Many paths to chlorophyll: the evolution of protochlorophyllide oxidoreductases in the algae

Heather Marie Hunsperger

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Rose Ann Cattolico, Chair

Richard Olmstead

Takato Imaizumi

Program Authorized to Offer Degree:

Department of Biology

University of Washington

Abstract

Many paths to chlorophyll: the evolution of protochlorophyllide oxidoreductases in the algae

Heather Marie Hunsperger

Chair of the Supervisory Committee:
Dr. Rose Ann Cattolico
Department of Biology

Cyanobacteria, eukaryotic algae and some land plants catalyze the second to last step of chlorophyll *a* synthesis with either of two non-homologous enzymes: the light-independent (LIPOR) and light-dependent (POR) protochlorophyllide oxidoreductases. Both taxa with and taxa without LIPOR genes are reported for the chlorophyte, rhodophyte, cryptophyte, heterokont, and chromerid algal lineages. Haptophyte and chlorarachniophyte algal taxa appear to lack LIPOR genes. In contrast, genes encoding POR were found in all algal nuclear genomes sequenced to date. Moreover, many algal genomes and transcriptomes encode multiple POR enzymes. Phylogenetic analysis suggests that POR gene duplications occurred early in the establishment of the dinoflagellate and chlorarachniophyte lineages. Furthermore, stramenopiles and haptophytes share POR gene duplicates obtained via horizontal gene transfer from a chlorophytic alga. Phylogenetic evidence indicates that haptophytes may have obtained these genes from the stramenopiles, either by horizontal or endosymbiotic gene transfer.

To elucidate the contribution of dual POR enzymes to chlorophyll synthesis, the regulation of both POR proteins (POR1 and POR2) and their mRNAs (*por1* and *por2*) was explored for the stramenopile alga *Phaeodactylum tricornutum*. This alga was exposed to daily light:dark cycles at various irradiances, continuous illumination, and a light stepdown. *Por*1 mRNA and POR1 protein abundances were highly responsive to changes in photoperiodicity and light intensity. Transition to continuous light altered *por*1 mRNA regulation and suppressed POR1 protein abundance. A significantly higher abundance of POR1 protein was observed following a transition from high to low light, when algae typically photoacclimate by increasing cellular chlorophyll levels. In contrast, *por*2 mRNA and POR2 protein levels maintained a marked diurnal rhythmicity despite transfer to continuous light and a shift from high to low light. The sensitivity of *por*1/POR1 to fluctuating lighting conditions and perpetual diurnal rhythmicity *por*2/POR2 suggest that diatom POR1 plays a role in photoacclimation and diatom POR2 sustains daily chlorophyll synthesis. Collectively, these data evidence *por* gene expansions in many algal taxa and provide first insights into the *por* regulatory scheme of a diatom utilizing two *por* genes to optimize chlorophyll synthesis in response to both daily photoperiod and fluctuating environmental parameters.

Table of Contents

i Abstract

01 Introduction

06 Chapter One: Extensive horizontal gene transfer, duplication, and loss of chlorophyll synthesis genes in the algae

- 07 Abstract
- 09 Introduction
- 14 Results and discussion
- 34 Conclusions
- 35 Methods
- 39 Availability of supporting data
- 40 Acknowledgements
- 41 Literature cited
- 52 Figures
- 57 Tables

62 Chapter Two: Differential regulation of duplicate light-dependent protochlorophyllide oxidoreductases in the diatom *Phaeodactylum tricornutum*

- 63 Abstract
- 65 Introduction
- 70 Results
- 82 Discussion
- 91 Conclusions

92 Methods

101 Acknowledgements

102 Literature cited

109 Figures

116 Supplementary figures

119 Supplementary table

121 Conclusions

125 Other manuscripts

125 Appendix A. Abstract from: The mitochondrial and chloroplast genomes of the haptophyte *Chrysochromulina tobin* contain unique repeat structures and gene profiles

127 Appendix B. Abstract from: Genome sequence and transcriptome analyses of *Chrysochromulina tobin*: metabolic tools for enhanced algal fitness in the prominent order Prymnesiales (Haptophyceae)

129 Appendix C. Abstract from: Cell morphology, mixotrophy and genetics of *Chrysochromulina tobin* sp. nov. (Haptophyta) - a new model system for analyzing oleaginous algae

131 Appendix D. Abstract from: Evolution and maintenance of haploid-diploid life cycles in natural populations: the case of the marine brown alga *Ectocarpus*

INTRODUCTION

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 [1, 2]. Either of these enzymes can reduce the C17=C18 bond of the D-ring of the protochlorophyll tetrapyrrole ring (Pchlide), immediately preceding the addition of a phytol tail by chlorophyll synthase to form the complete chlorophyll *a* molecule. The LIPOR enzyme is a hetero-octamer composed of subunits encoded in the chloroplast genome of eukaryotes by the *chlL, chlN,* and *chlB* genes. In contrast, the POR enzyme is a globular protein encoded in the nucleus by the *por* gene and then targeted to the chloroplast stroma, where most enzymes involved in chlorophyll synthesis catalyze their reactions.

The LIPOR enzyme arose in anoxygenic photosynthetic organisms and is present in the chlorobacteria, chloroflexi, proteobacteria, firmicutes and acidobacteria as part of the bacteriochlorophyll synthesis pathway [2, 3]. Structural analyses indicate that LIPOR likely evolved from a nitrogenase-like enzyme [4, 5] and, as such, possesses iron-sulfur clusters which confer an extreme sensitivity to oxygen [6, 7]. It has been postulated that oxygenic photosynthesis and contemporary atmospheric oxygen levels are incompatible with such an oxygen-sensitive enzyme [8, 9]. Accordingly, several studies have found that atmospheric oxygen and/or oxygen produced during photosynthesis inhibits chlorophyll production by LIPOR [9–11 and references therein].

Cyanobacteria represent the first oxygenic photosynthetic organisms and it is in this taxon that the POR enzyme first arose to compliment the oxygen-sensitive LIPOR enzyme [12]. However,

the POR ‘photoenzyme’ can only catalyze its reaction when its substrate (the chlorophyll precursor pigment Pchlide) absorbs light [13]. Thus the LIPOR enzyme may be favored in the dark, whereas the POR enzyme may be favored in the light.

Both LIPOR and POR were transferred to eukaryotic photosynthetic lineages via the primary endosymbiosis that established the chloroplasts of both rhodophytic algae and well as chlorophytic algae and land plants [14, 15]. Most genes necessary for chloroplast homeostasis, including those required for chlorophyll biosynthesis, were transferred to the host nucleus (e.g., *por*), whereas others remained in a reduced chloroplast genome (e.g., *chlL, chlN, chlB*). In the course of evolution, additional algal lineages were formed via secondary and higher order endosymbioses, each of which was followed by gene transfer to the new host nucleus. For example, the cryptophyte, stramenopile, haptophyte and dinoflagellate algal lineages are believed to have obtained their chloroplasts from the same secondary endosymbiotic uptake of a red alga [16]. This red-algal chloroplast was then transferred among these four lineages in subsequent tertiary and higher order endosymbioses whose order has not yet been ascertained [17, 18].

It is known that cyanobacteria generally possess both POR and LIPOR enzymes whereas some land plants (e.g., angiosperms) lack LIPOR and rely solely on POR [2, 19]. In fact, it has been observed that the multiple POR enzymes of angiosperms are each under a unique regulatory scheme to optimize chlorophyll synthesis under changing developmental and environmental conditions [19, 20]. The presence and quantity of LIPOR and POR enzymes encoded in eukaryotic algal genomes has not been cataloged since the recent increase in available nuclear and chloroplast genome as well as transcriptome sequence data. Some algal lineages, such as

those of the stramenopile class Bacillariophyceae (diatoms), lack LIPOR genes. Interestingly, each of three recently sequenced diatom genomes were found to encode multiple POR enzymes.

The aims of this PhD thesis were to: (a) survey the presence/absence of LIPOR genes from a burgeoning array of algal chloroplast genomes; (b) explore the evolutionary origins of the multiple diatom *por* genes in a phylogenetic context; and (c) determine whether diatoms possess regulatory schemes unique to each of their dual POR enzymes at both the levels of transcription and protein abundance. These studies will: (a) provide insight to the specific enzyme(s) that drive the penultimate step of chlorophyll synthesis in a broad range of algal taxa; (b) enhance our understanding of the fates of seemingly redundant non-homologous as well as homologous iso-functional enzymes and (c) enable comparative analyses concerning the regulation of Pchlide reduction among and between algae and land plants. These studies will be of interest to phycologists as well as botanists, physiologists as well as enzymologists, and especially to evolutionary biologists.

Works Cited

1. Suzuki JY, Bollivar DW, Bauer CE: **Genetic analysis of chlorophyll biosynthesis.** Annu Rev Genet 1997, **31**:61–89.

2. Fujita Y, Bauer CE: **The light-independent protochlorophyllide reductase: a nitrogenase-like enzyme catalyzing a key reaction for greening in the dark**. In Porphyr Handb. Volume 13. Edited by Kadish K, Smith K, Guilard R. Elsevier Science; 2003:109–156.

3. 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**:200–216.

4. Sarma R, Barney BM, Hamilton TL, Jones A, Seefeldt LC, Peters JW: **Crystal structure of the L protein of *Rhodobacter sphaeroides* light-independent protochlorophyllide reductase with MgADP bound: a homologue of the nitrogenase Fe protein.** Biochemistry 2008, **47**:13004–15.

5. Muraki N, Nomata J, Ebata K, Mizoguchi T, Shiba T, Tamiaki H, Kurisu G, Fujita Y: **X-ray crystal structure of the light-independent protochlorophyllide reductase.** Nature 2010, **465**:110–114.

6. 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**:6151–6154.

7. 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**:1663–1673.

8. Reinbothe S, Reinbothe C, Apel K, Lebedev N: **Evolution of chlorophyll biosynthesis—the challenge to survive photooxidation**. Cell 1996, **86**:703–705.

9. Schoefs B, Franck F: **Protochlorophyllide reduction: mechanisms and evolution**. Photochem Photobiol 2003, **78**:543–557.

10. 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**:177–185.

11. 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**:23583–23588.

12. Suzuki J, Bauer C: **A prokaryotic origin for light-dependent chlorophyll biosynthesis of plants**. Proc Natl Acad Sci U S A 1995, **92**:3749–3753.

13. Griffiths WT: **Characterization of the terminal stages of chlorophyll (ide) synthesis in etioplast membrane preparations.** Biochem J 1975, **152**:623–35.

14. 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. Yale University Press; 1970.

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

16. Yoon HS, Hackett JD, Pinto G, Bhattacharya D: **The single, ancient origin of chromist plastids.** Proc Natl Acad Sci U S A 2002, **99**:15507–15512.

17. Baurain D, Brinkmann H, Petersen J, Rodríguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H: **Phylogenomic evidence for separate acquisition of**

plastids in cryptophytes, haptophytes, and stramenopiles. Mol Biol Evol 2010, **27**:1698–1709.

