophytic *por* signal. Phylogenetically, stramenopile *pors* appear ancestral to those found in haptophytes, suggesting transfer from stramenopiles to haptophytes by either horizontal or endosymbiotic gene transfer. In dinoflagellates whose plastids have been replaced by those of a haptophyte or diatom, the ancestral *por* genes seem to have been lost whereas those of the new symbiotic partner are present. Furthermore, many chlorarachniophytes and peridinin-containing dinoflagellates possess *por* gene duplicates. In contrast to the retention, gain, and frequent duplication of algal *por* genes, the LIPOR gene complement (chloroplast-encoded *chlL*, *chlN*, and *chlB* genes) is often absent. LIPOR genes have been lost from haptophytes and potentially from the euglenid and chlorarachniophyte lineages. Within the chlorophytes, rhodophytes, cryptophytes, heterokonts, and chromerids, some taxa possess both POR and LIPOR genes while others lack LIPOR. The gradual process of LIPOR gene loss is evidenced in taxa possessing pseudogenes or partial LIPOR gene complements. No horizontal transfer of LIPOR genes was detected.

**Conclusions:** We document a pattern of *por* gene acquisition and expansion as well as loss of LIPOR genes from many algal taxa, paralleling the presence of multiple *por* genes and lack of LIPOR genes in the angiosperms. These studies present an opportunity to compare the regulation and function of *por* gene families that have been acquired and expanded in patterns unique to each of various algal taxa.

**Keywords:** Chlorophyll synthesis, Horizontal gene transfer, Endosymbiotic gene transfer, Gene duplication, Algae, Protochlorophyllide

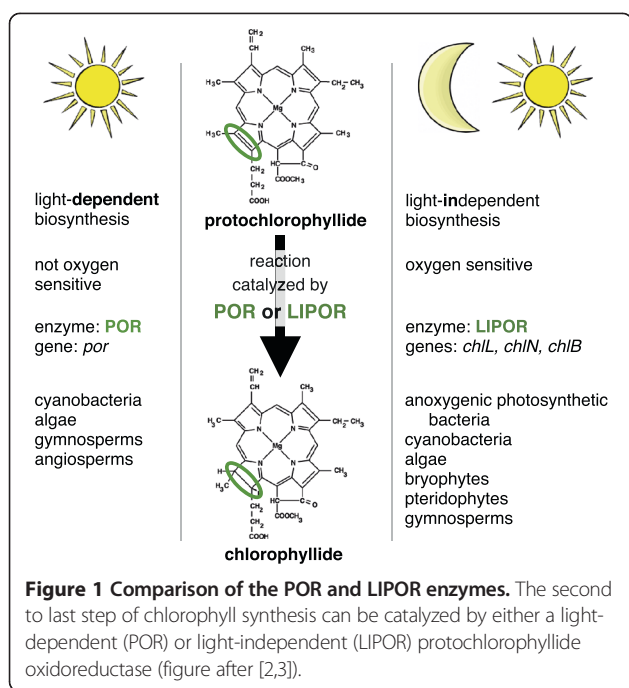
## Background

Chlorophyll *a* is synthesized entirely within the chloroplast, progressing in a series of enzymatic steps from the first committed precursor, 5-aminolevulinate, to the end product chlorophyll *a* [1]. The second to last step of this reaction sequence transforms the pigment protochlorophyllide to chlorophyllide via the reduction of a double bond. This step can be catalyzed by either of two non-

homologous, isofunctional enzymes: the light-independent (LIPOR) or the light-dependent (POR) protochlorophyllide oxidoreductase (Figure 1) [2,3].

The evolutionary origins and occurrence of POR and LIPOR oxidoreductases differ. LIPOR first arose in anoxygenic photosynthetic bacteria, likely evolving from a nitrogenase [4,5]. Similar to nitrogenase in structure, this enzyme is comprised of one or two L-protein homodimers encoded by the *chlL* gene and an NB-protein heterotrimer encoded by the genes *chlN* and *chlB*. Also like nitrogenase, the LIPOR holoenzyme contains iron-sulfur

\* Correspondence: racat@uw.edu  
Department of Biology, University of Washington, Seattle, WA, USA



clusters that confer sensitivity to oxygen [6-9]. In contrast, the POR enzyme arose in cyanobacteria [3], the first oxygenic photosynthesizers which are also thought to be responsible for the oxygenation of Earth's atmosphere [10]. It is postulated that the POR enzyme arose under strong selective pressures for an enzyme that would be unaffected by oxygen [8,11]. The POR enzyme, encoded by the *por* gene, is a globular protein with high sequence similarity to other members of the short-chain dehydrogenase-reductase (SDR) family. Although the POR enzyme is insensitive to oxygen, it has its own Achilles' heel. The enzyme is only active when its pigment substrate absorbs light [12] and thus, unlike LIPOR, POR cannot facilitate chlorophyll synthesis in the dark.

Endosymbiotic theory holds that chloroplasts originated when a non-photosynthetic protist engulfed and maintained cyanobacteria-like cells [13]. It is hypothesized that a single 'primary' endosymbiotic event generated the glaucophytic, rhodophytic and chlorophytic algae, as well as the viridiplantae ([14] but see [15,16]). In subsequent 'secondary' endosymbioses, the ancestors of euglenid and chlorarachniophyte algae each phagocytized and retained chlorophytes as plastids [17]. The origins of the cryptophyte, alveolate (e.g., dinoflagellate), stramenopile and haptophyte algae (collectively termed CASH) are less clear. Whereas nuclear genes show CASH host lineages to be polyphyletic [18], plastidial genes support a single, rhodophytic origin for their chloroplasts [19-22]. Synthesizing earlier views [23-25], the rhodoplex hypothesis describes any number of scenarios in which the initial CASH plastid was obtained via a

secondary endosymbiotic event and transferred between or even within CASH lineages by tertiary and potentially higher order endosymbioses [26].

During the establishment of the proto-chloroplast, and in those organisms of serial endosymbiotic origin, most of the genes required for photosynthesis and organellar homeostasis were transferred from the endosymbiont to the host nucleus in a process known as endosymbiotic gene transfer (EGT). In fact, ~18% of *Arabidopsis* genes are of cyanobacterial origin [27]. In extant cyanobacteria, POR and LIPOR genes are each present as single copies. In eukaryotic algae, the gene encoding POR appears to have been transferred to the nucleus, whereas LIPOR genes remain chloroplast-localized when present (these three genes are lost in many photosynthetic organisms).

Regardless of coding location, genetic restructuring can also be catalyzed by horizontal gene transfer (HGT), the process whereby xenologs (foreign genes) are incorporated into the genome of an organism. Although HGT was once thought to occur rarely, it is now recognized as a potentially pervasive force in genetic restructuring [28]. Transferred genes can originate from a variety of sources, including phagocytized prey, symbioses, viral transfection, and potentially other sources not yet identified [29,30]. Recent, intense sequencing efforts across a broad representation of prokaryotes and eukaryotes have resulted in extensive documentation of horizontal gene transfer (e.g., [31-35]). For example, Archibald et al. [36] analyzed nuclear-encoded, plastid-targeted genes of a chlorarachniophyte and found that up to 21% of the studied genes were derived from foreign sources. Such high rates of HGT observed in microbes are hypothesized to result from a 'gene transfer ratchet', wherein a small probability of gene incorporation multiplied by many gene uptake events over time results in many orthologous as well as novel genes in microbial genomes [29]. Apart from chance incorporation, the successful integration of a transferred gene has been shown to be highest for genes: (a) involved in few or no protein-protein interactions [37,38]; (b) not involved in information processing (e.g., DNA replication, RNA transcription, and protein translation; [39]); and (c) expressed at low levels [40]. *Por* genes (see below) appear to fulfill these criteria.

Whether of ancestral or foreign origin, the duplication of resident genes serves as an additional source of genetic novelty. Gene duplication (i.e., the generation of gene paralogs) can potentially impact metabolic processes on several levels. Most simply, gene dosage is increased for the duplicated gene. Alternatively, mutations in regulatory regions or coding sequences can effectively partition a gene's ancestral roles among the paralogs. If one copy mutates extensively, a novel protein may be generated.

In many cases, genetic change arising from gene duplication provide an adaptive advantage and become fixed in the population (reviewed in [41]). Recent studies [42,43] suggest that the diatom *Phaeodactylum tricornutum* uses more than one POR enzyme for chlorophyll synthesis. Multiple *por* genes have also been annotated in three additional diatom genomes (*Thalassiosira pseudonana*, *Fragilariopsis cylindrus*, and *Pseudo-nitzschia multiseriata* [44]). These observations generate many questions concerning POR duplication in diatoms as well as other algal species: Did extra *por* genes arise from HGT or gene duplication? Could the maintenance of redundant *por* genes account for the apparent loss of the genes encoding LIPOR (*chlL*, *chlN*, and *chlB*) in some chloroplast genomes [4,45,46]? Under what circumstances would the presence of both non-homologous, physiologically distinct protochlorophyllide oxidoreductases be advantageous to an organism?

In this paper we explore the nature of genetic novelty and genetic redundancy with respect to both the light-dependent (POR) and light-independent (LIPOR) enzymes that catalyze the penultimate step of chlorophyll synthesis. We find a reticulate *por* gene history within the algae, evidencing multiple horizontal gene transfer events including one that offers evidence of a close association of the plastids of haptophyte and stramenopile algae. Furthermore, we identify several *por* gene duplications and the presence of both ancestral and xenologous *pors* in some algal taxa. We also show a propensity for algae maintaining multiple *por* genes to lose their chloroplastic LIPOR genes (*chlL*, *chlN*, and *chlB*). Genetic redundancy, whether arising from non-homologous iso-functional enzymes, gene duplication, or horizontal gene transfer, fosters metabolic innovation. These data present an exciting opportunity to compare the fate of uniquely redundant protochlorophyllide oxidoreductase genes across evolutionarily close and distant lineages.

## Results and discussion

### POR protein phylogeny inference

To explore the evolution of *por* genes, a database was compiled from: (a) in-house amplification and sequencing of stramenopile *por* genes; (b) the recently sequenced genome of the haptophyte *Chrysochromulina tobin* (Hovde, Starckenburg and Cattolico, in prep.) and (c) publically available genomes, transcriptomes, and sequences. Because the POR protein is affiliated with the large, fairly conserved SDR protein family [47,48], e-values alone were not used to identify *por* genes. Instead all putative *por* sequences were screened for the presence of specific motifs encompassing particular lysine, tyrosine, and cysteine residues. These amino acids were experimentally shown to be essential to POR enzyme catalytic function in cyanobacteria and land plants (Figure 2; [49-53]). Analyses showed that these criteria eliminate

homologs from cyanobacteria as well as chlorophytic and CASH algae that comprise strongly supported branches that cannot be placed within a POR phylogeny with statistical certainty. These protein clusters potentially represent closely-related SDR proteins that perform distinct, as-yet-undescribed functions [54]. The use of specific amino acid diagnostic characters also eliminates two *por* genes that were putatively identified in microarray-based studies of the *P. tricornutum* chlorophyll synthesis pathway ([42,43]; their *por3* and *por4*).

The resultant alignment of the 274 amino acid conserved core of 275 POR proteins from cyanobacteria, eukaryotic algae and land plants represents 162 taxa (Figure 3). The Whelan and Goldman matrix for globular proteins (WAG, [55]), a gamma shape parameter and a proportion of invariable sites were found to best fit the data and were therefore used in Bayesian and maximum-likelihood phylogenetic inference. Both methods returned nearly identical topologies. The entirety of the Bayesian phylogeny is shown in Figure 3A. Details of this gene tree are shown in Figures 3B and 4. Bayesian posterior probabilities and maximum-likelihood bootstrap values are indicated on all principal branches (Figures 3B and 4), with dashed branches indicating less than 0.95 posterior probability throughout the tree. The amino acid alignment and Bayesian and maximum-likelihood trees (with sequence accessions) are available in Additional files 1, 2, and 3. The identities of all sequences are tabulated in Additional file 4.