18. 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.

19. 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.

20. Masuda T, Fusada N, Shiraishi T, Kuroda H, Awai K, Shimada H, Ohta H, Takamiya K-I: **Identification of two differentially regulated isoforms of protochlorophyllide oxidoreductase (POR) from tobacco revealed a wide variety of light- and development-dependent regulations of POR gene expression among angiosperms**. Photosynth Res 2002, **74**:165–172.

EXTENSIVE HORIZONTAL GENE TRANSFER, DUPLICATION, AND LOSS OF CHLOROPHYLL SYNTHESIS GENES IN THE ALGAE

Heather M Hunsperger[1], Tejinder Randhawa[1], Rose Ann Cattolico[1]

[1]Department of Biology, University of Washington, Seattle, WA, USA

Status: Published in BMC Evolutionary Biology 2015, **15**:16.

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 *por*s 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 compliments. 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.

INTRODUCTION

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 (Fig. 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 heterotetramer encoded by the genes *chlN* and *chlB*. Also like nitrogenase, the LIPOR holoenzyme contains iron-sulfur 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 multiseries* [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 *por*s 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 isofunctional 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, Starkenburg 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 1-3. 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 eight sampled cryptophyte species (*Chroomonas* cf. *mesostigmatica, Guillardia theta, Hanusia phi, Hemiselmis andersenii* [2 strains]*, Proteomonas sulcata, Rhodomonas abbreviata, Rhodomonas lens, Rhodomonas* sp.) and two stramenopiles (*Ectocarpus siliculosus* and *Mallomonas* sp.) demonstrate affinity with rhodophytic PORs. Previous studies identified a member of the Porphyridiales as the progenitor of all CASH plastids [59]. The fact that the phaeophycean, synurophycean, and cryptophyte POR proteins are sister to the *Porphyridium purpureum* POR protein suggests that the POR proteins found in these taxa may represent the original rhodophyte-derived enzyme. We note, however, that the *por* genes of the two stramenopiles *E. siliculosus* and *Mallomonas* sp. are not sister to one another as would be expected given the shared origin of their plastids. Barring a complex scenario of two unique HGT events from the rhodophytes to the phaeophyceans and synurophyceans, the polyphyletic placement of these sequences may be due to insufficient phylogenetic signal.

In contrast, a chlorophytic POR protein origin is detected for most stramenopiles, all sampled haptophytes, several peridinin-containing dinoflagellates, and many cryptophytes. This relationship is indicated by the emergence of their proteins within the prasinophytic POR branches (indicated by arrows in Figures 3 and 4). These data suggest that the original rhodophytic *por* gene of these lineages has been replaced (or supplemented in some cases) by a prasinophytic *por* gene obtained by HGT. The loss of the ancestral rhodophytic *por* gene is further supported by analyses of the whole genome sequences of two haptophytes and five stramenopiles. Only *por* genes of chlorophytic origin are recovered from these complete

genomes (*Emiliania huxleyii, Thalassiosira pseudonana*, *Phaeodactylum tricornutum, Fragilariopsis cylindrus, Pseudo-nitszchia multiseries* genomes: http://genome.jgi.doe.gov; *Nannochloropsis gaditana*: http://nannochloropsis.genomeprojectsolutions-databases.com; *Chrysochromulina tobin* genome: Hovde, Starkenburg and Cattolico, unpublished).

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 study that possess *por* genes of prasinophytic origin include members of SI (Raphidophyceae and Xanthophyceae), SII (Chrysophyceae, Eustigmatophyceae and Pinguiophyceae), 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*: por*1) 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 compliment 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 *Pinguiococcus* spp. and *Phaeomonas parva* (Stramenopila; Pinguiophyceae) 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, repectively). 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, 80, 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 amoebeformis* possesses one of the paralogs and represents an early diverging branch of chlorarachniophytes [95]. It is unclear whether *Amorphochlora amoebeformis, 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 Pchlide. 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 lands 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 tricornutum por1* and *por2* were found to be initially down-regulated in response to a transition from low (35 μM photons m^{-2} s^{-1}) to high (500 μM photons m^{-2} s^{-1}) light levels [42]. Our own RT-qPCR measurements of *P. tricornutum 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. tricornutum por1* and *por2* (Hovde and Cattolico, unpub).

Loss of chloroplastic genes encoding LIPOR

At least one *por* gene, encoding the light-dependent protochlorophyllide 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 protochlorophyllide 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 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 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 tricornutum, 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 incompatibile 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^{-2} s^{-1}). However, the mutant showed depressed growth and chlorophyll synthesis at medium light intensities (85 μmol photons m^{-2} s^{-1}). At high light intensities (130 μmol photons m^{-2} s^{-1}), 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 Pchlide enables POR to catalyze its conversion [107]. The Pchlide 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 Pchlide to Chlide by the POR enzyme was three to seven times as efficient when Pchlide absorbed red light (647nm) rather than blue light (407nm; 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 *por*1 and *por*2 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 *por*1 gene and one stramenopile/haptophyte *por*2 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 *por*s, whereas *Aureoumbra lagunensis* possesses LIPOR genes and maintains just one *por* gene. Similarly, within the Raphidophyceae, *Heterosigma akashiwo* lacks LIPOR but maintains two *por*s whereas *Chattonella subsalsa* maintains LIPOR genes and possesses just one *por* gene. The chlorarachniophyte *Bigelowiella 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 Pchlide 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, *por*1 transcript abundance predominates in light-grown vegetative cells, whereas *por*2 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$_{RAC}$ genome (Additional file 1). 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 2). 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 23bp of additional, non-degenerate nucleotides for ease of sequencing. POR genes were amplified in 25µL reactions containing 0.1U/µL Lamda Biotech Tsg Plus DNA polymerase (St. Louis, MO), 1X Tsg Plus reaction buffer, 0.2mM dNTPs, 1.25mM $MgCl_2$, 1ng/µL gDNA, and 1.2µM 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 4min; followed by 40 cycles of 30s denaturation at 94°C; 30s annealing at 50°C-58°C (gradient); a 2min extension at 72°C; then 10min 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 ~200bp 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 $1/8^{th}$ 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 Pchlide 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 + Γ 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 tricornutum* (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 **(Additional files 1-5)** are available in the Dryad repository at http://dx.doi.org/10.5061/dryad.3ss6p [125].

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 (.xlxs) of identifying information of POR protein sequences used, organized by taxon and clade in the POR phylogeny

Additional file 5 – Sequencing primers

Excel spreadsheet (.xlxs) listing degenerate and species-specific primers used to amplify and sequence *por* genes

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.

LITERATURE CITED

1. Willows RD: **Chlorophyll synthesis**. In Struct Funct Plast. Edited by Wise RR, Hoober JK. Springer; 2006:295–313.

2. 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.

3. Suzuki J, Bauer C: **A prokaryotic origin for light-dependent chlorophyll biosynthesis of plants**. Proc Natl Acad Sci U S A 1995, **92**:3749–3753.

4. Fujita Y, Bauer CE: **The light-dependent protochlorophyllide reductase: a nitrogenase-like enzyme catalyzing a key reaction for greening in the dark**. In Porphyr Handb. Volume 13. Edited by Kadish K, Smith K, Guilard R. Elsevier Science; 2003:109–156.

5. Muraki N, Nomata J, Ebata K, Mizoguchi T, Shiba T, Tamiaki H, Kurisu G, Fujita Y: **X-ray crystal structure of the light-independent protochlorophyllide reductase.** Nature 2010, **465**:110–114.

6. 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**:23583–23588.

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**:6151–6154.

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**:911–922.

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**:1663–1673.

10. Blankenship RE: Molecular Mechanisms of Photosynthesis. Blackwell Science Ltd; 2002:336.

11. Reinbothe S, Reinbothe C, Apel K, Lebedev N: **Evolution of chlorophyll biosynthesis—the challenge to survive photooxidation**. Cell 1996, **86**:703–705.

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**:235–238.

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. Yale University Press; 1970:349.

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

15. Stiller JW, Hall BD: **The origin of red algae: implications for plastid evolution.** Proc Natl Acad Sci U S A 1997, **94**:4520–4525.

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**:527–539.

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

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

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**:15507–15512.

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**:1772–1782.

21. Khan H, Parks N, Kozera C, Curtis B a, Parsons BJ, Bowman S, Archibald JM: **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**:1832–1842.

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**:885–897.

23. Bodył A: **Do plastid-related characters support the Chromalveolate hypothesis?** J Phycol 2005, **41**:712–719.

24. Sanchez-Puerta MV, Delwiche CF: **A hypothesis for plastid evolution in Chromalveolates**. J Phycol 2008, **44**:1097–1107.

25. Bodył 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, Brinkmann H: ***Chromera velia*, endosymbioses and the rhodoplex hypothesis—plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages).** Genome Biol Evol 2014, **6**:666–684.

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

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**:307–311.

30. Gogarten JP: **Gene transfer: gene swapping craze reaches eukaryotes.** Curr Biol 2003, **13**:R53–R54.

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**:143–157.

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**:1439–1447.

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**:367–379.

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**:680–687.

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 *Bigelowiella natans*.** Proc Natl Acad Sci U S A 2003, **100**:7678–7683.

37. Jain R, Rivera MC, Lake JA: **Horizontal gene transfer among genomes: the complexity hypothesis.** Proc Natl Acad Sci U S A 1999, **96**:3801–3806.

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**:1481–1489.

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**:6239–6244.

40. Park C, Zhang J: **High expression hampers horizontal gene transfer.** Genome Biol Evol 2012, **4**:523–532.

41. Zhang J: **Evolution by gene duplication: an update**. Trends Ecol Evol 2003, **18**:292–298.

42. Nymark M, Valle KC, Brembu T, Hancke K, Winge P, Andresen K, Johnsen G, Bones AM: **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, Bones AM, Brembu T: **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, Sims EH, Gillett WG, Zhou Y, Haugen E, Rocap G, Cattolico R: **Analyses of the complete chloroplast genome sequences of two members of the Pelagophyceae: *Aureococcus anophagefferens* CCMP1984 amd *Aureumbra lagunensis* CCMP1507**. J Phycol 2010, **46**:602–615.

46. Fong A, Archibald JM: **Evolutionary dynamics of light-independent protochlorophyllide oxidoreductase genes in the secondary plastids of cryptophyte algae.** Eukaryot Cell 2008, **7**:550–553.

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**:3895–3906.

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**:724–728.

50. Lebedev N, Karginova O, McIvor 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**:12562–12574.

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**:18160–18166.

53. Menon BRK, Davison P a, Hunter CN, Scrutton NS, Heyes DJ: **Mutagenesis alters the catalytic mechanism of the light-driven enzyme protochlorophyllide oxidoreductase.** J Biol Chem 2010, **285**:2113–2119.

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**:200–216.

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**:691–699.

56. Marin B, Nowack ECM, Melkonian M: **A plastid in the making: evidence for a second primary endosymbiosis.** Protist 2005, **156**:425–432.

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

58. Kaschner M, Loeschcke A, Krause J, Minh BQ, Heck A, Endres S, Svensson V, Wirtz A, von Haeseler A, Jaeger K-E, Drepper T, Krauss U: **Discovery of the first light-dependent protochlorophyllide oxidoreductase in anoxygenic phototrophic bacteria.** Mol Microbiol 2014, **93**:1066–1078.

59. Shalchian-Tabrizi K, Skånseng M, Ronquist F, Klaveness D, Bachvaroff TR, Delwiche CF, Botnen A, Tengs T, Jakobsen KS: **Heterotachy processes in rhodophyte-derived secondhand plastid genes: implications for addressing the origin and evolution of dinoflagellate plastids.** Mol Biol Evol 2006, **23**:1504–1515.

60. Yang EC, Boo GH, Kim HJ, Cho SM, Boo SM, Andersen R a, Yoon HS: **Supermatrix data highlight the phylogenetic relationships of photosynthetic stramenopiles.** Protist 2012, **163**:217–231.

61. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler L a, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, Beszteri B, Bidle KD, Cameron CT, Campbell L, Caron D a, Cattolico RA, Collier JL, Coyne K, Davy SK, Deschamps P, Dyhrman ST, Edvardsen B, Gates RD, Gobler CJ, Greenwood SJ, Guida SM, Jacobi JL, Jakobsen KS, James ER, Jenkins B, 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, Erdner DL, Bhattacharya D: **Analysis of *Alexandrium tamarense* (Dinophyceae) genes reveals the complex evolutionary history of a microbial eukaryote.** J Phycol 2012, **48**:1130–1142.

63. Wisecaver JH, Brosnahan ML, Hackett JD: **Horizontal gene transfer is a significant driver of gene innovation in dinoflagellates.** Genome Biol Evol 2013, **5**:2368–2381.

64. Imanian B, Keeling PJ: **Horizontal gene transfer and redundancy of tryptophan biosynthetic enzymes in dinotoms.** Genome Biol Evol 2014, **6**:333–343.

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 B a, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, Hopkins JF, Kuo A, Rensing S a, Schmutz J, Symeonidi A, Elias M, Eveleigh RJM, Herman EK, Klute MJ, Nakayama T, Oborník M, Reyes-Prieto A, Armbrust EV, Aves SJ, Beiko RG, Coutinho P, Dacks JB, Durnford DG, Fast NM, Green BR, et al.: **Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs.** Nature 2012, **492**:59–65.

67. Kawachi M, Inouye I, Honda D, Kelly CJO, Bailey JC, Bidigare RR, Andersen RA: **The Pinguiophyceae 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**:41–44.

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

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**:1724–1726.

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, Woyke T, Welsh RM, Ishoey T, Lee J-H, Binder BJ, DuPont CL, Latasa M, Guigand C, Buck KR, Hilton J, Thiagarajan M, Caler E, Read B, Lasken RS, Chavez FP, Worden AZ: **Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton.** Proc Natl Acad Sci U S A 2010, **107**:14679–14684.

73. Burki F, Flegontov P, Oborník M, Cihlár J, Pain A, Lukes J, Keeling PJ: **Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin.** Genome Biol Evol 2012, **4**:626–635.

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

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**:161–176.

76. Gillott M, Gibbs S: **The cryptomonad nucleomorph: its ultrastructure and evolutionary significance**. J Phycol 1980, **16**:558–568.

77. Lane CE, Archibald JM: **The eukaryotic tree of life: endosymbiosis takes its TOL.** Trends Ecol Evol 2008, **23**:268–275.

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**:103–115.

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 Corguillé G, Pearson G, Valente M, Viegas C, Gschloessl B, Corre E, Bailly X, Peters AF, Jubin C, Vacherie B, Cock JM, Leblanc C: **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**:10949–10954.

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, Starkenburg SR, Hunsperger HM, Mercer LD, Deodato CR, Jha RK, Chertkov O, Monnat RJ, Cattolico RA: **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**:809–818.

86. Berney C, Pawlowski J: **A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record.** Proc Biol Sci 2006, **273**:1867–1872.

87. Belyaeva OB, Litvin FF: **Photoactive pigment—enzyme complexes of chlorophyll precursor in plant leaves**. Biochem 2007, **72**:1458–1477.

88. Matsumoto T, Shinozaki F, Chikuni T, Yabuki A, Takishita K, Kawachi M, Nakayama T, Inouye I, Hashimoto T, Inagaki Y: **Green-colored plastids in the dinoflagellate genus *Lepidodinium* are of core chlorophyte origin.** Protist 2011, **162**:268–276.

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**:1523–1534.

91. Patron NJ, Waller RF, Keeling PJ: **A tertiary plastid uses genes from two endosymbionts.** J Mol Biol 2006, **357**:1373–1382.

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**:11724–11729.

93. Takahashi F, Okabe Y, Nakada T, Sekimoto H, Ito M, Kataoka H, Nozaki H: **Origins of the secondary plastids of Euglenophyta and Chlorarachniophyta as revealed by an analysis of the plastid-targeting, nuclear-encoded gene psbO 1**. J Phycol 2007, **43**:1302–1309.

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**:631–648.

95. Ota S, Vaulot 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**:438–448.

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**:543–557.

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**:177–185.

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**:1346–1350.

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)2.** J Biol Chem 2010, **285**:27336–27345.

102. Behrenfeld MJ, Worthington K, Sherrell RM, Chavez FP, Strutton P, McPhaden M, Shea DM: **Controls on tropical Pacific Ocean productivity revealed through nutrient stress diagnostics.** Nature 2006, **442**:1025–1028.

103. Bowler C, Vardi A, Allen AE: **Oceanographic and biogeochemical insights from diatom genomes**. Ann Rev Mar Sci 2010, **2**:333–365.

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**:565–575.

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**:767–777.

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**:520–530.

107. Heyes DJ, Hunter CN: **Making light work of enzyme catalysis: protochlorophyllide oxidoreductase.** Trends Biochem Sci 2005, **30**:642–649.

108. Koski VM, Smith JHC: **The isolation and spectral absorption properties of protochlorophyll from etiolated barley seedlings**. J Am Chem Soc 1948, **70**:3558–3562.

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**:2013–2015.

110. Björn LO: **Comment on "Catalytic efficiency of a photoenzyme--an adaptation to natural light conditions" by J. Popp et al.** Chemphyschem 2013, **14**:2595–2597; 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**:304–317.

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. Ralser 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**:747–751.

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**:803–816.

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**:549–555.

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

119. Abascal F, Zardoya R, Posada D: **ProtTest: selection of best-fit models of protein evolution.** Bioinformatics 2005, **21**:2104–2105.

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**:2688–2690.

122. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models**. Bioinformatics 2003, **19**:1572–1574.

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.

FIGURES

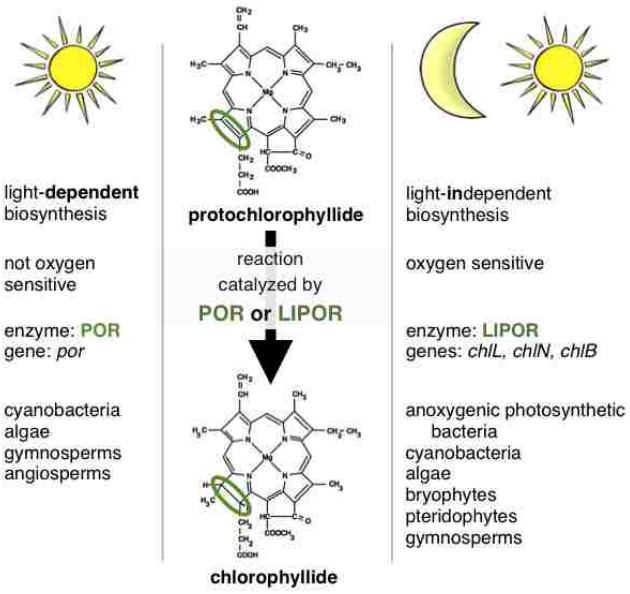


Figure 1 – Comparison of the POR and LIPOR enzymes
The second to last step of chlorophyll synthesis can be catalyzed by either a light-dependent (POR) or light-independent (LIPOR) protochlorophyllide oxidoreductase (figure after [2, 3]).

Figure 2 – Sequence logos of cyanobacterial PORs and diatom POR1 and POR2 proteins
Alignment of sequence logos of cyanobacterial POR proteins with diatom POR1 and POR2 proteins. Amino acid position indicated at the right of each line, corresponding to cyanobacterium *Synechocystis elongatus* and diatom *Phaeodactylum tricornutum*. Boxes indicate characteristic motifs, with diagnostic amino acids marked with asterisks: (a) Rossman fold essential to NADPH binding; (b) Y, K residues essential to enzyme-cofactor-substrate coordination and proton donation; (c) cysteine essential to catalysis. Amino acids are colored according to their chemical properties: green are polar (GSTYC); purple are neutral (QN), blue are positively charged (KRH); red are negatively charged (DE); and black are hydrophobic (AVLIPWFM).