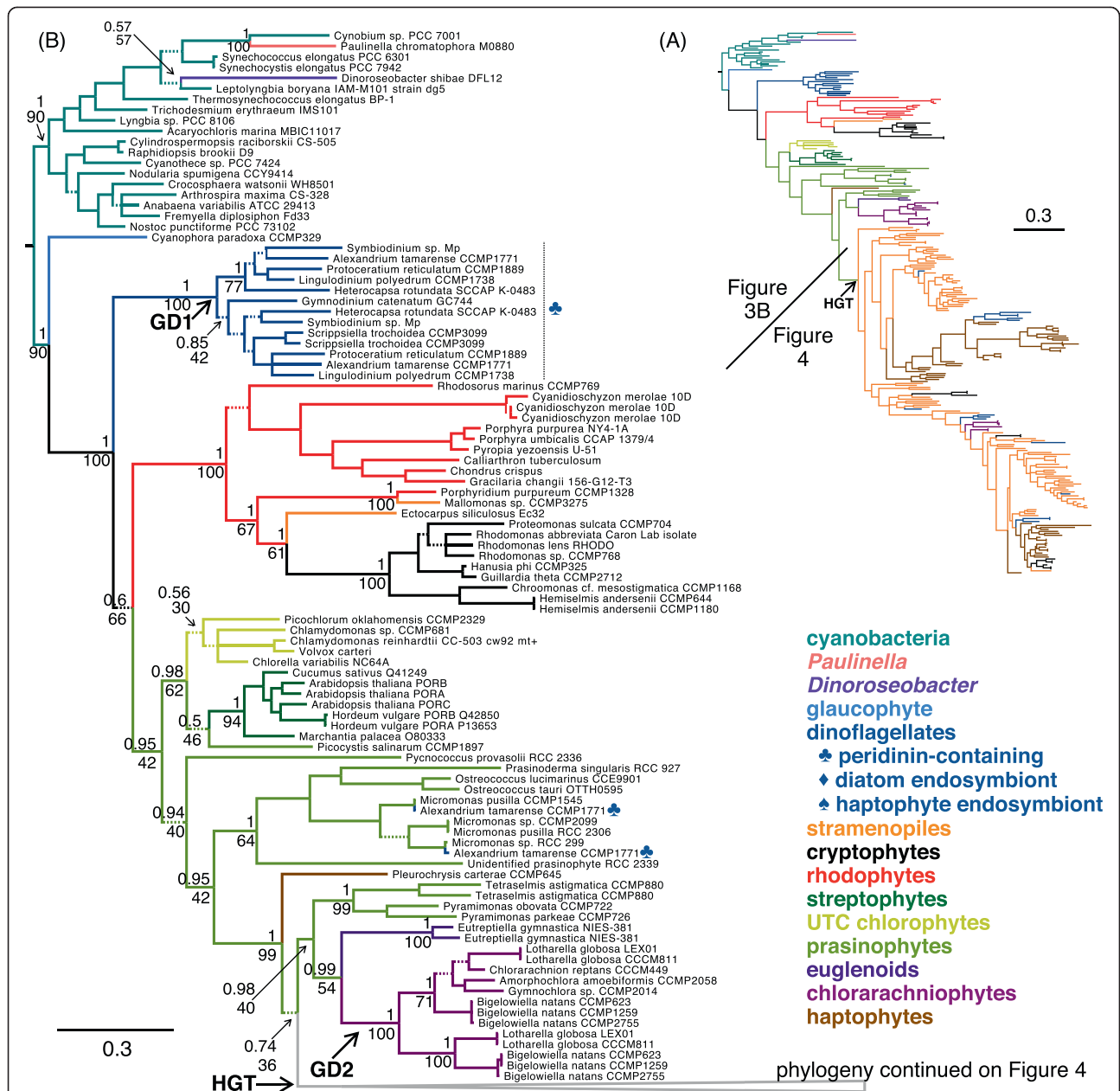
Given the cyanobacterial origin of the POR enzyme [3,47], POR proteins from this taxon were used to root the phylogeny shown in Figure 3. The POR protein phylogeny backbone follows the expected pattern based on current knowledge of the relationships among algal taxa originating from primary endosymbiosis. The *Paulinella chromatophora* POR clusters within the cyanobacterial outgroup, reflecting its close association with cyanobacteria as an alga derived from a unique primary endosymbiosis [56]. Among the Archaeplastida, the glaucophytic, rhodophytic and chlorophytic (including streptophyte, ulvophyte-trebouxiophyte-chlorophyte (UTC) clade, and prasinophyte) lineages branch deeply as expected [14,57], interrupted only by a branch of dinoflagellate POR proteins (discussed below). Our extensive POR protein phylogeny also confirms the cyanobacterial origin of the POR of *Dinoroseobacter shibae*, an anoxygenic phototroph believed to have obtained a *por* gene via horizontal transfer [58].

### Replacement of ancestral *por* gene with horizontally transferred *por* gene in stramenopile and haptophyte algae

Because the CASH algal lineages obtained their plastids from a rhodophytic source, the POR proteins of these lineages are expected to nest within the rhodophytic POR branch. As shown in Figure 3B, POR proteins of



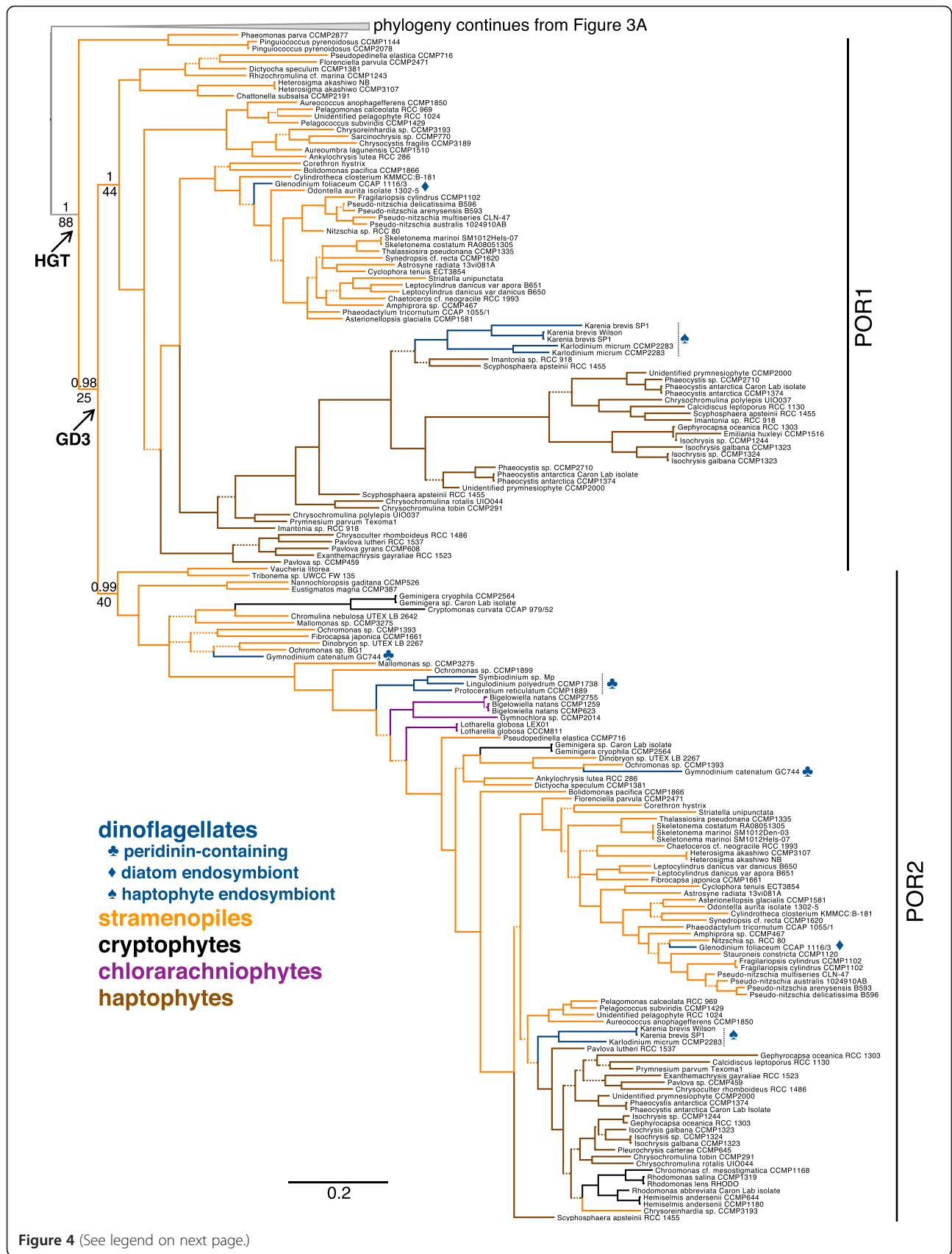




**Figure 3** *Por* gene tree: rhodophytic identity of cryptophyte and some stramenopile *por*s; duplication of dinoflagellate and chlorarachniophyte *por*s. (A) Outline of the full *por* gene tree inferred from the 274 amino acid conserved core of 275 POR proteins from cyanobacteria, eukaryotic algae and land plants, representing 162 taxa. Branches are colored according to algal lineage (see legend). The corresponding, detailed phylogeny is split between Figures 3B and 4. Scale bar indicates 0.3 amino acid substitutions per site. (B) Basal portion of *por* gene tree. Branches are colored according to algal lineage (see legend), with symbols indicating origin of endosymbiont in dinoflagellate taxa whose ancestral plastids have been replaced. Bayesian and maximum-likelihood analyses recovered nearly identical trees. Posterior probabilities are shown above branches and bootstrap support is shown below branches. All dashed branches have less than 0.95 posterior probability. Scale bar indicates 0.3 amino acid substitutions per site. Gene duplication (GD) and horizontal gene transfer (HGT) events are indicated with arrows.

The presence of only two stramenopiles that exhibit a POR protein in the rhodophyte clade (Figure 3B) is enigmatic. It is unclear whether most phaeophyceans (e.g., *E. siliculosus*) or synurophyceans (e.g., *Mallomonas* sp.) retain a rhodophytic-type POR. These two stramenopile

classes are not closely related to one another, but rather belong to the stramenopile SI and SII clades (of three total clades), respectively [60]. They are each more closely related to classes whose members appears to solely possess the prasinophytic *por* gene. Stramenopile classes in this



(See figure on previous page.)

**Figure 4 Por gene tree: duplication of xenologous stramenopile/haptophyte *por* genes.** Bottom half of the *por* gene tree outlined in Figure 3A. Branches colored by lineage, with symbols indicating origin of endosymbiont in dinoflagellate taxa whose ancestral plastid has been replaced (see legend). Posterior probabilities are shown above branches and bootstrap support are shown below branches. All dashed branches have less than 0.95 posterior probability. Scale bar indicates 0.3 amino acid substitutions per site. Arrows indicate the inferred horizontal gene transfer (HGT) of a *por* gene from prasinophytes to the stramenopiles, and the subsequent gene duplication (GD3) to create *por1* and *por2*.

study that possess *por* genes of prasinophytic origin include members of SI (Raphidophyceae and Xanthophyceae), SII (Chrysophyceae, Eustigmatophyceae and Pinguicophyceae), as well as the SIII (Bacillariophyceae, Bolidophyceae, Dictyochophyceae, and Pelagophyceae) clades. The presence of the prasinophytic *por* gene in all three stramenopiles clades, the derived position of the Phaeophyceae and Synurophyceae within the stramenopiles, and the presence of both rhodophytic and prasinophytic *por* genes in *Mallomonas* sp. suggest that the xenologous *por* gene was obtained by stramenopiles early in their evolution. Similarly, the absence of rhodophytic *por* genes in haptophytes and presence in each sampled species of the xenologous *por* gene suggests that the rhodophytic *por* gene of haptophytes was replaced early in their evolution.