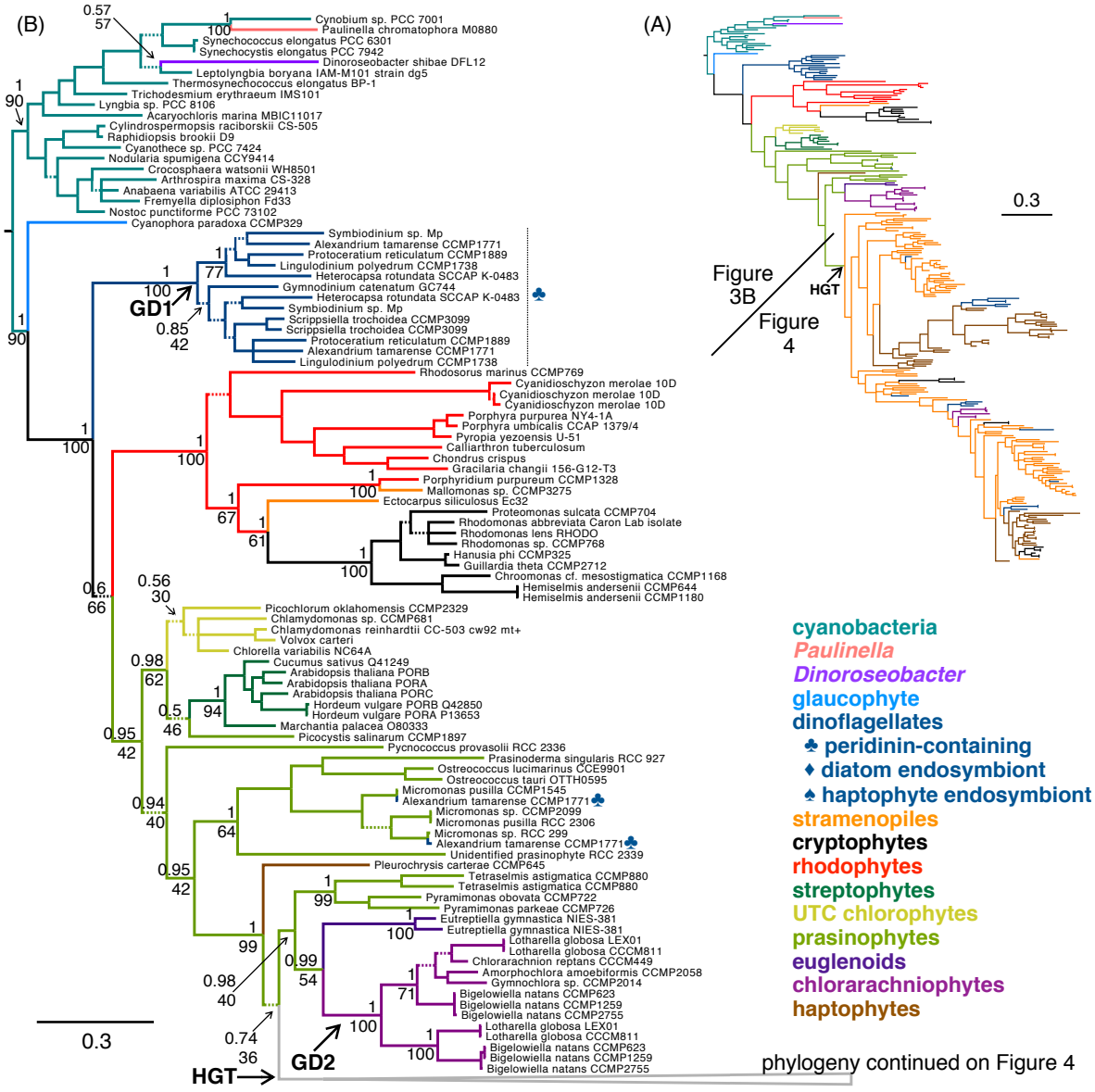


Figure 3 – *por* gene tree: rhodophytic identity of cryptophyte and some stramenopile *por*s; duplication of dinoflagellate and chlorarachniophyte *por*s

(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 Figure 3B and Figure 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.

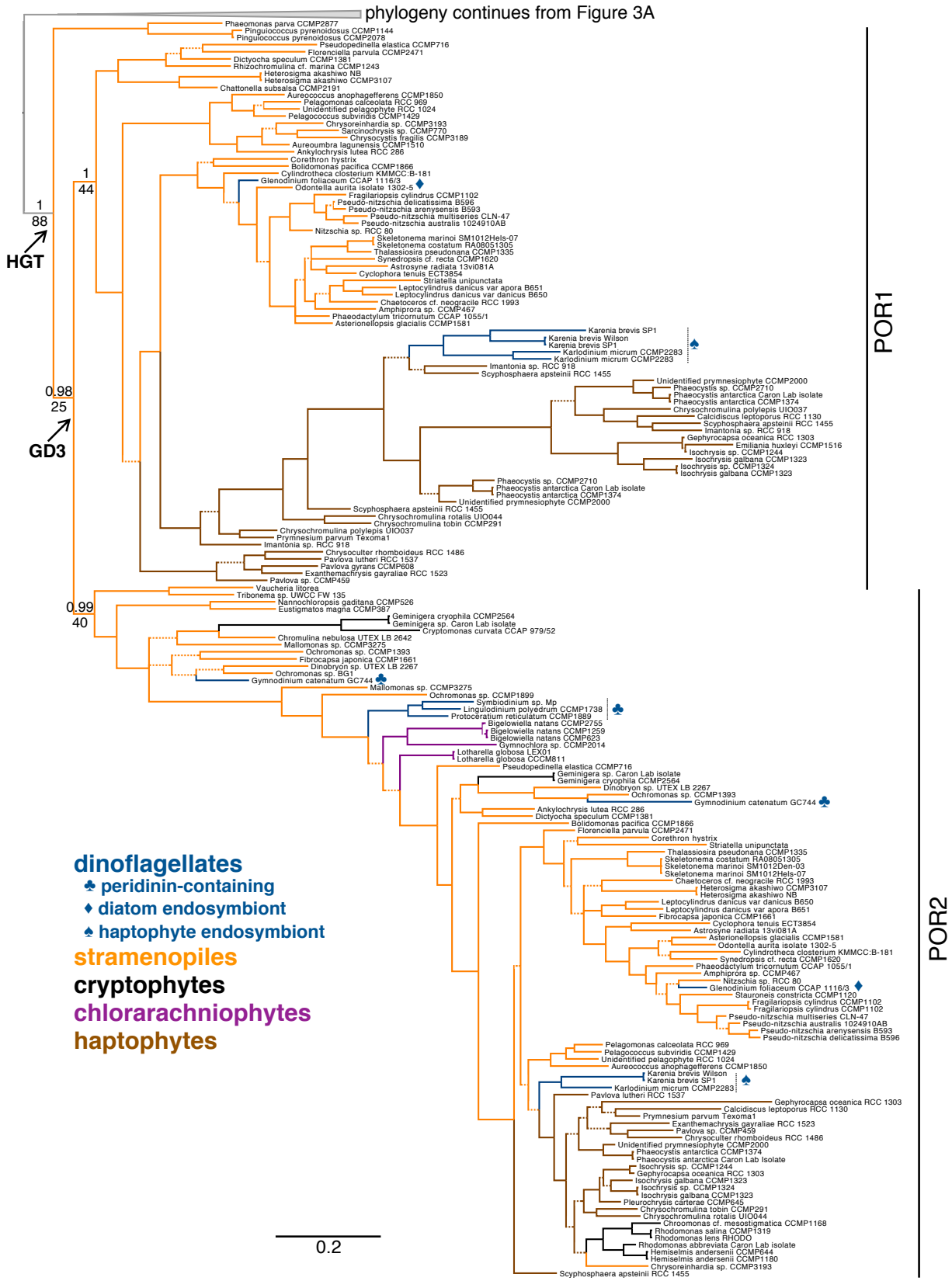


Figure 4 – Legend on next page.

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 *por*1 and *por*2.

TABLES

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

taxon	species	culture ID	*por* genes in genus (this paper)	chloroplast-encoded LIPOR genes			chloroplast genome accession
				chlL	*chlN*	*chlB*	
Chlorophyta							
Chlorophyceae	*Acutodesmus obliquus*	UTEX 393		+	+	+	NC_008101
Chlorophyceae	*Chlamydomonas reinhardtii*	n/a	1	+	+	+	NC_005353
Chlorophyceae	*Dunaliella salina*	CCAP 19/18		+	+	+	NC_016732
Chlorophyceae	*Floydiella terrestris*	UTEX 1709		+	+	+	NC_014346
Chlorophyceae	*Gonium pectorale*	K3-F3-4		+	+	+	NC_020438
Chlorophyceae	*Oedogonium cardiacum*	SAG 575-1b		+	+	+	NC_011031
Chlorophyceae	*Pleodorina starrii*	NIES 1363		+	+	+	NC_021109
Chlorophyceae	*Schizomeris leibleinii*	UTEX LB 1228		+	+	+	NC_015645
Chlorophyceae	*Stigeoclonium helveticum*	UTEX 441		+	+	+	NC_008372
Mamiellophyceae	*Micromonas pusilla*	RCC299	1	-	-	-	NC_012575
Mamiellophyceae	*Monomastix sp.*	OKE-1		-	-	-	NC_012101
Mamiellophyceae	*Ostreococcus tauri*	OTTH0595	1	-	-	-	NC_008289
Nephroselmidophyceae	*Nephroselmis olivacea*	NIES 484		+,2	+,2	+,2	NC_000927
Prasinophyceae	*Pycnococcus provasolii*	CCMP1203	1	+	+	-	NC_012097
Prasinophyceae	*Pyramimonas parkeae*	CCMP726	1	+,2	+,2	+	NC_012099
Trebouxiophyceae	*Chlorella variabilis*	NC64A	1	+	+	+	NC_015359
Trebouxiophyceae	*Chlorella vulgaris*	C-27	1	+	+	+	NC_001865
Trebouxiophyceae	*Coccomyxa subellipsoidea*	C-169		+	+	+	NC_015084
Trebouxiophyceae	*Leptosira terrestris*	UTEX 333		+	+	+	NC_009681
Trebouxiophyceae	*Parachlorella kessleri*	SAG 211/11g		+	+	+	NC_012978
Trebouxiophyceae	*Pedinomonas minor*	UTEX LB 1350		-	-	-	NC_016733
Trebouxiophyceae	*Trebouxiophyceae sp.*	MX-AZ01		+	+	+	NC_018569
Ulvophyceae	*Bryopsis hypnoides*	n/a		+	+	+	NC_013359
Ulvophyceae	*Oltmannsiellopsis viridis*	NIES 360		+	+	+	NC_008099
Ulvophyceae	*Pseudendoclonium akinetum*	UTEX 1912		-	-	-	NC_008114

Cercozoa							
Chlorarachniophyceae	*Bigelowiella natans*	CCMP621	3	-	-	-	NC_008408
Chlorarachniophyceae	*Lotharella oceanica*	CCMP622	3	-	-	-	KF438023
Euglenozoa							
Euglenophyceae	*Euglena gracilis*	Z		-	-	-	NC_001603
Euglenophyceae	*Euglena viridis*	ATCC PRA-110		-	-	-	NC_020460
Euglenophyceae	*Eutreptiella gymnastica*	K-0333	2	-	-	-	NC_017754
Euglenophyceae	*Monomorphina aenigmatica*	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.