The *por* gene identities of cryptophytes are highly variable. A rhodophytic *por* appears to be the only *por* in *Guillardia theta*, *Hanusia phi*, *Proteomonas sulcata*, and *Rhodomonas* sp. Cryptophytes bearing rhodophytic *por*s that also have *por*s related to the clade of xenologous stramenopile/haptophyte *por*s include *Chroomonas* cf. *mesostigmatica*, *Hemiselmis andersenii*, *Rhodomonas abbreviata*, and *Rhodomonas lens*. Lastly, *Cryptomonas curvata*, *Geminigera cryophila*, *Geminigera* sp., and *Rhodomonas salina* appear to have only the xenologous stramenopile/haptophyte *por*. However, it is possible that not all *por* genes were present in the sampled transcriptomes. Notably, within the branch of xenologous stramenopile *por* genes (Figure 4), stramenopile and haptophytic *por* genes generally cluster together. This association is not true for cryptophyte *por* genes, which are spread among three branches far apart from one another, suggesting that the xenologous stramenopile/haptophyte *por* genes of cryptophytes might originate from several independent HGTs or possibly represent phylogenetic artifacts or transcriptome contamination [61].

Similar to the cryptophytes, several peridinin-containing dinoflagellate species (*Gymnodinium catenatum*, *Symbiodinium* sp., *Lingulodinium polyedrum* and *Protoceratium reticulatum*) and several chlorarachniophytes (*Lotharella globosa*, *Gymnochlora* sp., *Bigelowiella natans*) possess a copy of the xenologous stramenopile/haptophyte *por*. The punctate nature of these POR protein identities within the phylogeny (Figure 4) suggests that the genes were obtained by HGT, though artifacts of poor phylogenetic signal or contamination must be considered. Extensive

gene acquisition via HGT from a variety of sources has been documented to occur in dinoflagellates (e.g., [32,62-64]) and the chlorarachniophyte *B. natans* [36,65,66]. Given the propensity of dinoflagellates to obtain exotic genes, we also note that the peridinin-containing dinoflagellate *Alexandrium tamarense* appears to harbor two prasinophytic POR proteins, possibly obtained from a unique HGT event sourced from a *Micromonas*-like species (Figure 3B).

#### Duplication of the xenologous stramenopile/haptophyte *por* gene

The branches of the POR protein phylogeny pertaining to the xenologous POR enzymes are shown in Figure 4. Multiple POR xenologs are recovered from many of the surveyed stramenopile and haptophyte species. In most cases, each of two gene copies is distributed between two principal branches in the tree—evidence that a gene duplication event has occurred. Whether this duplication event took place: (a) in the lineage that first obtained the xenolog (most parsimonious); (b) in an unsampled prasinophyte lineage prior to the HGT of both paralogs, or (c) resulted from a near simultaneous incorporation of two copies of the same gene, cannot be determined with certainty.

In Figure 4a, the node representing the gene duplication event is labeled as GD3, and resultant POR paralogs are annotated as POR1 (gene: *por1*) and POR2 (gene: *por2*). The maintenance of both POR1 and POR2 is fairly-well conserved: 16 of 22 haptophyte taxa and 34 of 56 stramenopile taxa possess both *por1* and *por2* genes. Those lacking the full *por1/por2* gene complement possess solely one paralog or occasionally two of one paralog (see Additional file 4 for data in tabular format). Note that incomplete transcriptomic data or poor gene predictions in genomic datasets may obscure the identification of a second paralog for some species in this study.

One of two *Pleurochrysis carterae* (Haptophyta; Coccolithales) POR proteins, as well as the sole POR proteins of *Pinguicoccus* spp. and *Phaeomonas parva* (Stramenopila; Pinguicophyceae) branch before the duplication event shown in Figure 4. The placement of the *P. carterae* *por* within prasinophytic PORs (Figure 3) may simply be an artifact of phylogenetic uncertainty, since its placement changed as taxa were added to the phylogeny (data not shown). Alternatively, the *P. carterae* *por* could represent

a unique HGT event from a prasinophyte to just this taxon. The second *P. carterae por* placed as expected within the stramenopile/haptophyte POR2 clade. The Pinguiophyceae are not expected to be sister to the rest of the stramenopiles [60,67], thus the placement of their PORs at the base of the duplication event is enigmatic. Support for the monophyly of the xenologous stramenopile/haptophyte PORs is high (posterior probability 1, bootstrap 88). The POR1 and POR2 branches are distinguished with high posterior support (1 and 0.99, respectively) but low bootstrap support (44 and 40, respectively). Given the subsampling algorithm used in bootstrapping, the low bootstrap support at these nodes likely reflects the fact that a small subset of amino acid positions are diagnostic for the stramenopile/haptophyte POR1 versus POR2 proteins, as shown in Figure 2 [68].

#### Evolutionary significance of the xenologous stramenopile/haptophyte *por* genes

Researchers studying algae bearing plastids of rhodophytic origin frequently document the occurrence of nuclear-encoded genes of chlorophytic origin. For example, a phosphoribulokinase of chlorophytic origin was found in CASH algae [31], and Frommolt et al. [69] reported that five of the 16 carotenoid biosynthesis genes in cryptophyte, haptophyte, and stramenopile algae were also from chlorophytes. Recent meta-analyses of genomic data have reported the presence of many chlorophytic genes in diatoms (Stramenopila; [70]), a chromerid (Alveolata; [71]) and pico-prymnesiophytes (Haptophyta; [72]). Some researchers have attributed high levels of green genes in CASH lineages to putative cryptic endosymbiotic gene transfer (EGT) events (e.g., [69,70,72]), while others have invoked poor taxon sampling, a lack of manual curation, and phylogenetic error to explain these findings [71,73,74]. We note that, although the possibility for phylogenetic error is omnipresent, our study benefits from: (a) the inclusion POR protein sequences from many rhodophytic (including mesophilic), chlorophytic, and other algal taxa; (b) manual curation of sequence data and alignments; (c) special attention paid to support values at key nodes on the tree when making inferences about sequence origin; as well as (d) data exploration [e.g., in a POR protein phylogeny inferred without chlorophytic PORs, the stramenopile/haptophyte POR clade remained sister to rather than derived from rhodophytic PORs (data not shown)]. Furthermore, the horizontal transfer of at least some genes can be expected for phagocytotic algae (or algae with phagocytotic ancestors) [29]. Representatives within the cryptophytes, haptophytes, stramenopiles and dinoflagellates are commonly phagocytotic.

Whereas the prasinophytic origin of these xenologous POR proteins is unambiguous, the history of these

xenologs among CASH taxa is less clear. Like chloroplast-encoded genes, nuclear-encoded chloroplast-targeted genes serve as markers for plastid origin because they are transferred from the symbiont to the host during endosymbiosis. However, HGT presents another potential route of transfer between lineages that can obscure relationships among groups.

As discussed above, the punctate nature of the xenologous *por* gene distribution in cryptophytes, dinoflagellates, and chlorarachniophytes suggests that these genes were obtained from several unique HGT events to these groups. In contrast, the xenologous *por* genes appear in all but one stramenopile and all haptophytes in our extensive sampling of 11 classes and all three clades of stramenopiles as well as six orders of haptophytes including the basal lineage Pavlovales [75]. The ubiquitous presence of the xenolog in stramenopiles and haptophytes suggests that this prasinophytic *por* was acquired early in the evolution of these taxa.

Fascinatingly, stramenopile *por*s are found ancestral to haptophyte *por*s in the phylogeny presented in Figure 4, especially those from members of the Pelagophyceae. Although statistical support for the exact placement of haptophyte PORs (and PORs of dinoflagellates with haptophyte-derived plastids) in the phylogeny is weak, stramenopile PORs occupy strongly supported basal nodes within both the POR1 and POR2 branches. The derived position of the haptophyte PORs suggests transfer of the xenologous *por* genes from stramenopiles to haptophytes. Under the aforementioned rhodoplex hypothesis, an endosymbiotic origin for the *por* xenolog duplicates of haptophytes would necessarily invoke plastid transfer from the stramenopiles to the haptophytes, likely after the stramenopile plastid lineage diverged from that of cryptophytes (which retain a relic nucleomorph unlike other CASH plastids [76]).

The relationship between the plastids of stramenopile and haptophytic algae is presently unresolved [18,25,26,77,78]. Using a BLAST-based statistical approach, Stiller and colleagues recently documented strong support for a model of serial endosymbiosis wherein the plastid was transferred from rhodophytes → cryptophytes → stramenopiles → haptophytes [79]. Furthermore, some plastid phylogenies find stramenopiles and haptophytes sister to one another to the exclusion of cryptophytes (e.g., [17,78]). Assuming serial endosymbiosis from stramenopiles to haptophytes, limited taxon sampling may explain why haptophytes were observed sister to rather than derived from the stramenopiles in these plastid phylogeny studies. The present study, although limited to just one gene, incorporates a diverse array of haptophytes and stramenopiles and may therefore be expected to better resolve such a relationship. Under this scenario, low support for the exact placement of



haptophytic *por*s may be due to extinction of the stramenopile donor taxon.

In contrast to the above findings, other plastid phylogenies find cryptophytes and haptophytes more closely related to one another than they are to stramenopiles (e.g., [79-81]). Importantly, a shared, horizontally transferred *rpl36* gene encoded in the chloroplasts of only cryptophytes and haptophytes strongly indicates a sister relationship between these two taxa [82,83]. If haptophyte and cryptophyte plastids are indeed more closely related to one another than to stramenopile plastids, the xenologous *por* genes would have to have been transferred from stramenopiles to haptophytes via HGT early in the evolution of the haptophytes. However, just as a consensus concerning the relationships of CASH plastids has not yet been reached, the relative ages of the various CASH lineages remain unresolved [84-86].

#### **POR protein identity post-duplication**

Stramenopile/haptophyte POR protein identity post-duplication is demonstrated in the sequence logos of Figure 2. Diatom POR1 and POR2 amino acid sequences were used to best represent these xenologous POR proteins without the many small gaps present in an alignment of all stramenopile/haptophyte PORs. High sequence conservation is shown when POR1 and POR2 are compared to ancestral, cyanobacterial PORs. The core region of diatom PORs (excluding signal and transit peptides and a C-terminal extension on diatom POR2) share 60% sequence similarity with cyanobacterial PORs. Each diatom POR appears to maintain conserved regions particular to that POR protein as well as to all POR proteins; sequence similarity is 75% within diatom POR1, whereas sequence similarity is lower at 66% within diatom POR2, principally due to a poorly conserved C-terminal extension. This C-terminal extension results in a predicted protein of ~60kD rather than the typical ~40kD (Figure 2, [87]). Sequencing the *Phaeodactylum por2* cDNA amplified by 3' RACE shows that this extension is transcribed (Hunsperger and Cattolico, unpub.). Notably, antibodies raised against heterologously expressed, full-length *Phaeodactylum* POR2 proteins show cross reactivity to a 40kD product in *Phaeodactylum* cell extracts. This observation suggests that POR2 is post-transcriptionally truncated to a conventional POR2 size (Hunsperger and Cattolico, unpub.). Thus both proteins are expected to be functional.

#### **Identity of *por* genes in dinoflagellates with haptophyte and diatom endosymbionts**

The POR proteins of dinoflagellates whose plastids have been replaced by those of a haptophyte (*Karenia brevis*, *Karlodinium micrum*) or diatom (*Glenodinium foliaceum*, a "dinotom") appear to originate from the haptophyte or

diatom endosymbiont, respectively (Figure 4). Just as diatoms and haptophytes each have two unique POR proteins, dinoflagellates that bear plastids originating from these algal sources also possess these same two unique POR proteins. Given that the haptophyte and diatom plastids replaced the ancestral peridinin-containing plastids, it is expected that ancestral POR proteins were already integrated into the dinoflagellate nuclear genome. These ancestral POR proteins were lost, however, rather than re-targeted to the new chloroplast. One might speculate that regulatory or functional schemes unique to each of the new endosymbiont's two *por* genes favored the retention of these new *por* genes. It would be interesting to determine whether this *por* gene substitution pattern also holds for dinoflagellates bearing chlorophyte (e.g., *Lepidodinium*; [88]) or ephemeral cryptophyte (e.g., *Dinophysis*; [89]) derived plastids. Whereas haptophytic endosymbionts no longer possess nuclei, identifiable nuclei remain in diatom endosymbionts [90]. As the dinotom *por* genes used in this study were obtained from transcriptomes, it is unclear whether they are encoded within the endosymbiont's nucleus or have been transferred to the dinoflagellate nucleus. Nonetheless, the diverse origins of dinoflagellate POR proteins reflect the propensity of members of this taxon for foreign gene acquisition, endosymbiont replacement and genetic remodeling (e.g., [32,62-64,90,91]).

#### **Additional *por* gene duplications**

##### ***Duplicated dinoflagellate-specific por genes***

Because chloroplasts of ancestral, peridinin-containing dinoflagellates have been shown to be of rhodophytic origin [20,59,92], the *por* genes of dinoflagellates can be expected to group within the rhodophytes. Instead, a dinoflagellate-specific group of POR proteins is found sister to rhodophytic and chlorophytic algae, indicating an unresolved origin for this unique group of enzymes (Figure 3).

The recurrence of five dinoflagellate taxa in each of the two main branches of the dinoflagellate POR subtree is classic evidence of a gene duplication event (annotated in Figure 3A as GD1). Low support values for one of these branches, however, makes the nature of the gene duplication less clear. Because all seven taxa in these branches utilize peridinin, which is thought to be the ancestral photosynthetic dinoflagellate pigment [90], these paralogous POR proteins of uncertain origin may have been acquired early in the evolution of the dinoflagellates.

##### ***Duplication of chlorarachniophyte and euglenoid por genes***

The euglenids and chlorarachniophytes are algal lineages originating from two separate secondary endosymbioses involving chlorophytic algae. The euglenid chloroplast originates from the Pyramimonadales lineage of the

prasinophyte algae, whereas the chlorarachniophytes engulfed an alga of uncertain identity from the UTC clade [17,88,93,94]. Phylogenetically, the POR proteins of a euglenid and several chlorarachniophytes are found sister to one another and nested within prasinophyte algae closest to a branch containing pyramimonads (*Pyramimonas* spp.) and a chlorodendrophyte (*Tetraselmis astigmatica*). The placement of the euglenid PORs is broadly congruent with the known origin of their plastids from pyramimonads. We note that the sole Euglenoid included in these studies, *Eutreptiella gymnastica*, appears to possess two *por* genes, perhaps indicating that a *por* gene duplication event occurred in this taxon.

A sister relationship between the euglenids and chlorarachniophytes, however, is inconsistent with the separate origins of the plastids of these two groups. Improper placement of the chlorarachniophyte POR proteins may be due to the inclusion of very few UTC chlorophyte species and few euglenids in the POR protein tree. Alternatively, POR placement could reflect a horizontal gene transfer from the prasinophytes to the chlorarachniophytes, though additional taxon sampling would be necessary to verify such an event.

These chlorarachniophyte-specific POR proteins show evidence of gene duplication (labeled GD2 in Figure 3B). Three strains of *Bigelowiella natans* and two strains of *Lotharella globosa* appear to each have two chlorarachniophyte-specific POR proteins that are divided between the two main branches in this clade. The basal position of the split between the two POR paralogs supports a gene duplication event in the common ancestor of most chlorarachniophytes, given that *Amorphochlora amoebiformis* possesses one of the paralogs and represents an early diverging branch of chlorarachniophytes [95]. It is unclear whether *Amorphochlora amoebiformis*, *Gymnochlora* sp., and *Chlorarachnion reptans* then lost one paralog, or whether incomplete transcriptomic data impeded the recovery of the second paralog.

#### Physiological significance of multiple *por* genes

The discovery of multiple *por* genes in a species is not without precedent. Although some species of Viridiplantae are confirmed to have just one *por* gene, numerous representatives within this taxon encode multiple *por* genes (reviewed in [47]). Phylogenetic analysis suggests that some of the angiosperm POR paralogs are shared among select plant species, while other paralogs arose more recently and are unique to a particular taxon. In vascular plants, light and developmental stage appears to regulate the expression of each *por* gene. For example, in the angiosperm *Arabidopsis thaliana*, two POR isoenzymes, PORA and PORB, accumulate in dark-adapted seedlings in concert with increasing levels of the

pigment substrate Pchl<sub>id</sub>. As a result, the plant is poised for chlorophyll synthesis upon illumination of the seedling. PORA is quickly degraded upon seedling exposure to light, while PORB continues to be expressed in mature tissues and thus is primarily responsible for continued chlorophyll production. A third POR, PORC, is up-regulated with increasing light intensity, enabling higher chlorophyll abundances under high light (reviewed in [47,87]).

Given intrinsic differences between the life histories, physiologies and ecologies of land plants and algae, it will be interesting to compare their regulatory and functional schemes for chlorophyll biosynthesis. One might anticipate that, similar to land plants, the regulation of multiple *por* homologs in algae may be tied to light availability (varying with time of day, season, water turbidity and depth) and developmental stage (e.g., encystment/excystment). For example, whereas many land plants increase their chlorophyll levels in response to high light [96], algal chlorophyll levels decrease as light intensity increases [97]. As expected, the transcription of both *Phaeodactylum tricorutum por1* and *por2* were found to be initially down-regulated in response to a transition from low (35  $\mu\text{M photons m}^{-2} \text{ s}^{-1}$ ) to high (500  $\mu\text{M photons m}^{-2} \text{ s}^{-1}$ ) light levels [42]. Our own RT-qPCR measurements of *P. tricorutum por1* and *por2* mRNA abundance shows a unique oscillatory pattern for each gene over a 12 hour light:12 hour dark photoperiod (Hunsperger and Cattolico, in prep), suggesting independent regulation of these two genes. Similarly, transcriptomic analysis of 12 hour light:12 hour dark synchronized *Chrysochromulina tobin* (Haptophyta; Prymnesiales) cultures indicates that the two *por* genes independently respond to the imposed light/dark cues in a pattern that differs from that seen for *P. tricorutum por1* and *por2* (Hovde and Cattolico, unpub).

#### Loss of chloroplastic genes encoding LIPOR

At least one *por* gene, encoding the light-dependent prochlorophyllide oxidoreductase (POR), has been documented in all sequenced algal nuclear genomes. In contrast, chloroplast genome sequencing has revealed the loss or degradation of the three chloroplast-localized genes encoding the light-independent prochlorophyllide oxidoreductase (LIPOR; *chlL*, *chlN*, and *chlB*) in members of the chlorophytic, euglenoid and chlorarachniophyte algae (Table 1) as well as rhodophytic and CASH algae (Table 2) (see also [4,45,46,80,94]). Furthermore, the loss of these genes is well documented for angiosperms (reviewed in [4]).

Typically, the three LIPOR genes are either entirely present or completely absent from a chloroplast genome. Notably, both sampled chlorarachniophytes, all four euglenoids, and all five haptophytes lack LIPOR genes in

**Table 1 Distribution of *por* genes and chloroplast-encoded LIPOR genes in chlorophytic algae, chlorarachniophytes, and euglenids**

Taxon	Species	Culture ID	<i>Por</i> genes in genus (this paper)	Chloroplast-encoded LIPOR genes			Chloroplast genome accession
				<i>chlL</i>	<i>chlN</i>	<i>chlB</i>	
<b>Chlorophyta</b>							
Chlorophyceae	<i>Acutodesmus obliquus</i>	UTEX 393		+	+	+	NC_008101
Chlorophyceae	<i>Chlamydomonas reinhardtii</i>	n/a	1	+	+	+	NC_005353
Chlorophyceae	<i>Dunaliella salina</i>	CCAP 19/18		+	+	+	NC_016732
Chlorophyceae	<i>Floydiella terrestris</i>	UTEX 1709		+	+	+	NC_014346
Chlorophyceae	<i>Gonium pectorale</i>	K3-F3-4		+	+	+	NC_020438
Chlorophyceae	<i>Oedogonium cardiacum</i>	SAG 575-1b		+	+	+	NC_011031
Chlorophyceae	<i>Pleodorina starii</i>	NIES 1363		+	+	+	NC_021109
Chlorophyceae	<i>Schizomeris leibleinii</i>	UTEX LB 1228		+	+	+	NC_015645
Chlorophyceae	<i>Stigeoclonium helveticum</i>	UTEX 441		+	+	+	NC_008372
Mamiellophyceae	<i>Micromonas pusilla</i>	RCC299	1	-	-	-	NC_012575
Mamiellophyceae	<i>Monomastix</i> sp.	OKE-1		-	-	-	NC_012101
Mamiellophyceae	<i>Ostreococcus tauri</i>	OTTH0595	1	-	-	-	NC_008289
Nephroselmidophyceae	<i>Nephroselmis olivacea</i>	NIES 484		+2	+2	+2	NC_000927
Prasinophyceae	<i>Pycnococcus provasolii</i>	CCMP1203	1	+	+	-	NC_012097
Prasinophyceae	<i>Pyramimonas parkeae</i>	CCMP726	1	+2	+2	+	NC_012099
Trebouxiophyceae	<i>Chlorella variabilis</i>	NC64A	1	+	+	+	NC_015359
Trebouxiophyceae	<i>Chlorella vulgaris</i>	C-27	1	+	+	+	NC_001865
Trebouxiophyceae	<i>Coccomyxa subellipsoidea</i>	C-169		+	+	+	NC_015084
Trebouxiophyceae	<i>Leptosira terrestris</i>	UTEX 333		+	+	+	NC_009681
Trebouxiophyceae	<i>Parachlorella kessleri</i>	SAG 211/11 g		+	+	+	NC_012978
Trebouxiophyceae	<i>Pedinomonas minor</i>	UTEX LB 1350		-	-	-	NC_016733
Trebouxiophyceae	<i>Trebouxiophyceae</i> sp.	MX-AZ01		+	+	+	NC_018569
Ulvophyceae	<i>Bryopsis hypnoides</i>	n/a		+	+	+	NC_013359
Ulvophyceae	<i>Oltmannsiellopsis viridis</i>	NIES 360		+	+	+	NC_008099
Ulvophyceae	<i>Pseudoclonium akinetum</i>	UTEX 1912		-	-	-	NC_008114
<b>Cercozoa</b>							
Chlorarachniophyceae	<i>Bigelowiella natans</i>	CCMP621	3	-	-	-	NC_008408
Chlorarachniophyceae	<i>Lotharella oceanica</i>	CCMP622	3	-	-	-	KF438023
<b>Euglenozoa</b>							
Euglenophyceae	<i>Euglena gracilis</i>	Z		-	-	-	NC_001603
Euglenophyceae	<i>Euglena viridis</i>	ATCC PRA-110		-	-	-	NC_020460
Euglenophyceae	<i>Eutreptiella gymnastica</i>	K-0333	2	-	-	-	NC_017754
Euglenophyceae	<i>Monomorphina aenigmatica</i>	UTEX 1284		-	-	-	NC_020018

(+) Present in chloroplast genome; (-) not present in fully-sequenced chloroplast genome. The number of *por* genes found in this study for a particular genus (not necessarily the same species or strain) is also indicated.

their chloroplasts, suggesting that LIPOR gene loss may have occurred early in the establishment of these lineages. In contrast, species with and species without chloroplastic LIPOR genes are documented for the chlorophytes, rhodophytes, cryptophytes, heterokonts, and chromerids. Such heterogeneity is seen even at the level of taxonomic

class, with some members maintaining and other members having lost LIPOR genes in the Trebouxiophyceae, Ulvophyceae and Prasinophyceae (Chlorophyta), Bangiophyceae and Florideophyceae (Rhodophyta), as well as the Dictyochophyceae, Pelagophyceae, and Raphidophyceae (Stramenopila). The prasinophycean *Pycnococcus*