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
				chlL	*chlN*	*chlB*	
Rhodophyta							
Bangiophyceae	*Cyanidioschyzon merolae*	strain 10D	3	-	-	-	NC_004799
Bangiophyceae	*Cyanidium caldarium*	RK1		-	-$^{\Delta}$	-	NC_001840
Bangiophyceae	*Porphyra purpurea*	Avonport	1	+	+	+	NC_000925
Bangiophyceae	*Pyropia haitanensis*	PH-38 (voucher)	1	+	+	+	NC_021189
Bangiophyceae	*Pyropia yezoensis*	U-51	1	+	+	+	NC_007932
Florideophyceae	*Calliarthron tuberculosum*		1	+	+	+	NC_021075
Florideophyceae	*Chondrus crispus*		1	-	-	-	NC_020795
Florideophyceae	*Gracilaria tenuistipitata var. liui*		1	-	-	-	NC_006137
Florideophyceae	*Gracilaria salicornia*	ARS08332 (voucher)	1	-	-	-	KF861575
Halymeniaceae	*Grateloupia taiwanensis*			-	-	-	NC_021618
Porphyridiophyceae	*Porphyridium purpureum*	NIES 2140	1	-	-	-	AP012987
Cryptophyta							
Chroomonadaceae	*Chroomonas mesostigmatica*	CCMP1168	2	+	ψ	?	EU233753; EU233756
Chroomonadaceae	*Chroomonas pauciplastida*	CCMP268	2	+	+	+	EU233754; EU233755; EU233748
Chroomonadaceae	*Hemiselmis andersenii*	CCMP644	2	+	+	+	EU233749; EU233750; EU233747
Chroomonadaceae	*Hemiselmis tepida*	CCMP443	2	+	+	?	EU233751; EU233752
Geminigeraceae	*Guillardia theta*		1	-	-	-	NC_000926
Pyrenomonadaceae	*Rhodomonas salina*	CCMP1319	1-2	ψ	ψ	ψ	NC_009573
Haptophyta							
Isochrysidales	*Emiliania huxleyi*	CCMP373	1	-	-	-	NC_007288
Pavlovales	*Pavlova lutheri*	ATCC 50092	1-2	-	-	-	NC_020371

Phaeocystales	*Phaeocystis antarctica*	CCMP1374	2-3	-	-	-	NC_016703
Phaeocystales	*Phaeocystis globosa*	Pg-G(A)	2-3	-	-	-	NC_021637
Prymnesiales	*Chrysochromulina tobin*	CCMP291	2	-	-	-	KJ201907
Stramenopila							
Bacillariophyceae	*Fistulifera sp.*	JPCC DA0580		-	-	-	NC_015403
Bacillariophyceae	*Odontella sinensis*		2	-	-	-	NC_001713
Bacillariophyceae	*Phaeodactylum tricornutum*	CCAP1055/1	2	-	-	-	NC_008588
Bacillariophyceae	*Synedra acus*			-	-	-	NC_016731
Bacillariophyceae	*Thalassiosira oceanica*	CCMP 1005	2	-	-	-	NC_014808
Bacillariophyceae	*Thalassiosira pseudonana*	CCMP 1335	2	-	-	-	NC_008589
Dictyochophyceae	*Apedinella radians*	CCMP1767		-	-	-	unpublished data*
Dictyochophyceae	*Rhizochromulina marina*	CCAP950/1	1	+	+	+	unpublished data*
Eustimatophyceae	*Nannochloropsis gaditana*	CCMP526	1	+	+	+	KJ410682
Eustimatophyceae	*Nannochloropsis oceanica*	LAMB0001	1	+	+	+	KJ410683
Eustimatophyceae	*Nannochloropsis oculata*	CCMP525	1	+	+	+	KJ410684
Eustimatophyceae	*Nannochloropsis salina*	CCMP1776	1	+	+	+	KJ410685
Pelagophyceae	*Aureococcus anophagefferens*	CCMP1984	2	-	-	-	NC_012898
Pelagophyceae	*Aureoumbra lagunensis*	CCMP1507	1	+	+	+	NC_012903
Pelagophyceae	*Pelagomonas calceolata*	CCMP1756	2	-	-	-	unpublished data*
Phaeophyceae	*Desmarestia aculeata*	KU-1141		+	+	?	unpublished data*
Phaeophyceae	*Ectocarpus siliculosus*	Ec32 (CCAP1310/4)	1	+	+	+	NC_013498
Phaeophyceae	*Fucus vesiculosus*			+	+	+	NC_016735
Phaeophyceae	*Nereocystis lutkeana*	UWCC MA 708		+	+	+	unpublished data*
Phaeophyceae	*Saccharina japonica*			+	+	+	NC_018523
Pinguiophyceae	*Pinguiococcus pyrenoidosus*	CCMP2188	1	+	+	+	unpublished data*
Raphidophyceae	*Chattonella subsalsa*	CCMP217	1	+	+	+	unpublished

							data*
Raphidophyceae	*Heterosigma akashiwo*	CCMP452	2	-	-	-	EU168191
Raphidophyceae	*Heterosigma akashiwo*	NIES293	2	-	-	-	NC_010772
Synurophyceae	*Synura petersenii*	CCMP854		-	-	-	unpublished data*
Xanthophyceae	*Botrydium cytosum*	UTEX 157		+	+	+	unpublished data*
Xanthophyceae	*Tribonema aequale*	CCMP1275	1	+	+	+	unpublished data*
Xanthophyceae	*Vaucheria litorea*	CCMP2940	1	+	+	+	NC_011600
Dinophyta							
Dinotrichales (dinotom)	*Durinskia baltica*	CS-38		-	-	-	NC_014287
Dinotrichales (dinotom)	*Kryptoperidinium foliaceum*	CCMP1326		-	-	-	NC_014267
Chromerida							
Chromeraceae	*Chromera velia*	CCMP2878		-	-	-	NC_014340
Vitrellaceae	*Vitrella brassicaformis*	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; ($^{\Delta}$) *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.

DIFFERENTIAL REGULATION OF DUPLICATE LIGHT-DEPENDENT PROTOCHLOROPHYLLIDE OXIDOREDUCTASES IN THE DIATOM *PHAEODACTYLUM TRICORNUTUM*

Heather M Hunsperger[1], Christopher J Ford[1], James S Miller[1], Rose Ann Cattolico[1]

[1]Department of Biology, University of Washington, Seattle, WA, USA

Status: Manuscript in preparation.

ABSTRACT

Diatoms (Bacilliariophyceae) encode two light-dependent protochlorophyllide oxidoreductases (POR) that catalyze the penultimate step of chlorophyll biosynthesis. Thus these algae only synthesize chlorophyll in the light. Algae live in dynamic environments whose changing light levels induce photoacclimative metabolic shifts, including altered cellular chlorophyll levels. To elucidate why two POR enzymes (POR1 and POR2 encoded by genes *por*1 and *por*2) are maintained in *Phaeodactylum tricornutum*, differences in *por* gene expression and POR protein abundance were examined when this diatom was grown on an alternating light:dark cycles at different irradiances, exposed to continuous light, and challenged by a significant decrease in light availability.

For cultures maintained on a 12hr light:12hr dark photoperiod at 200μE m^{-2} s^{-1} ($_{200}$L/D), both *por* genes were up-regulated during the light and down-regulated in the dark, though *por*1 transcript abundance rose and fell earlier than that of *por*2. Little concordance occurred between *por*1 mRNA and POR1 protein abundance. In contrast, *por*2 mRNA and POR2 protein abundances followed similar diel patterns. When $_{200}$L/D *P. tricornutum* cultures were transferred to continuous light ($_{200}$L/L), the diel regulatory pattern of *por*1 mRNA abundance but not of *por*2 was disrupted, and POR1 but not POR2 protein abundance dropped steeply. Cultures maintained at 1200μE m^{-2} s^{-1} ($_{1200}$L/D), exhibited a weaker diel oscillatory pattern than observed under $_{200}$L/D, though a diel pattern of POR2 abundance was maintained. When cells grown at $_{1200}$L/D were then shifted to 50μE m^{-2} s^{-1} ($_{50}$L/D), *por*1 and *por*2 mRNA levels decreased swiftly but

briefly upon light reduction. Thereafter, POR1 but not POR2 protein levels rose significantly in response to this light stepdown.

Given the sensitivity of *por*1/POR1 to real-time light cues and adherence of *por*2/POR2 regulation to the diel cycle, we suggest that diatom POR1 supports photoacclimation, whereas POR2 is the workhorse of daily chlorophyll synthesis.

INTRODUCTION

Two functionally equivalent but non-homologous enzymes catalyze the penultimate step of chlorophyll synthesis: the light-dependent (POR) and light-independent (LIPOR) protochlorophyllide oxidoreductases [1, 2]. Both proteins reduce the C17=C18 double bond of the chlorophyll precursor protochlorophyllide (Pchlide) to form chlorophyllide (Chlide). The addition of a phytol tail to chlorophyllide by chlorophyll synthetase results in a mature chlorophyll *a* molecule (Fig. 1). The long held assumption that most algae use both POR and LIPOR to generate chlorophyll was recently revised by the discovery that many algal taxa, including stramenopile, haptophyte, chlorarachniophyte and euglenid representatives, lack LIPOR and instead maintain multiple POR isoenzymes [3]. A similar pattern of LIPOR gene loss and POR expansion has also been documented to occur in diverse angiosperms such as *Arabidopsis,* barley, tobacco, tomato, corn, rice, as well as gymnosperms within the genus *Pinus* (reviewed in [4]). Phylogenetic analyses suggest that the *por* gene duplicates of some land plants may be shared among species, whereas other duplicates are specific to individual species [4]). Importantly, the origins of algal and land plant *por* gene duplicates differ.

The origin of the first *por* gene and all other photosynthesis-related genes in algae can be traced to the endosymbiotic entrainment of a proto-cyanobacterium in a eukaryotic host cell [5]. The rhodophytic (red) and chlorophytic (green) algal lineages diverged ~1,500 million years ago from this eukaryote-prokaryote chimera [6–8]. Whereas modern green algae and land plants emerged directly from the green algal lineage, other 'green-lineage' algal taxa have been established via unique secondary endosymbioses of green algae (and therefore green algal genes)

that were incorporated into previously non-photosynthetic eukaryotic taxa (e.g., chlorarachniophyte and euglenid algae) [9, 10]. Similarly, secondary or potentially higher order endosymbioses involving the assimilation of red algae as chloroplasts have produced additional 'red-lineage' algae (e.g., the stramenopiles, haptophytes, cryptophytes and dinoflagellates) [11, 12]. Phylogenetic analyses suggest that the duplication events leading to two *por* genes in both euglenids and chlorarachniophytes occurred after each of these algal lineages were established [3]. Uniquely, nearly all stramenopiles and haptophytes appear to have lost their native red algal *por* genes. Instead, both of these algal lineages share duplicates of a *por* gene obtained via horizontal gene transfer from the ancestral prasinophyte lineage of green algae. Phylogenetic analyses suggest that the stramenopiles first incorporated and duplicated the prasinophytic *por* gene, and that this dual gene set was then transferred to haptophytes in a separate horizontal or potentially endosymbiotic gene transfer event [3].

The maintenance of redundant gene sets for extended evolutionary time periods is ascribed to divergences in biochemistry or regulation of their resultant product(s) that offer adaptive advantages [13]. Gene duplication can increase gene dosage. Alternatively, mutations of the coding or regulatory sequences can divide enzymatic responsibility between gene duplicates and or enable the rise of novel functions. Because the *por* gene families of stramenopiles/haptophytes, chlorarachniophytes, euglenids, and land plants were individually established via unique gene duplication events, each of these expanded *por* gene families evolved separately. Because each of these taxa possesses vastly different evolutionary histories, nuclear gene compliments, regulatory networks, physiologies and ecologies, one may posit that each POR isoenzyme fulfills different needs for each organism. Alternatively, given the

universality of *por* gene duplication across evolutionarily distant lineages, the possibility of convergent evolution in POR enzyme regulation in response to similar environmental stimuli warrants consideration.