**Table 2 Distribution of *por* genes and chloroplast-encoded LIPOR genes in rhodophytes and CASH algae**

Taxon	Species	Culture ID	Por genes in genus (this paper)	Chloroplast-encoded LIPOR genes			Chloroplast genome accession
				<i>chlL</i>	<i>chlN</i>	<i>chlB</i>	
<b>Rhodophyta</b>							
Bangiophyceae	<i>Cyanidioschyzon merolae</i>	Strain 10D	3	-	-	-	NC_004799
Bangiophyceae	<i>Cyanidium caldarium</i>	RK1		-	Δ	-	NC_001840
Bangiophyceae	<i>Porphyra purpurea</i>	Avonport	1	+	+	+	NC_000925
Bangiophyceae	<i>Pyropia haitanensis</i>	PH-38 (voucher)	1	+	+	+	NC_021189
Bangiophyceae	<i>Pyropia yezoensis</i>	U-51	1	+	+	+	NC_007932
Florideophyceae	<i>Calliarthron tuberculosum</i>		1	+	+	+	NC_021075
Florideophyceae	<i>Chondrus crispus</i>		1	-	-	-	NC_020795
Florideophyceae	<i>Gracilaria tenuistipitata</i> var. <i>liui</i>		1	-	-	-	NC_006137
Florideophyceae	<i>Gracilaria salicornia</i>	ARS08332 (voucher)	1	-	-	-	KF861575
Halymeniaceae	<i>Grateloupia taiwanensis</i>			-	-	-	NC_021618
Porphyridiophyceae	<i>Porphyridium purpureum</i>	NIES 2140	1	-	-	-	AP012987
<b>Cryptophyta</b>							
Chroomonadaceae	<i>Chroomonas mesostigmatica</i>	CCMP1168	2	+	ψ	?	EU233753; EU233756
Chroomonadaceae	<i>Chroomonas pauciplastida</i>	CCMP268	2	+	+	+	EU233754; EU233755; EU233748
Chroomonadaceae	<i>Hemiselmis andersenii</i>	CCMP644	2	+	+	+	EU233749; EU233750; EU233747
Chroomonadaceae	<i>Hemiselmis tepida</i>	CCMP443	2	+	+	?	EU233751; EU233752
Geminigeraceae	<i>Guillardia theta</i>		1	-	-	-	NC_000926
Pyrenomonadaceae	<i>Rhodomonas salina</i>	CCMP1319	1–2	ψ	ψ	ψ	NC_009573
<b>Haptophyta</b>							
Isochrysidales	<i>Emiliana huxleyi</i>	CCMP373	1	-	-	-	NC_007288
Pavloales	<i>Pavlova lutheri</i>	ATCC 50092	1–2	-	-	-	NC_020371
Phaeocystales	<i>Phaeocystis antarctica</i>	CCMP1374	2–3	-	-	-	NC_016703
Phaeocystales	<i>Phaeocystis globosa</i>	Pg-G(A)	2–3	-	-	-	NC_021637
Prymnesiales	<i>Chrysochromulina tobin</i>	CCMP291	2	-	-	-	KJ201907
<b>Stramenopila</b>							
Bacillariophyceae	<i>Fistulifera</i> sp.	JPCC DA0580		-	-	-	NC_015403
Bacillariophyceae	<i>Odontella sinensis</i>		2	-	-	-	NC_001713
Bacillariophyceae	<i>Phaeodactylum tricorutum</i>	CCAP1055/1	2	-	-	-	NC_008588
Bacillariophyceae	<i>Synedra acus</i>			-	-	-	NC_016731
Bacillariophyceae	<i>Thalassiosira oceanica</i>	CCMP 1005	2	-	-	-	NC_014808
Bacillariophyceae	<i>Thalassiosira pseudonana</i>	CCMP 1335	2	-	-	-	NC_008589
Dictyochophyceae	<i>Apedinella radians</i>	CCMP1767		-	-	-	unpublished data*
Dictyochophyceae	<i>Rhizochromulina marina</i>	CCAP950/1	1	+	+	+	unpublished data*
Eustimatophyceae	<i>Nannochloropsis gaditana</i>	CCMP526	1	+	+	+	KJ410682
Eustimatophyceae	<i>Nannochloropsis oceanica</i>	LAMB0001	1	+	+	+	KJ410683
Eustimatophyceae	<i>Nannochloropsis oculata</i>	CCMP525	1	+	+	+	KJ410684
Eustimatophyceae	<i>Nannochloropsis salina</i>	CCMP1776	1	+	+	+	KJ410685
Pelagophyceae	<i>Aureococcus anophagefferens</i>	CCMP1984	2	-	-	-	NC_012898
Pelagophyceae	<i>Aureombra lagunensis</i>	CCMP1507	1	+	+	+	NC_012903
Pelagophyceae	<i>Pelagomonas calceolata</i>	CCMP1756	2	-	-	-	unpublished data*
Phaeophyceae	<i>Desmarestia aculeata</i>	KU-1141		+	+	?	unpublished data*



**Table 2 Distribution of *por* genes and chloroplast-encoded LIPOR genes in rhodophytes and CASH algae (Continued)**

Phaeophyceae	<i>Ectocarpus siliculosus</i>	Ec32 (CCAP1310/4)	1	+	+	+	NC_013498
Phaeophyceae	<i>Fucus vesiculosus</i>			+	+	+	NC_016735
Phaeophyceae	<i>Nereocystis luteana</i>	UWCC MA 708		+	+	+	unpublished data*
Phaeophyceae	<i>Saccharina japonica</i>			+	+	+	NC_018523
Pinguiphyceae	<i>Pinguicoccus pyrenoidosus</i>	CCMP2188	1	+	+	+	unpublished data*
Raphidophyceae	<i>Chattonella subsalsa</i>	CCMP217	1	+	+	+	unpublished data*
Raphidophyceae	<i>Heterosigma akashiwo</i>	CCMP452	2	-	-	-	EU168191
Raphidophyceae	<i>Heterosigma akashiwo</i>	NIES293	2	-	-	-	NC_010772
Synurophyceae	<i>Synura petersenii</i>	CCMP854		-	-	-	unpublished data*
Xanthophyceae	<i>Botrydium cytosum</i>	UTEX 157		+	+	+	unpublished data*
Xanthophyceae	<i>Tribonema aequale</i>	CCMP1275	1	+	+	+	unpublished data*
Xanthophyceae	<i>Vaucheria litorea</i>	CCMP2940	1	+	+	+	NC_011600
<b>Dinophyta</b>							
Dinotrichales (dinotom)	<i>Durinskia baltica</i>	CS-38		-	-	-	NC_014287
Dinotrichales (dinotom)	<i>Kryptoperidinium foliaceum</i>	CCMP1326		-	-	-	NC_014267
<b>Chromerida</b>							
Chromeraceae	<i>Chromera velia</i>	CCMP2878		-	-	-	NC_014340
Vitrellaceae	<i>Vitrella brassicaformis</i>	CCMP3155/RM11		+, 2	+, 2	+	NC_014345

(+) Present in chloroplast genome or Fong and Archibald [46] study; (-) not present in fully-sequenced chloroplast genome; (ψ) present as pseudogene; (?) unknown; (\*) Cattolico, Rocap and McKay; (♯) *C. caldarium* re-annotated as per [6]. The number of *por* genes found in this study for a particular genus (not necessarily the same species or strain) is also indicated.

*provasoli* appears to have lost solely the *chlB* gene and some cryptophytes are documented to possess LIPOR pseudogenes, showcasing the gradual process of LIPOR gene loss ([46,94]; Table 2).

These data from chloroplast genomes do not exclude the possibility that the three genes encoding LIPOR have, in some species, been moved to the nuclear genome via endosymbiotic gene transfer. BLASTp searches of all accessible, completely sequenced algal nuclear genomes for which chloroplastic *chlL*, *chlN*, or *chlB* genes are absent did not reveal nuclear homologs to these three genes (chlorophytes *Micromonas pusilla* and *Ostreococcus tauri*; cryptophyte *Guillardia theta*; stramenopiles *Aureococcus anophagefferens*, *Fragilariopsis cylindrus*, *Phaeodactylum tricorutum*, *Thalassiosira pseudonana*; haptophytes *Chrysochromulina tobin* and *Emiliania huxleyi*).

Bayesian and maximum-likelihood phylogenetic analysis of each LIPOR gene was also performed. Adding the genes in Tables 1 and 2 to the expansive survey of Sousa et al. [54], homologs were sampled from extant phyla known to possess *chlL*, *chlN* and *chlB*: those in Tables 1 and 2, cyanobacteria, chlorobacteria, chloroflexi, proteobacteria, firmicutes and acidobacteria. Similar to previous findings [46,54], the LIPOR genes of eukaryotic algae and a particular subset of cyanobacteria formed a monophyletic group. Resolution among phyla was correlated with protein length, with phyla well resolved only by the longest protein, CHLB (404 amino

acids in alignment). Convincing evidence of horizontal transfer of any LIPOR gene was not detected for any eukaryotic alga (data not shown).

#### Maintenance of non-homologous, isofunctional enzymes

*Why are por genes seemingly ubiquitous in plant and algal genomes, whereas LIPOR genes are lost in some lineages?* It has been suggested that oxygenic photosynthesis and present-day atmospheric oxygen levels are incompatible with the oxygen-sensitive LIPOR enzyme [4,11,98]. Studies utilizing the cyanobacterium *Leptolyngbia boryana* (formerly *Plectonema boryanum*) and a *L. boryana por* knockout mutant were performed to probe this enzymatic constraint. Both the wild-type (encoding both POR and LIPOR proteins) and *por* knockout mutant grew equally well under low light intensities (10–25 μmol photons m<sup>-2</sup> s<sup>-1</sup>). However, the mutant showed depressed growth and chlorophyll synthesis at medium light intensities (85 μmol photons m<sup>-2</sup> s<sup>-1</sup>). At high light intensities (130 μmol photons m<sup>-2</sup> s<sup>-1</sup>), the *por* knockout mutant failed to grow whereas the wild-type flourished ([99]). An increased rate of photosynthesis at high light intensities causes increased oxygen production that could impede LIPOR enzyme function. In support of this reasoning, later research determined that the *por* knockout mutant could grow when oxygen was continuously removed from the growth medium, although at only two-thirds the rate of the wild-type [8]. *In vitro*

studies have identified the iron-sulfur clusters of LIPOR L-proteins as the primary targets of molecular oxygen [7,9]. The iron-sulfur cluster of the NB-proteins are much less vulnerable to oxygen [5,100,101].

Furthermore, the synthesis of an iron-requiring protein such as LIPOR may prove metabolically disadvantageous to phytoplankton living in iron-depleted regions such as the high-nutrient, low-chlorophyll regions of the subarctic and equatorial Pacific Ocean as well as the Southern Ocean [102,103]. Iron deficiencies have been shown to trigger a reduction in the synthesis of iron-rich proteins [104] and induce the substitution of functionally similar proteins that do not rely on this element, such as the replacement of ferredoxin by flavodoxin under iron-limiting conditions [105,106]. Future studies might determine whether low-iron conditions favor a switch between LIPOR and POR synthesis in algae possessing both enzymes.

*Why have some lineages maintained LIPOR genes?* Although the POR enzyme neither possesses iron moieties nor is sensitive to oxygen, light quantity and quality may affect the catalytic capacity of this enzyme. Studies in land plants have long identified that the absorption of light energy by Pchl<sub>id</sub> enables POR to catalyze its conversion [107]. The Pchl<sub>id</sub> pigment has absorbance maxima in both red and blue regions of the light spectrum [108]. Recently, Hanf et al. [109] showed that the photoconversion of Pchl<sub>id</sub> to Chl<sub>id</sub> by the POR enzyme was three to seven times as efficient when Pchl<sub>id</sub> absorbed red light (647 nm) rather than blue light (407 nm; though their choice of blue excitation wavelength for this experiment was controversial [110]). Due to its long wavelengths and concomitant lower energy, red light is attenuated rapidly from the water column, whereas green and especially blue light penetrates deeper. It is therefore possible that in deep or turbid waters or during an algal bloom, the POR enzyme may not efficiently enable chlorophyll production whereas the enzymatic ability of the LIPOR enzyme would not be expected to decrease under these conditions. In addition to enabling greening in the dark and low light, LIPOR would then also enable greening under red-light limited conditions. Future physiological experiments are warranted to explore whether a wavelength bias of the POR enzyme exists and, if so, to determine whether LIPOR provides a compensatory advantage.