A particularly well-studied *por* gene expansion is that observed in the land plant *Arabidopsis*. This organism maintains three nuclear-encoded *por* genes (*porA, porB, porC*) encoding unique POR enzymes (PORA, PORB, PORC) that each fulfill a specific role upon transit to the chloroplast thylakoid membrane. *A. thaliana por*A is highly transcribed and translated in dark-adapted seedlings, poising tissues for rapid greening upon exposure to light [14]. The gene *por*B is under circadian regulation, supporting daily chlorophyll synthesis [15]. The third gene, *por*C, is up-regulated in response to increasing light intensities and is postulated to enable elevated rates of chlorophyll synthesis under high light [15]. Similar to *Arabidopsis*, the multiple *por* genes of other angiosperms as well as those of gymnosperms display unique regulatory schemes for each gene copy. Some species appear to share similar *por* gene regulatory programs even though their *por* genes arose from unique duplication events during evolution [4]. For example, *H. vulgare* also has a *por* gene specialized to seedling greening and another for daily chlorophyll synthesis although these *por* gene duplicates arose independently from those of *A. thaliana* [16].

Just as some land plants regulate their unique *por*s in a similar manner, it is reasonable to suspect that the regulatory patterns of some algal *por* gene duplicates converge upon plant regulatory schemes in response to commonly experienced environmental cues. Diurnal light/dark cues are a good example. Analysis of pigment synthesis in the diatom *P. tricornutum* (a stramenopile) showed an increase in cellular chlorophyll content in the light and decrease in the dark

concomitant with cellular division [17]. Upon continuous illumination, cellular chlorophyll content rose in the subjective day, decreased in the subjective night, and increased considerably just prior to subjective dawn (whereas synthesis would normally be inhibited due to the light-requirement of the POR enzyme)—indicating circadian regulation of pigment synthesis in this alga.

However, physiological differences between land plants and algae suggest that their expanded *por* gene families may be functionally divergent. For example, whereas land plants generally increase chlorophyll levels under high light conditions [18], cyanobacteria and eukaryotic algae (including diatoms, cryptophytes, prymnesiophytes, dinoflagellates and chlorophytes) have been found to increase chlorophyll concentrations under low light intensities as part of longer-term (hours to days) photoacclimative responses [19–21]. Additionally, phytoplankton lack the elaborate reproductive structures (e.g., seeds) and complex tissue differentiation processes that require specialized metabolic programs during development.

The well-studied stramenopile class Bacillariophyceae (diatoms) is an ideal taxon to initiate studies probing how algal taxa utilize their unique *por* gene sets. Genomes and transcriptomes of diatoms have been particular well sampled, revealing that at least 22 of 24 sampled diatom species possess *por* gene duplicates [3]. All of six of the fully sequenced diatom chloroplast genomes lack LIPOR genes [3], indicating the dependence of these algae on their light-dependent POR proteins. Using the model diatom *P. tricornutum*, we examine *por1* and *por2* mRNA abundance as well as POR1 and POR2 protein abundance for cultures maintained on alternating light:dark photoperiods or under constant illumination at moderate light intensities.

These experiments provide insight into diurnal and potentially circadian oscillation of *por* gene regulation at both the transcriptional and post-translational levels. Diatoms photoacclimate when challenged by shifts in light intensity (e.g., during transition from brightly lit surface waters to poorly lit deeper waters). A key process in photoacclimation is the adjustment of cellular chlorophyll abundance, which can require hours to days to achieve (e.g., [19]). By monitoring *por* gene regulation and POR protein abundance over several days after transition to low light, we investigate how each *por* gene potentially contributes to photoacclimation. The results of our studies suggest that each *P. tricornutum por* gene is regulated differently over the diel cycle, that *por*1/POR1 may play a role in photoacclimation to low light by diatoms, and that *por*2 transcription and POR2 protein abundance are potentially under diel regulation.

RESULTS

Characterization of *P. tricornutum* PORs

The hypothetical sizes of the POR1 and POR2 proteins are ~47kD and ~61kD, respectively, including putative bipartite signal/transit peptides. Using the HECTAR algorithm [22] to identify the bipartite targeting peptide of stramenopile chloroplast-targeted proteins, POR1 was predicted to be chloroplast-localized with high confidence, demonstrating both canonical signal and transit peptides (Fig. 2). In contrast, only the signal peptide was identified for POR2, indicating that although the protein enters the outermost of the four membranes that enclose stramenopile chloroplasts (additional membranes are due to chloroplast acquisition via secondary or higher order endosymbiosis), its mode of passage through the remaining three membranes is presently unknown (see [23, 24]). Alternative methods for protein movement (e.g., vesicular transport) within complex plastids of *P. tricornutum* have been reported [25, 26].

P. tricornutum POR1 and POR2 protein sequences show similar levels of sequence conservation to one another (65% biochemical similarity) as to other POR enzymes (e.g., 54-65% similarity to that of the cyanobacterium *Plectonema boryanum*) when the N-terminal signal and transit peptides of POR1 and POR2 and the long C-terminal tail of POR2 are excluded (Fig. 2). For comparison, the three *A. thaliana* POR enzymes and two *H. vulgare* POR enzymes show 84-94% and 100% biochemical similarity to one another, respectively. The greater sequence divergence of the duplicate diatom POR enzymes compared to land plant POR duplicates likely reflects the ancient origins of the diatom POR enzyme duplication event, pre-dating the formation of extant stramenopile (and haptophyte) algal lineages ~800-1100 million years ago [6, 8, 33].

Most mature land plant and chlorophytic algal POR proteins are ~36kD [34]. The larger hypothetical size of the mature *P. tricornutum* POR2 enzyme (~59 kD without signal peptide) is due to an unusual C-terminal extension (Fig. 2). This C-terminal extension is also present in POR2 enzymes encoded in the genomes of the diatoms *Fragilariopsis cylindrus*, *Pseudo-nitzschia multiseries*, *Thalassiosira pseudonana* as well as another stramenopile, *Aureococcus anophagefferens* (Pelagophyceae). All these algae share the same duplication of a horizontally-transferred prasinophyte *por* gene [35]. The C-terminal extensions of the diverse diatom POR2s show 50-68% biochemical similarity to one another and 39-50% similarity to that of *A. anophagefferens*.

Heterologously expressed *P. tricornutum* POR1 and POR2 proteins were used for antibody production. cDNAs for expression constructs were obtained by 3' RACE, and both *por*1 and *por*2 cDNAs matched protein translations predicted in the genome annotations [35], confirming the hypothetical protein sizes of POR1 and POR2. The cDNAs were then cloned into a pET-15-HE vector and expressed in bacteria (see Methods). The resultant, affinity purified anti-POR1 antibody detected heterologously expressed POR1 but not POR2 proteins, and cross-reacted with a 42.5kD band of approximately expected size for POR1 (Fig. 3). The anti-POR2 antibody cross-reacted with heterologously expressed POR2 proteins but not POR1 proteins, and detected a single protein band in *P. tricornutum* extracts (Fig. 3). Interestingly, this band was 44kD, which is much smaller than the expected ~59kD protein based on 3' RACE of the mRNA transcript. These results suggest that the unusual, ~15kD C-terminal extension of this protein is cleaved, leaving the protein at a typical size for POR enzymes. It should be noted that the C-terminus of

non-extended POR proteins (e.g, as found in cyanobacteria, red algae, green algae and land plants) enables the enzyme to associate with the thylakoid membrane [36]. In this context, the cleaved C-terminal extension of some stramenopile PORs may possibly function in regulation or subcellular localization.

Culture maintenance

To ensure a uniform culture source for the multiple samples required for each experiment, *P. tricornutum* cells were grown as a semi-continuous culture in a 15L photobioreactor containing 12L of medium (Supplemental file 1). To eliminate issues of self-shading or nutrient depletion over the course of extended sampling periods, photobioreactor contents were diluted each 24hr period at the onset of the light (L0) with fresh medium. Depending on the sampling regime, 20-30% of the culture was harvested every 24hrs (15% culture depletion occurred during the light period). *P. tricornutum* growth responses in this large photobioreactor are shown in Supplementary file 2 for cells grown either under a 12h light:12h dark photoperiod (L/D) or under constant illumination (L/L) at 50μE m^{-2} s^{-1} ($_{50}$L/D), 200μE m^{-2} s^{-1} ($_{200}$L/D and $_{200}$L/L) and 1,200μE m^{-2} s^{-1} ($_{1200}$L/D). The intermediate light intensity of 200μE m^{-2} s^{-1} ($_{200}$L/D) was chosen to serve as a reference control for the high and low light studies described below. Growth responses in the semi-continuous cultures, regardless of light regime, were shown to parallel those obtained batch cultures (Supplemental file 2).

Diel rhythmicity: $_{200}$L/D

Cell growth responses and pigment production: To determine the effect of a moderate light regime on *P. tricornutum* POR1 and POR2 production, cells were sampled over a two day period

when grown under a 12h light:12h dark light/dark cycle at 200μE m^{-2} s^{-1} ($_{200}L_1/D_1$ and $_{200}L_2/D_2$; Fig. 4). Cultures were partially synchronized by this light/dark regime, dividing approximately once daily. Division for both days initiated by the 7th hour in the light (i.e., $L_1$7 and $L_2$7) and ceased by the 7th hour in the dark (i.e., $D_1$7 and $D_2$7; Fig. 4a), in a pattern that repeated in both phase and amplitude. Average cell size was maximal in the middle of the light period (~4.9μm) and minimal at the termination of cell division (~4.2μm; Fig. 4b). *P. tricornutum* is only lightly silicified and so size is not reduced with each division cycle as for other diatoms [37]). Maximal cell sizes likely represent dividing cells prior to separation into daughter cells. Changes in both total and cellular pigment abundance (cellular fluorescence; Figs. 4c and 4d) showed a highly reproducible phase period, and amplitude when monitored over the two-day sampling period. Total pigment abundance increased linearly over the course of each day (e.g., $_{200}L_1$2 to $_{200}L_1$11). Mean pigment abundance per cell, however, slowed upon the onset of cell division (e.g., $_{200}L_1$7 to $_{200}L_1$11) and thereafter declined until the termination of the dark cycle (e.g., $_{200}D_1$12). Average cellular protein levels for both L_1/D_1 and L_2/D_2 cycles remained high in the light declined to lowest values by the end of the dark portion of the cycle, coincident with the termination of cell division (Fig. 4e). Taken together, *P. tricornutum* cellular division, as well as pigment and protein abundance appeared to be highly consistent across photoperiods and tightly coupled to light cues.

Transcriptional regulation of *por*1 and *por*2: Given the potential redundancy of *por*1 and *por*2, it was of interest to know whether transcription abundance of these two genes differed as *P. tricornutum* progressed through the diel cycle. Housekeeping genes proposed for qPCR abundance normalization standards in *P. tricornutum* light/dark cycle studies [38] were evaluated

for use in this analysis. Genes *cdkA* (cyclin-dependent kinase A); *RPS* (30S ribosomal protein subunit), and *TBP* (TATA box binding protein) showed acceptable, low variability (ΔC_ts ranging from 2.9-3.5; Supplemental file 3) over the experimental conditions chosen. The ΔC_ts of *por1* and *por2* were 5.9 and 6.9 in the $_{200}$L/D experiment.

Figure 5a shows that *por*1 mRNA abundance oscillated reproducibly over two 24hr $_{200}$L/D cycles. A small increase in *por*1 product appeared during the dark (perhaps indicating anticipation of the onset of light), and ramped up to a high abundance with the onset of light. This peak was followed by a very rapid decline soon after hour 3 in the light, reaching a restored minimum by the onset of darkness. The regulation of *por*2 transcript abundance contrasted with that of *por*1. Though transcript abundance of *por*2 was also maximal in the light, this RNA peaked in abundance from mid to late in the light period. However, similar to *por*1, as soon as RNA transcript load peaked, a rapid decline ensued that extended into the dark period. In summary, *por*1 and *por*2 RNA transcripts were both up-regulated in the light and down-regulated in the dark, but displayed a phase shift relative to one another.

<u>POR1 and POR2 enzyme accumulation:</u> POR1 protein abundance showed a different pattern in the light versus the dark. This pattern was maintained when measured as POR1 per cell or POR1 per μg protein (Fig. 5b). During the $_{200}L_1$, $_{200}L_2$ and $_{200}L_3$ light phases, a distinct 'peak' was seen wherein the amount of POR1 protein was elevated (~L7). However, the POR1 abundance pattern did not have a reproducible periodicity. In the $_{200}D_1$ and $_{200}D_2$ dark phases, POR1 protein levels appeared to be constant and approximately as high as peak levels observed in the light phases of cell growth. In contrast, POR2 enzyme accumulation per μg protein remained fairly stable

throughout the light/dark cycle. At the cellular level, however, a shallow oscillation was seen in response to light/dark cues; cellular POR2 levels increased modestly during $_{200}L_1$, $_{200}L_2$ and $_{200}L_3$, then declined throughout $_{200}D_1$ and $_{200}D_2$, reaching a minimum at the end of the dark phase (Fig. 5c). These cell-level oscillations likely reflect both POR protein anabolism/catabolism over the diel photoperiod as well as cellular protein reduction due to cell division (Figs. 4a, 4e). Data show that there was no strong correspondence between *por*1 RNA abundance and POR1 protein levels over the diel cycle – suggesting that the regulation of POR1 protein level most likely occurs post-transcriptionally. A closer, though imperfect, correspondence between *por2* mRNA and cellular POR2 enzyme levels was seen. Most intriguing, given the light requirement of the POR enzyme for catalysis, is the retention of high levels of POR1 and POR2 enzymes in the dark phases of the diel cycle when compared to the light phases.

Photoacclimation: $_{200}$L/D to $_{200}$L/L

Cell growth responses and pigment production: To determine the impact of eliminating the dark cue on POR enzyme biogenesis, the *P. tricornutum* culture previously maintained for two 12h light:12h dark cycles (L_1/D_1 and L_2/D_2) was shifted to a continuous light regime while maintaining the same light intensity (200μE m^{-2} s^{-1}). As seen in Figure 4a, in the first 24hr on this new program, cell division proceeded from $_{200}L_3$7 to subjective night $_{200}sD_3$7 as expected under a normal light/dark cycle, but achieved only a 0.89 cell divisions per day rather than the 1.0 per day observed in previous $_{200}L/_{200}D$ cycles. Additionally, the continued lack of a dark cue most likely resulted in the early onset of cell division during the ensuing $_{200}L_4/sD_4$ cycle. Division began by $_{200}L_4$0, and appeared to persist through the termination of the experiment at $_{200}sD_4$2.

The shift in the timing of cell division for cultures grown in continuous light also impacted cell size (Fig. 4b). Although a maximum average cell size of 5.0μm was attained by $_{200}L_3$7 in the first continuous light cycle, average cell size only declined to 4.6μm by $_{200}sD_3$7 rather than the expected 4.2μm minimum. Subsequently, cell size no longer achieved the $_{200}$L/D maximum, but continued to slowly and linearly decrease until the cells achieved a size (4.3μm) similar to that normally observed at $_{200}$L/D post-division (4.2μm). These data plus the linear increase in protein per cell beginning at $_{200}L_4$2 (Fig. 4e) suggest that the culture began to lose a phased division response and started to divide continuously.

Under the continuous light regime, total pigment abundance increased in the $_{200}L_3$ culture similarly to that observed in $_{200}L_1$ and $_{200}L_2$ (Figure 4c). Pigment accumulation continued in the subjective dark of $_{200}sD_3$, reaching fluorescence levels approximately 10% higher than previous dark periods, reflecting the increase in cell size. Pigment abundance per cell at $_{200}sD_3$7 was equal to that seen in $_{200}D_1$7 and $_{200}D_2$7. However, from $_{200}sD_3$7 to $_{200}sD_3$11, average pigment per cell did not decline as rapidly as anticipated for a normal dark period, nor increased from $_{200}L_4$2 to $_{200}L_4$11 as expected for a normal light period. Overall pigment decline per cell most likely was due to the early cell division that took place in the L_4 cycle combined with the decrease in total pigment abundance.

Transcriptional regulation of *por*1 and *por*2: Despite a continuous light regime, *por*1 transcript levels reached an expected low at the onset of the subjective dark at sD_3 and then slowly increased in abundance in sD_4, as expected on a $_{200}$L/D cycle (Fig. 5a). Lacking the dark reset

however, *por*1 mRNA abundance was not highest at the onset of the light phase as found under the $_{200}$L/D regime. Instead, *por*1 transcript increase was delayed until the middle of L_4, indicating a phase shift had occurred. Transcript accumulation for *por*2 appeared to maintain a normal oscillation whether the cultures were subject to a $_{200}$L/D or $_{200}$L/L light program. Interestingly though, *por2* transcript abundance remained slightly elevated during the $_{200}sD_3$ when compared to levels observed during a true dark period (e.g. $_{200}D_1$ and $_{200}D_2$).

POR1 and POR2 enzyme accumulation: Despite the transition from an L/D regime to continuous illumination, POR1 proteins accumulated as expected during L_3 (Fig. 5b). Thereafter, the concentration of this protein began to drop precipitously after $L_3$7 and was only marginally detectible for almost 24hrs when, at $L_4$11, a slight up-tick in enzyme signal was seen. This signal was further augmented in sD_4, suggesting a re-establishment of protein complement was potentially occurring. POR2 protein maintained the shallow oscillatory response in abundance per cell that peaked at ~L7 for the $_{200}$L/D regime (Fig. 5c). However, POR2 abundance (both per cell and per μg protein) was much lower by $sD_3$2 than at $D_1$2 and $D_2$2. Furthermore, POR2 abundance per μg protein remained at approximately half of the levels seen during a normal $_{200}$L/D cycle through $L_4$2. In the next expected light period, POR2 levels returned to typical $_{200}$L/D levels.

Photoacclimation: $_{1200}$L/D to $_{50}$L/D

Cell growth responses and pigment production in high light cultures: To determine whether the regulation of *por*1/POR1 and *por*2/POR2 differ as *P. tricornutum* cells adjust to a light stepdown, cells were sampled as they transitioned from a 1200μE m^{-2} s^{-1} L/D cycle ($_{1200}$L/D) to a

50μE m^{-2} s^{-1} L/D cycle ($_{50}$L/D) (Fig. 6). As seen in Figure 6a, the *P. tricornutum* population acclimated to $_{1200}$L/D grew slightly faster (1.2 divisions per day) than the control culture grown at $_{200}$L/D (1.0 division per day; Fig. 4a). In the middle of the light period (L7, before cell division), the mean cell fluorescence of *P. tricornutum* grown at $_{1200}$L/D was on average 60% as high as that seen for the diatom culture maintained at $_{200}$L/D. Thus the high light culture displayed the expected increase in cell division rates and decrease in cellular chlorophyll levels [17, 20]. Although pigment levels were reduced, the amount of pigment per cell and protein per cell present during the two $_{1200}$L/D cycles displayed a similar period and phase in product abundance to that observed for $_{200}$L/D grown cultures. However, under $_{1200}$L/D, but not under $_{200}$L/D or $_{50}$L/D, total pigment levels increased during the dark period. This observation may reflect an increased accumulation of additional pigments (e.g., fucoxanthin, as previously observed in dark-adapted *P. tricornutum* [39]). Maximum protein levels were approximately twice those observed in the $_{200}$L/D reference control.

Transcriptional regulation of *por*1 and *por*2 and POR1 and POR2 enzyme accumulation: Given that the ΔC_ts of genes *cdkA* and *TBP* varied little over the light stepdown experiment (ΔC_ts from 3.1-3.5), these genes were also used as housekeeping controls for this study. The ΔC_ts of *por1* and *por2* were both 6.3 in this $_{1200}$L/D:$_{50}$L/D experiment.

The high light exposure of 1200μE m^{-2} s^{-1} appeared to impact both *por* mRNA as well as POR protein abundance patterns when observed over the L/D cycle. Although both *por*1 and *por*2 gene products were most abundant in the light, the reproducible amplitude and periodicity seen in the $_{200}$L/D culture was lost (Fig. 7a). In contrast to the differences observed in both POR1 and

POR2 protein $_{200}$L/D patterns, those observed for $_{1200}$L/D showed a strong, tightly coordinated oscillation at the cellular level: high in the light and low in the dark (Fig. 7b,c).

Cell growth responses and pigment production upon transition to low light: As seen in Figure 6a, transition to the $_{50}$L/D light regime caused an immediate decrease in the rate of cell division to 0.26 divisions per day on $_{50}L_3/D_3$. Cell division rose slowly as cells acclimated to the new light condition, reaching 0.33 divisions per day by $_{50}L_5/D_5$ and maintaining approximately 0.38 divisions per day on $_{50}L_8/D_8$, the same cell division rate seen in $_{50}$L/D acclimated batch cultures in exponential growth (Supplemental figure 2). In addition to this decrease in the amplitude of the cell division response, the timing of cell division shifted so that this process was primarily constrained to the dark period of the diel cycle rather than extending from L7 to D7 as seen for the $_{200}$L/D and $_{1200}$L/D cultures. Average cell size initially remained high under $_{50}$L/D (~5.2µm; Fig. 6b), most likely because cells had not yet reached an adequate level of growth to enable cell division. Mean cell size then decreased linearly as the cellular division rate slowly increased, with daytime maximum cell size averages much smaller than those observed under $_{1200}$L/D (~4.9µm on the sixth day at $_{50}$L/D compared to ~5.6µm under $_{1200}$L/D). Protein abundance retained a predictable though imperfect oscillatory pattern due to cell growth and division over the $_{50}$L/D cycle.