*Could a por gene duplication compensate for a loss of LIPOR genes?* Interestingly, Tables 1 and 2 document a potential association between the loss of genes encoding LIPOR and the presence of duplicated *por* genes in both the haptophytes and stramenopiles (Tables 1 and 2). All five sequenced chloroplast genomes of haptophytes lack LIPOR genes, and 20 out of 22 sampled haptophytes maintain multiple *por* genes (stramenopile/haptophyte *por1* and *por2* genes, occasionally multiples copies of

one paralog; Additional file 4). In those stramenopiles for which LIPOR and *por* gene complements are known, those species that lack LIPOR genes maintain multiple *por* genes (one stramenopile/haptophyte *por1* gene and one stramenopile/haptophyte *por2* gene; Additional file 4). The pattern of duplicated *por* genes in the absence of the isofunctional LIPOR enzyme is maintained even within taxonomic class. Within the Pelagophyceae, *Aureococcus anophagefferens* and *Pelagomonas calceolata* both lack LIPOR and each possesses two *pors*, whereas *Aureoumbra lagunensis* possesses LIPOR genes and maintains just one *por* gene. Similarly, within the Raphidophyceae, *Heterosigma akashiwo* lacks LIPOR but maintains two *pors* whereas *Chattonella subsalsala* maintains LIPOR genes and possesses just one *por* gene. The chlorarachniophyte *Bigeloviella natans* and the euglenid *Eutreptiella gymnastica* lack LIPOR genes, and both maintain multiple *por* genes. In contrast, two chlorophytes (*Micromonas pusilla* and *Ostreococcus tauri*), four rhodophytes (*Chondrus crispus*, *Gracilaria tenuistipitata* and *salicornia*, and *Grateloupia taiwanensis*) and two cryptophytes (*Guillardia theta* and some *Rhodomonas* spp.) lack LIPOR genes but possess just one *por* gene. A *por* gene duplication has not been documented, however, in these three taxa. A possible relationship between the maintenance of *por* gene duplicates and the loss of the LIPOR enzyme should be clarified as more algal genomes are sequenced.

Given that chlorophyll is only used in the light, the forestalling of chlorophyll synthesis due to the light-dependency of the POR enzyme may not prove problematic. For example, as in the etiolated seedlings of angiosperms discussed above, a dark-adapted alga that lacks LIPOR might accumulate POR enzymes complexed with Pchl<sub>id</sub> substrate and therefore be poised to produce large quantities of chlorophyll upon illumination. Our preliminary data also suggests that the capacity to differentially regulate *por* genes may be critical to algal cells as they progress through an alternate life history phase where light plays a seminal role. In transcriptomes developed from samples of the harmful-bloom forming alga *Heterosigma akashiwo* (Stramenopila; Raphidophyceae) which lacks LIPOR genes, *por1* transcript abundance predominates in light-grown vegetative cells, whereas *por2* appears to be highly up-regulated when resting phase cells are maintained in the cold and dark. Though hypothetical, these data suggest that stockpiling POR2 proteins may enable rapid chlorophyll synthesis upon the re-activation of resting cells initiated by light ([111,112]; Deodato and Cattolico, unpub.). These preliminary data merit rigorous study to determine if algae lacking LIPOR genes but possessing multiple *por* genes utilize one *por* gene copy to enable swift chlorophyll production upon a return to light.

## Conclusions

This study identifies conserved *por* gene duplications in: (a) dinoflagellates, (b) chlorarachniophytes, as well as (c) stramenopiles and haptophytes. These three *por* gene expansions offer a unique opportunity to study whether and how expanded gene sets with independent origins converge on similar regulatory schemes among evolutionarily divergent taxa. Even within the shared stramenopile and haptophyte *por* gene family, the ancient divergence of these two taxa may mean that they use their *por* gene sets differently—especially for those stramenopiles maintaining the LIPOR enzyme rather than multiple *por* genes. Given the loss of LIPOR genes from many species in various taxa, future studies are also warranted to clarify possible advantages of maintaining the LIPOR enzyme and whether iron limitation affects LIPOR synthesis.

The *por* gene duplicates of stramenopiles and haptophytes appear to arise from a horizontal gene transfer from a prasinophytic (chlorophytic) alga early in the evolution of the stramenopiles. The derived position of even basal haptophytes in comparison to stramenopiles evidences a possible gene transfer from the stramenopiles to the haptophytes, whether via EGT or HGT. Our data suggest that a thorough phylogenetic examination of chloroplast-targeted genes originally existing as single copies and shared among CASH lineages (e.g., *por*) may be a boon to the determination of CASH plastid relationships. The recent surge of publically available genomic and transcriptomic datasets should be mined for such informative genes [61].

## Methods

### *por* gene recovery

The following sources were used to retrieve POR genes for use in phylogenetic studies: (a) *public and private datasets*: *por* genes were identified by blast searching against public databases; in-house databases compiled from publically available genomes and transcriptomes, as well as the *Chrysochromulina tobin* CCMP291<sub>RAC</sub> genome (Additional file 4). Transcriptomes reported to be derived from co-cultures (e.g., predator-prey experiments) were excluded from the analyses, although bacterized cultures were permitted because the *por* gene is not expected in non-photosynthetic organisms. (b) *algal samples*: Genomic DNA was extracted from algal cell pellets using Genomic-tip 500/G and 100/G DNA extraction kits (Qiagen, Valencia, CA) and targeted genes were recovered by PCR amplification and sequencing (Additional file 5). Degenerate primers were designed to universally amplify *por* sequence from diverse algal taxa (Additional file 5). Conserved protein regions for primer design were identified by aligning POR proteins from diverse algal taxa with the MUSCLE sequence alignment software (Edgar 2004).

Degenerate primers were flanked with 23 bp of additional, non-degenerate nucleotides for ease of sequencing. POR genes were amplified in 25  $\mu$ L reactions containing 0.1U/ $\mu$ L Lamda Biotech Tsg Plus DNA Polymerase (St. Louis, MO), 1X Tsg Plus reaction buffer, 0.2 mM dNTPs, 1.25 mM MgCl<sub>2</sub>, 1 ng/ $\mu$ L gDNA, and 1.2  $\mu$ M each primer, with the addition of CES PCR additive when amplification proved problematic (described in Ralser et al. [113]). Cycling reactions were performed in an Eppendorf Mastercycler gradient thermocycler as follows: initial denaturation was at 94°C for 4 min; followed by 40 cycles of 30s denaturation at 94°C; 30s annealing at 50°C–58°C (gradient); a 2 min extension at 72°C; then 10 min final elongation at 72°C. When the only successful gene amplification for a given species occurred with an internal degenerate primer (i.e., not the degenerate primers closest to the 5' or 3' ends of the gene), a species-specific primer was designed ~200 bp from the appropriate sequence end and PCR was repeated with this new primer and the degenerate primer closest to the desired gene end.

When multiple bands or primer dimers were present in a PCR product, the desired band was gel extracted from a 1% agar Tris-Acetate-EDTA (TAE) gel stained with ethidium bromide. Gel extraction was performed using the QIAquick gel extraction kit (Qiagen). When sequencing yielded multiple products, the gene was re-amplified and extracted from a TAE gel stained with SeqJack GreenGene nucleic acid stain as per manufacturer's recommendations (Mt. Baker Bio, Everett, WA) and visualized with blue light rather than UV light to retain DNA integrity. The extracted PCR product was cloned for re-sequencing using the TOPO TA cloning kit following manufacturer's directions (Invitrogen, Carlsbad, CA). All sequencing was performed on an ABI 3130xl Genetic Analyzer using the ABI BigDye Terminator v3.1 Cycle Sequencing kit with 1/8th the manufacturer's recommended reaction size (Applied BioSystems, Inc., Foster City, CA). pGEX primers (flanking the degenerate primer), species-specific internal primers, or M13 primers (cloned products) were used in the sequencing reactions.

When necessary, cDNA sequences were deduced from intron-containing gene sequences using GenomeScan [114–116], or by alignment with known POR protein sequences.

### POR protein curation

BLASTp searches for POR proteins returned many homologs, likely reflecting their origins in the conserved SDR (short-chain dehydrogenase-reductase) protein family [47,48]. Mutagenic studies of cyanobacterial and plant *por* genes have revealed several essential features of POR proteins: (a) the N-terminal Rossman fold (Gly-X-X-X-Gly-X-GLY) that is essential to NADPH binding ([117];

Figure 2a), (b) the Try-X-X-X-Lys array that stabilizes the enzyme-cofactor-substrate complex and whose Tyr donates a proton to Pchl<sub>id</sub> during the enzymatic reaction ([49–52]; Figure 2b), as well as (c) the cysteine residue determined to be essential to POR enzyme catalysis by Menon et al. ([53]; Figure 2c). Putative POR proteins were aligned with MUSCLE [118] and omitted if they lacked these diagnostic motifs. Sequences missing their N-termini (e.g., transcriptomic sequences) were not eliminated for lacking the N-terminal Rossman motif. Duplicate, short, and low-quality transcriptomic sequences (those with many undetermined amino acids) were removed.

### Phylogenetic inference

Curated POR protein sequences were aligned with MUSCLE [118] and trimmed to remove gaps and ambiguously aligned regions, resulting in a 274 amino acid alignment of 275 sequences representing 162 taxa. Available protein matrices were evaluated for appropriateness using ProtTest 2.4 [119]. The WAG + I +  $\Gamma$  model of protein sequence evolution was found to best suit the data. Trees were inferred in the CIPRES Science Gateway [120] using RAxML 8.0.24 [121] with 1000 bootstraps, as well as MrBayes 3.2.2 [122] with two runs each of four chains, 10,000,000 generations and 25% burn-in. The Whelan and Goldman matrix for globular proteins (WAG, [55]), a Gamma shape parameter, and an empirical estimation of invariable sites was used for both the Bayesian and maximum-likelihood analyses. Stationarity and convergence of the Bayesian analysis were assessed with Tracer v1.5 [123]. Maximum likelihood and Bayesian methods recovered nearly identical topologies (see Figures 3B, 4). Trees were visualized in FigTree [124].

### Multiple sequence alignment (MSA) sequence logo construction

Cyanobacterial and diatom POR protein sequences used in the POR gene tree were aligned with MUSCLE [118] and incomplete sequences were removed, resulting in an alignment of 18 cyanobacterial sequences (18 genera) and 21–22 diatom sequences (15–16 genera) (Additional file 4). The alignment was trimmed to the N-terminus of the cyanobacterial POR proteins and gaps pertaining to two or fewer sequences were removed. Numbering is in accordance with the cyanobacterium *Synechocystis elongatus* (YP\_401520) and the diatom *Phaeodactylum tri-cornutum* (XP\_002179689; XP\_002180992).

### Availability of supporting data

Gene sequences obtained in course of this study have been deposited in GenBank under accessions KJ408437–45. The data sets supporting the results of this article

are available in the Dryad repository at <http://dx.doi.org/10.5061/dryad.3ss6p> [125].

### Additional files

**Additional file 1: POR protein alignment for phylogenetic inference.** FASTA format (.fa).

**Additional file 2: Bayesian phylogenetic tree.** Nexus file (.nex), formatted for FigTree [124].

**Additional file 3: Maximum-likelihood tree.** Newick format (.tre).

**Additional file 4: Complete list of POR protein sequences.** Excel spreadsheet (.xlsx) of identifying information of POR protein sequences used, organized by taxon and clade in the POR phylogeny.

**Additional file 5: Sequencing primers.** Excel spreadsheet (.xlsx) listing degenerate and species-specific primers used to amplify and sequence *por* genes.

### Abbreviations

CASH: Assemblage comprising Cryptophyte, Alveolate (dinoflagellate), Stramenopile and Haptophyte algae; EGT: Endosymbiotic gene transfer; HGT: Horizontal gene transfer; LI-POR: Light-independent protochlorophyllide oxidoreductase; Pchl<sub>id</sub>: Protochlorophyllide; POR: Light-dependent protochlorophyllide oxidoreductase; UTC: Ulvophyte-trebouxiophyte-chlorophyte.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

HMH and RAC conceived the study. HMH and TR collected and analyzed the data. HMH and RAC wrote the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

We thank J. Collèn and C. Boyen for the *Chondrus crispus por* gene sequence, S. Pierce and J. Schwartz for the *Vaucheria litorea por* gene sequence, and R. A. Andersen for algal cell pellets. We also thank G. Rocap, M. Jacobs, and C. McKay for advance access to unpublished chloroplast genomes. HMH was supported by the NSF GRFP (DGE-0718124; DGE-1256082) and a NHGRI ITGS grant (T32 HG00035). This research was funded by a Grant In Aid of Research from the Phycological Society of America to HMH, by the US Department of Energy under contract DE-EE0003046 awarded to RAC as part of the National Alliance for Advanced Biofuels and Bioproducts, and by NOAA NA070AR4170007 to RAC.

Received: 7 October 2014 Accepted: 15 January 2015

Published online: 10 February 2015

### References

- Willows RD. Chlorophyll synthesis. In: Wise RR, Hooper JK, editors. The structure and function of plastids. Dordrecht: Springer; 2006. p. 295–313.
- Armstrong G. Greening in the dark: light-independent chlorophyll biosynthesis from anoxygenic photosynthetic bacteria to gymnosperms. *J Photochem Photobiol B Biol.* 1998;43:87–100.
- Suzuki J, Bauer C. A prokaryotic origin for light-dependent chlorophyll biosynthesis of plants. *Proc Natl Acad Sci U S A.* 1995;92:3749–53.
- Fujita Y, Bauer CE. The light-dependent protochlorophyllide reductase: a nitrogenase-like enzyme catalyzing a key reaction for greening in the dark. In: Kadish K, Smith K, Guillard R, editors. The porphyrin handbook. vol. 13. San Diego: Elsevier Science; 2003. p. 109–56.
- Muraki N, Nomata J, Ebata K, Mizoguchi T, Shiba T, Tamiaki H, et al. X-ray crystal structure of the light-independent protochlorophyllide reductase. *Nature.* 2010;465:110–4.
- Fujita Y, Bauer CE. Reconstitution of light-independent protochlorophyllide reductase from purified BchL and BchN-BchB subunits: in vitro confirmation of nitrogenase-like features of a bacteriochlorophyll biosynthesis enzyme. *J Biol Chem.* 2000;275:23583–8.



7. Nomata J, Kitashima M, Inoue K, Fujita Y. Nitrogenase Fe protein-like Fe-S cluster is conserved in L-protein (Bchl) of dark-operative protochlorophyllide reductase from *Rhodobacter capsulatus*. *FEBS Lett.* 2006;580:6151–4.
8. Yamazaki S, Nomata J, Fujita Y. Differential operation of dual protochlorophyllide reductases for chlorophyll biosynthesis in response to environmental oxygen levels in the cyanobacterium *Leptolyngbya boryana*. *Plant Physiol.* 2006;142:911–22.
9. Yamamoto H, Kurumiya S, Ohashi R, Fujita Y. Oxygen sensitivity of a nitrogenase-like protochlorophyllide reductase from the cyanobacterium *Leptolyngbya boryana*. *Plant Cell Physiol.* 2009;50:1663–73.
10. Blankenship RE. Molecular mechanisms of photosynthesis. Oxford: Blackwell Science Ltd; 2002. p. 336.
11. Reinbothe S, Reinbothe C, Apel K, Lebedev N. Evolution of chlorophyll biosynthesis—the challenge to survive photooxidation. *Cell.* 1996;86:703–5.
12. Griffiths WT, McHugh T, Blankenship RE. The light intensity dependence of protochlorophyllide photoconversion and its significance to the catalytic mechanism of protochlorophyllide reductase. *FEBS Lett.* 1996;398:235–8.
13. Margulis L. Origin of eukaryotic cells: evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian earth. New Haven: Yale University Press; 1970. p. 349.
14. Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, et al. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol.* 2005;15:1325–30.
15. Stiller JW, Hall BD. The origin of red algae: implications for plastid evolution. *Proc Natl Acad Sci U S A.* 1997;94:4520–5.
16. Stiller JW, Riley J, Hall BD. Are red algae plants? A critical evaluation of three key molecular data sets. *J Mol Evol.* 2001;52:527–39.
17. Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. The complete chloroplast genome of the chlorarachniophyte *Bigeloviella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Mol Biol Evol.* 2007;24:54–62.
18. Baurain D, Brinkmann H, Petersen J, Rodríguez-Ezpeleta N, Stechmann A, Demoulin V, et al. Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol Biol Evol.* 2010;27:1698–709.
19. Yoon HS, Hackett JD, Pinto G, Bhattacharya D. The single, ancient origin of chromist plastids. *Proc Natl Acad Sci U S A.* 2002;99:15507–12.
20. Bachvaroff TR, Sanchez Puerta MV, Delwiche CF. Chlorophyll c-containing plastid relationships based on analyses of a multigene data set with all four chromalveolate lineages. *Mol Biol Evol.* 2005;22:1772–82.
21. Khan H, Parks N, Kozera C, Curtis BA, Parsons BJ, Bowman S, et al. Plastid genome sequence of the cryptophyte alga *Rhodomonas salina* CCMP1319: lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny. *Mol Biol Evol.* 2007;24:1832–42.
22. Sanchez-Puerta MV, Bachvaroff TR, Delwiche CF. Sorting wheat from chaff in multi-gene analyses of chlorophyll c-containing plastids. *Mol Phylogenet Evol.* 2007;44:885–97.
23. Bodyl A. Do plastid-related characters support the chromalveolate hypothesis? *J Phycol.* 2005;41:712–9.
24. Sanchez-Puerta MV, Delwiche CF. A hypothesis for plastid evolution in chromalveolates. *J Phycol.* 2008;44:1097–107.
25. Bodyl A, Stiller JW, Mackiewicz P. Chromalveolate plastids: direct descent or multiple endosymbioses? *Trends Ecol Evol.* 2009;24:119–21. author reply 121–2.
26. Petersen J, Ludewig A-K, Michael V, Bunk B, Jarek M, Baurain D, et al. *Chromera velia*, endosymbioses and the rhodoplex hypothesis—plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). *Genome Biol Evol.* 2014;6:666–84.
27. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, et al. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci U S A.* 2002;99:12246–51.
28. Doolittle WF, Boucher Y, Nesbø CL, Douady CJ, Andersson JO, Roger AJ. How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philos Trans R Soc Lond B Biol Sci.* 2003;358:39–58.
29. Doolittle WF. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.* 1998;14:307–11.
30. Gogarten JP. Gene transfer: gene swapping craze reaches eukaryotes. *Curr Biol.* 2003;13:R53–4.
31. 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;62:143–57.
32. Waller RF, Patron NJ, Keeling PJ. Phylogenetic history of plastid-targeted proteins in the peridinin-containing dinoflagellate *Heterocapsa triquetra*. *Int J Syst Evol Microbiol.* 2006;56:1439–47.
33. Rogers MB, Watkins RF, Harper JT, Durnford DG, Gray MW, Keeling PJ. A complex and punctate distribution of three eukaryotic genes derived by lateral gene transfer. *BMC Evol Biol.* 2007;7:89.
34. Allen AE, Moustafa A, Montsant A, Eckert A, Kroth PG, Bowler C. Evolution and functional diversification of fructose bisphosphate aldolase genes in photosynthetic marine diatoms. *Mol Biol Evol.* 2012;29:367–79.
35. Qiu H, Price DC, Weber APM, Facchinelli F, Yoon HS, Bhattacharya D. Assessing the bacterial contribution to the plastid proteome. *Trends Plant Sci.* 2013;18:680–7.
36. Archibald JM, Rogers MB, Toop M, Ishida K-I, Keeling PJ. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigeloviella natans*. *Proc Natl Acad Sci U S A.* 2003;100:7678–83.
37. Jain R, Rivera MC, Lake JA. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A.* 1999;96:3801–6.
38. Cohen O, Gophna U, Pupko T. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol Biol Evol.* 2011;28:1481–9.
39. Rivera MC, Jain R, Moore JE, Lake JA. Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci U S A.* 1998;95:6239–44.
40. Park C, Zhang J. High expression hampers horizontal gene transfer. *Genome Biol Evol.* 2012;4:523–32.
41. Zhang J. Evolution by gene duplication: an update. *Trends Ecol Evol.* 2003;18:292–8.
42. Nymark M, Valle KC, Brembu T, Hancke K, Winge P, Andresen K, et al. An integrated analysis of molecular acclimation to high light in the marine diatom *Phaeodactylum tricornutum*. *PLoS One.* 2009;4:e7743.
43. Nymark M, Valle KC, Hancke K, Winge P, Andresen K, Johnsen G, et al. Molecular and photosynthetic responses to prolonged darkness and subsequent acclimation to re-illumination in the diatom *Phaeodactylum tricornutum*. *PLoS One.* 2013;8:e58722.
44. Joint Genome Institute: Genome Portal <http://genomeportal.jgi.doe.gov/>
45. Ong H, Wilhelm S, Gobler C, Bullerjahn G, Jacobs MA, McKay J, et al. Analyses of the complete chloroplast genome sequences of two members of the Pelagophyceae: *Aureococcus anophagefferens* CCMP1984 and *Aureumbrambra lagunensis* CCMP1507. *J Phycol.* 2010;46:602–15.
46. Fong A, Archibald JM. Evolutionary dynamics of light-independent protochlorophyllide oxidoreductase genes in the secondary plastids of cryptophyte algae. *Eukaryot Cell.* 2008;7:550–3.
47. Masuda T, Takamiya K-I. Novel insights into the enzymology, regulation and physiological functions of light-dependent protochlorophyllide oxidoreductase in angiosperms. *Photosynth Res.* 2004;81:1–29.
48. Kavanagh KL, Jörnvall H, Persson B, Oppermann U. The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell Mol Life Sci.* 2008;65:3895–906.
49. Wilks HM, Timko MP. A light-dependent complementation system for analysis of NADPH:protochlorophyllide oxidoreductase: identification and mutagenesis of two conserved residues that are essential for enzyme activity. *Proc Natl Acad Sci U S A.* 1995;92:724–8.
50. Lebedev N, Karginova O, Mclvor W, Timko MP. Tyr275 and Lys279 stabilize NADPH within the catalytic site of NADPH:protochlorophyllide oxidoreductase and are involved in the formation of the enzyme photoactive state. *Biochemistry.* 2001;40:12562–74.
51. Heyes DJ, Hunter CN. Site-directed mutagenesis of Tyr-189 and Lys-193 in NADPH: protochlorophyllide oxidoreductase from *Synechocystis*. *Biochem Soc Trans.* 2002;30:601–4.
52. Menon BRK, Waltho JP, Scrutton NS, Heyes DJ. Cryogenic and laser photoexcitation studies identify multiple roles for active site residues in the light-driven enzyme protochlorophyllide oxidoreductase. *J Biol Chem.* 2009;284:18160–6.
53. Menon BRK, Davison PA, Hunter CN, Scrutton NS, Heyes DJ. Mutagenesis alters the catalytic mechanism of the light-driven enzyme protochlorophyllide oxidoreductase. *J Biol Chem.* 2010;285:2113–9.
54. Sousa FL, Shavit-Grievink L, Allen JF, Martin WF. Chlorophyll biosynthesis gene evolution indicates photosystem gene duplication, not photosystem merger, at the origin of oxygenic photosynthesis. *Genome Biol Evol.* 2013;5:200–16.

55. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol.* 2001;18:691–9.
56. Marin B, Nowack ECM, Melkonian M. A plastid in the making: evidence for a second primary endosymbiosis. *Protist.* 2005;156:425–32.
57. Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, Weber APM, et al. Cyanophora paradoxa genome elucidates origin of photosynthesis in algae and plants. *Science.* 2012;335:843–7.
58. Kaschner M, Loeschcke A, Krause J, Minh BQ, Heck A, Endres S, et al. Discovery of the first light-dependent protochlorophyllide oxidoreductase in anoxygenic phototrophic bacteria. *Mol Microbiol.* 2014;93:1066–78.
59. Shalchian-Tabrizi K, Skånseng M, Ronquist F, Klaveness D, Bachvaroff TR, Delwiche CF, et al. Heterotachy processes in rhodophyte-derived secondhand plastid genes: implications for addressing the origin and evolution of dinoflagellate plastids. *Mol Biol Evol.* 2006;23:1504–15.
60. Yang EC, Boo GH, Kim HJ, Cho SM, Boo SM, Andersen RA, et al. Supermatrix data highlight the phylogenetic relationships of photosynthetic stramenopiles. *Protist.* 2012;163:217–31.
61. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* 2014;12:e1001889.
62. Chan CX, Soares MB, Bonaldo MF, Wisecaver JH, Hackett JD, Anderson DM, et al. Analysis of *Alexandrium tamarense* (Dinophyceae) genes reveals the complex evolutionary history of a microbial eukaryote. *J Phycol.* 2012;48:1130–42.
63. Wisecaver JH, Brosnahan ML, Hackett JD. Horizontal gene transfer is a significant driver of gene innovation in dinoflagellates. *Genome Biol Evol.* 2013;5:2368–81.
64. Imanian B, Keeling PJ. Horizontal gene transfer and redundancy of tryptophan biosynthetic enzymes in dinotoms. *Genome Biol Evol.* 2014;6:333–43.
65. Yang Y, Matsuzaki M, Takahashi F, Qu L, Nozaki H. 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.* 2014;9:e101158.
66. Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, et al. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature.* 2012;492:59–65.
67. Kawachi M, Inoué I, Honda D, Kelly CJO, Bailey JC, Bidigare RR, et al. The Pinguiphyceae classis nova, a new class of photosynthetic stramenopiles whose members produce large amounts of omega-3 fatty acids. *Phycol Res.* 2002;50:31–47.
68. García-Sandoval R. Why some clades have low bootstrap frequencies and high Bayesian posterior probabilities. *Isr J Ecol Evol.* 2014;60:41–4.
69. Frommolt R, Werner S, Paulsen H, Goss R, Wilhelm C, Zauner S, et al. Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. *Mol Biol Evol.* 2008;25:2653–67.
70. Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science.* 2009;324:1724–6.
71. Woehle C, Dagan T, Martin WF, Gould SB. Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and apicomplexa-related *Chromera velia*. *Genome Biol Evol.* 2011;3:1220–30.
72. Cuvelier ML, Allen AE, Monier A, McCrow JP, Messié M, Tringe SG, et al. Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc Natl Acad Sci U S A.* 2010;107:14679–84.
73. Burki F, Flegontov P, Oborník M, Cihlár J, Pain A, Lukes J, et al. Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. *Genome Biol Evol.* 2012;4:626–35.
74. Deschamps P, Moreira D. Reevaluating the green contribution to diatom genomes. *Genome Biol Evol.* 2012;4:683–8.
75. Liu H, Aris-Brosou S, Probert I, de Vargas C. A time line of the environmental genetics of the haptophytes. *Mol Biol Evol.* 2010;27:161–76.
76. Gillott M, Gibbs S. The cryptomonad nucleomorph: its ultrastructure and evolutionary significance. *J Phycol.* 1980;16:558–68.
77. Lane CE, Archibald JM. The eukaryotic tree of life: endosymbiosis takes its TOL. *Trends Ecol Evol.* 2008;23:268–75.
78. Green BR. After the primary endosymbiosis: an update on the chromalveolate hypothesis and the origins of algae with Chl c. *Photosynth Res.* 2011;107:103–15.
79. Stiller JW, Schreiber J, Yue J, Guo H, Ding Q, Huang J. The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat Commun.* 2014;5:5764.
80. Le Corquillé G, Pearson G, Valente M, Viegas C, Gschloessl B, Corre E, et al. Plastid genomes of two brown algae. *Ectocarpus siliculosus* and *Fucus vesiculosus*: further insights on the evolution of red-algal derived plastids. *BMC Evol Biol.* 2009;9:253.
81. Janouškovec J, Horák A, Oborník M, Lukes J, Keeling PJ. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc Natl Acad Sci U S A.* 2010;107:10949–54.
82. Rice DW, Palmer JD. An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *BMC Biol.* 2006;4:31.
83. Hovde BT, Starckenburg SR, Hunsperger HM, Mercer LD, Deodato CR, Jha RK, et al. The mitochondrial and chloroplast genomes of the haptophyte *Chrysochromulina tobin* contain unique repeat structures and gene profiles. *BMC Genomics.* 2014;15:604.
84. Medlin L, Kooistra W, Potter D, Saunders G, Andersen R. Phylogenetic relationships of the “golden algae” (haptophytes, heterokont chromophytes) and their plastids. *Plant Syst Evol.* 1997;11:187–219.
85. Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol Biol Evol.* 2004;21:809–18.
86. Berney C, Pawlowski J. A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc Biol Sci.* 2006;273:1867–72.
87. Belyaeva OB, Litvin FF. Photoactive pigment—enzyme complexes of chlorophyll precursor in plant leaves. *Biochem.* 2007;72:1458–77.
88. Matsumoto T, Shinozaki F, Chikuni T, Yabuki A, Takishita K, Kawachi M, et al. Green-colored plastids in the dinoflagellate genus *Lepidodinium* are of core chlorophyte origin. *Protist.* 2011;162:268–76.
89. Wisecaver JH, Hackett JD. Transcriptome analysis reveals nuclear-encoded proteins for the maintenance of temporary plastids in the dinoflagellate *Dinophysis acuminata*. *BMC Genomics.* 2010;11:366.
90. Hackett JD, Anderson DM, Erdner DL, Bhattacharya D. Dinoflagellates: a remarkable evolutionary experiment. *Am J Bot.* 2004;91:1523–34.
91. Patron NJ, Waller RF, Keeling PJ. A tertiary plastid uses genes from two endosymbionts. *J Mol Biol.* 2006;357:1373–82.
92. Yoon HS, Hackett JD, Bhattacharya D. A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proc Natl Acad Sci U S A.* 2002;99:11724–9.
93. Takahashi F, Okabe Y, Nakada T, Sekimoto H, Ito M, Kataoka H, et al. Origins of the secondary plastids of Euglenophyta and Chlorarachniophyta as revealed by an analysis of the plastid-targeting, nuclear-encoded gene *psbO* 1. *J Phycol.* 2007;43:1302–9.
94. Turmel M, Gagnon M-C, O’Kelly CJ, Otis C, Lemieux C. The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Mol Biol Evol.* 2009;26:631–48.
95. Ota S, Vulot D. *Lotharella reticulosa* sp. nov.: a highly reticulated network forming chlorarachniophyte from the Mediterranean Sea. *Protist.* 2012;163:91–104.
96. Murchie EH, Horton P. Acclimation of photosynthesis to irradiance and spectral quality in British plant species: chlorophyll content, photosynthetic capacity and habitat preference. *Plant Cell Environ.* 1997;20:438–48.
97. Geider R, MacIntyre H, Kana T. Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a:carbon ratio to light, nutrient-limitation and temperature. *Mar Ecol Prog Ser.* 1997;148:187–200.
98. Schoefs B, Franck F. Protochlorophyllide reduction: mechanisms and evolution. *Photochem Photobiol.* 2003;78:543–57.
99. Fujita Y, Takagi H, Hase T. Cloning of the gene encoding a protochlorophyllide reductase: the physiological significance of the co-existence of light-dependent and -independent protochlorophyllide reduction systems in the cyanobacterium *Plectonema boryanum*. *Plant Cell Physiol.* 1998;39:177–85.
100. Nomata J, Ogawa T, Kitashima M, Inoue K, Fujita Y. NB-protein (BchN-BchB) of dark-operative protochlorophyllide reductase is the catalytic component containing oxygen-tolerant Fe-S clusters. *FEBS Lett.* 2008;582:1346–50.
101. Bröcker MJ, Schomburg S, Heinz DW, Jahn D, Schubert W-D, Moser J. Crystal structure of the nitrogenase-like dark operative protochlorophyllide oxidoreductase catalytic complex (ChlN/ChlB)<sub>2</sub>. *J Biol Chem.* 2010;285:27336–45.

102. Behrenfeld MJ, Worthington K, Sherrell RM, Chavez FP, Strutton P, McPhaden M, et al. Controls on tropical Pacific Ocean productivity revealed through nutrient stress diagnostics. *Nature*. 2006;442:1025–8.
103. Bowler C, Vardi A, Allen AE. Oceanographic and biogeochemical insights from diatom genomes. *Ann Rev Mar Sci*. 2010;2:333–65.
104. Greene RM, Geider RJ, Kolber Z, Falkowski PG. Iron-induced changes in light harvesting and photochemical energy conversion processes in eukaryotic marine algae. *Plant Physiol*. 1992;100:565–75.
105. La Roche J, Geider RJ, Graziano LM, Murray H, Lewis K. Induction of specific proteins in eukaryotic algae grown under iron-, phosphorus-, or nitrogen-deficient conditions. *J Phycol*. 1993;29:767–77.
106. La Roche J, Murray H, Orellana M, Newton J. Flavodoxin expression as an indicator of iron limitation in marine diatoms. *J Phycol*. 1995;31:520–30.
107. Heyes DJ, Hunter CN. Making light work of enzyme catalysis: protochlorophyllide oxidoreductase. *Trends Biochem Sci*. 2005;30:642–9.
108. Koski VM, Smith JHC. The isolation and spectral absorption properties of protochlorophyll from etiolated barley seedlings. *J Am Chem Soc*. 1948;70:3558–62.
109. Hanf R, Fey S, Schmitt M, Hermann G, Dietzek B, Popp J. Catalytic efficiency of a photoenzyme—an adaptation to natural light conditions. *ChemPhysChem*. 2012;13:2013–5.
110. Björn LO. Comment on “Catalytic efficiency of a photoenzyme—an adaptation to natural light conditions” by J Popp et al. *Chemphyschem*. 2013;14:2595–7. author reply 2598–2600.
111. Han M, Kim Y, Cattolico RA. *Heterosigma akashiwo* (Raphidophyceae) resting cell formation in batch culture: strain identity versus physiological response. *J Phycol*. 2002;317:304–17.
112. Tobin ED, Grünbaum D, Patterson J, Cattolico RA. Behavioral and physiological changes during benthic-pelagic transition in the harmful alga, *Heterosigma akashiwo*: potential for rapid bloom formation. *PLoS One*. 2013;8:e76663.
113. Ralsler M, Querfurth R, Warnatz H-J, Lehrach H, Yaspo M-L, Krobitsch S. An efficient and economic enhancer mix for PCR. *Biochem Biophys Res Commun*. 2006;347:747–51.
114. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol*. 1997;268:78–94.
115. Yeh R-F, Lim LP, Burge CB. Computational inference of homologous gene structures in the human genome. *Genome Res*. 2001;11:803–16.
116. GenomeScan Web Server at MIT <http://genes.mit.edu/genomescan.html>
117. Birve S, Selstam E, Johansson B. Secondary structure of NADPH: protochlorophyllide oxidoreductase examined by circular dichroism and prediction methods. *Biochem J*. 1996;317:549–55.
118. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.
119. Abascal F, Zardoya R, Posada D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*. 2005;21:2104–5.
120. CIPRES Science Gateway <http://www.phylo.org>
121. Stamatakis A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006;22:2688–90.
122. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19:1572–4.
123. Tracer <http://tree.bio.ed.ac.uk/software/tracer/>
124. FigTree <http://tree.bio.ed.ac.uk/software/figtree/>
125. Hunsperger HM, Randhawa T, Cattolico RA. Data from: extensive horizontal gene transfer, duplication, and loss of chlorophyll synthesis genes in the algae. *BMC Evol Biol* 2015.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