The transition from $_{1200}$L/D to $_{50}$L/D initially led to decreased total pigment levels at the end of the $_{50}$L3/D3 cycle (Fig. 6c). However, by the end of the second low light photoperiod ($_{50}$L4/D4), total pigment levels in the culture had rebounded to the levels observed at $_{1200}$L/D and continued rising. Similarly, a slow increase in pigment per cell began upon transfer to $_{50}$L/D. By the end of

six days at the new $_{50}$L/D regime, mean fluorescence per cell at L7 (before cellular division in the $_{1200}$L/D culture) was 173% higher than under the $_{1200}$L/D photocycle. At the level of the individual *P. tricornutum* cell, cellular pigment accumulation rates (measured from L2-L7 to exclude post-L7 cellular division under $_{1200}$L/D) dropped dramatically on the first day at $_{50}$L/D, recovered by the second day at low light, and increased thereafter to ~138% of the rate at $_{1200}$L/D. Given the higher rate of cellular pigment synthesis measured at $_{50}$L/D, it appears that both increased pigment production and decreased cellular division contribute to the increasing cellular pigment levels.

Transcriptional regulation of *por*1 and *por*2: When subject to a light step down from $_{1200}$L/D to $_{50}$L/D, *por*1 mRNAs failed to accumulate normally during the first day under the new reduced-light regime ($_{50}L_3/D_3$) whereas *por*2 mRNA was abundant at $L_3$2 but declined precipitously by $L_3$7 (Fig. 7a). Subsequent to this 24-hour acclimation period, a burst in both *por*1 and *por*2 mRNA accumulation was observed. The high levels of *por*1 mRNA observed at $_{50}L_4$11 dropped steadily through the end of the experiment at $_{50}D_5$, suggesting that an overcompensation had occurred. In contrast, *por*2 mRNA levels appeared to form a shallow but incomplete oscillation with peaks during the light period, a pattern similar to that seen for $_{1200}$L/D.

POR1 and POR2 enzyme accumulation: Despite a sharp drop in both *por*1 and *por*2 mRNA levels during the $_{50}$L3/D3 light exposure, Western blotting gives a completely different view of POR1 and POR2 regulation in cells adapting to a new light regime (Fig. 7b,c). POR1 protein levels per μg peaked after transition to $_{50}$L/D, with much greater abundance levels at $_{50}L_3$11 and $_{50}D_3$11 than in the two previous $_{1200}$L/D cycles ($p<0.0001$). Thereafter, POR1 protein abundance

decreased to a less-defined L/D pattern than that seen in the $_{1200}$L/D regime. Despite robust POR1 abundance peaks during transition from the $_{1200}$L/D to the $_{50}$L/D light regime, POR1 levels were not significantly different when the three $_{50}$L/D photoperiods were compared to the two $_{1200}$L/D photoperiods (*p=0.191*). These data show that most changes in POR protein levels occurred during the 24-hour transition to the lower light intensity. In contrast to these observations for POR1, the accumulation of POR2 enzyme under $_{50}$L/D appeared to be regulated in a similar manner as that observed for $_{1200}$L/D. The light phase-dominant increase in cellular levels of POR2 protein seen in the two $_{1200}$L/D cycles was repeated in all three $_{50}$L/D cycles.

DISCUSSION

Day to night fluctuations in photoperiod present a powerful and nearly universal environmental signal for metabolic entrainment [40]. However, in addition to diel changes in light intensity, algae must cope with unforeseen light fluctuations caused by vertical displacement within the water column, changes in cloud cover, shading during a bloom, and detritus within the water column. Significant changes in pigment production represent a primary metabolic response by which phytoplankton acclimate to such changes in light availability [20, 41]. The goal of this study was to observe how cells use their dual POR enzymes when exposed to changing light regimes. By tracking responses to changing light programs over multiple days, we monitored normal diel activities as well as those associated with photoacclimation. Important physiological parameters were carefully regulated in this experimental series to allow shifts in light intensity to be the sole variable under investigation. By diluting the culture at the onset of light each photoperiod, self-shading was alleviated and nutrients replenished. Additionally, the use of a large volume photobioreactor enabled repetitive sampling of cells from a single mother culture over a multi-day period. This approach avoids variability often observed when several small batch cultures are used for sourcing samples.

Physiological responses at various light intensities:

Repetitive sampling of *P. tricornutum* cultures maintained at either $_{200}$L/D or $_{1200}$L/D over several days, revealed moderately to tightly controlled oscillations in cell division response, cell size variation and pigment accumulation. Phased cell division occurred at least once daily, usually taking place from the middle of the light interval to the middle of the dark interval.

Cellular pigment levels were ~70% higher in moderate ($_{200}$L/D) than high light ($_{1200}$L/D) cultures, reflecting the increased pigment complement needed for effective light capture at reduced light fluences [19–21]. Upon transition of $_{200}$L/D cultures to constant illumination ($_{200}$L/L), the periodicity of cell division was disrupted and growth thereafter trended towards constant cell division [17, 42], reduced cell size and a decreased rate of pigment accumulation. In contrast, the first day of transition from $_{1200}$L/D (high light) to $_{50}$L/D (very low light) caused a rapid decline in cell division but maintenance of cellular pigment levels. Over the next 6 days, the $_{50}$L/D culture adjusted to the reduced light program by maintaining a reduced cellular division rate and diminished average cell size, while maintaining a steadily increasing cellular pigment content. The decline of average cell size as *P. tricornutum* transitioned from either a diel cycle to continuous light or from high to low light fluences is noteworthy given that light absorption efficiency has been shown to be greater in smaller cells (reviewed in [43]).

Regulatory differences among por1/POR1 and por2/POR2 suggest different roles for each protochlorophyllide oxidoreductase in *P. tricornutum*:

P. tricornutum cultures subject to12hr light:12hr dark diel cycles (either $_{200}$L/D or $_{1200}$L/D) show *por*1 and *por*2 mRNA abundances that oscillate between daytime maxima and nighttime minima, though *por*1 and *por*2 responses are shifted in phase. No concordance was observed between *por*1 transcription and POR1 protein accumulation under $_{200}$L/D, though moderate agreement was observed under $_{1200}$L/D. POR2 but not POR1 protein abundance per cell shows diel oscillation with photophase peaks and scotophase troughs under both $_{200}$L/D and $_{1200}$L/D. This result suggests that post-transcriptional regulation may play a greater role in the control of POR1 than POR2 enzyme levels.

When *P. tricornutum* cultures were exposed to changes in their ongoing diel light program, changes in transcription and enzyme abundance of each POR shifted with different response rates. For example, transfer from $_{200}$L/D to constant light delayed *por*1 transcription after only one subjective dark interval. In contrast, *por*2 mRNA abundance continued to maintain a L/D-like oscillation despite lacking a dark interval cue. Differences at the transcriptional level were manifest at the post-translational level. POR1 proteins all but disappeared during the first subjective dark interval and did not begin to recover until the end of the next expected light period. In contrast, POR2 levels maintained an expected daily oscillatory pattern, although with reduced abundance during and recovering from the subjective dark period. These data show *por*1/POR1 regulation to be more markedly influenced by real-time light cues than *por*2/POR2.

Given the apparent adherence of *por*2/POR2 regulation to the diel cycle but sensitivity of por1/POR1 regulation to changing light intensities, one may hypothesize that the POR1 enzyme plays a key role in photoacclimation. Indeed, POR1 enzyme abundance increased to 2.5X the peak levels observed under $_{1200}$L/D by the end of the first $_{50}$L/D light interval. Thereafter, POR1 abundance remained elevated as a percentage of cellular protein (density per μg protein). POR2 protein abundance, in contrast, oscillated normally during the high to low light transition.

Different responses to light by *P. tricornutum*'s two *por* genes suggests that each is most likely controlled by a different regulatory network. Additional clues come from microarray studies of *por* gene transcription wherein *P. tricornutum* cultures were grown under various light regimes. To eliminate influences of diel cues, cultures in these studies were maintained under constant

light (L/L) conditions [44, 45]. Similar to many other genes in the chlorophyll synthesis pathway, *P. tricornutum por*1 and *por*2 genes were briefly (<3hr) down-regulated during a light step-up from 35μE m^{-2} s^{-1} to 500μE m^{-2} s^{-1}, though *por*1 was less downregulated than *por*2 [44]. In a second experimental series [45], transfer of culture from moderate light (100μE m^{-2} s^{-1}) to 48 hours of darkness caused *por*2 transcript abundance to decline. Upon re-exposure to light, *por*2 transcript abundance slowly rose. In contrast, *por1* was not down-regulated during prolonged darkness but was quickly (<0.5hr) down-regulated upon exposure of *P. tricornutum* to light and remained depressed for at least 24 hours [45]. The high abundance of *por*1 transcripts in dark-adapted cultures and our finding that the regulation of *por*1/POR1 is responsive to real-time environmental changes is intriguing. In terms of cell survival, future studies may explore whether POR1 plays a role in diatom photoacclimation when sediment-dwelling resting cells return to the photic zone [46].

Light quality as well as light quantity appears to influence *por* gene transcriptional responses. For example, when the effects of spectral quality on *P. tricornutum* gene expression were analyzed, photosynthetically equivalent quantities of red, green, and blue light did not significantly alter the transcription of *por*1 in dark-adapted cells compared to cells exposed to white light 0.5hr after re-illumination [39]. In conjunction with our results, one may speculate that light intensity or duration rather than quality may influence the regulation of the *por*1 gene. In contrast, *por*2 was down-regulated under blue light and up-regulated under red and green light (though see [39] for discussion). As noted previously, it is well established that photoacclimating algae reduce cellular chlorophyll levels as light increases. Because the *por*2 repression was observed 0.5hr after re-illumination but not later, the reduced *por*2 mRNA level may represent a

transient down-regulation in response to an increased light availability. This conclusion is further supported by the fact that *por*2 transcription is down-regulated 0.5hr after transfer from 35μE m^{-2} s^{-1} to 500μE m^{-2} s^{-1}, but returns to normal levels within three hours [44].

A potential link between *por*2 transcription and blue light reception is especially interesting given (a) the universal roles played by blue-light receptors such as cryptochromes in the entrainment and maintenance of circadian rhythms [47, 48] and (b) the strict adherence to diel light cues we observe for *P. tricornutum por*2 transcription as well as cellular POR2 abundance. Further work is warranted to determine whether *por*2 is regulated by a cryptochrome or other blue-light receptor such as an aureochrome, a class of photoreceptor specific to the diatom crown-taxon Stramenopila [49]. Silencing of particular aureochromes in *P. tricornutum* leads to a high-light acclimation-like phenotype with reduced chlorophyll *a* levels [50].

Cumulatively, the data from these experiments suggest that the regulatory networks governing *por1* and *por2* transcription may be unique to each gene. POR1 appears to enable *P. tricornutum* to rapidly respond to changing environmental parameters. In contrast, the regulation of POR2 is intimately tied to the diel cycle and thus this enzyme may function as the 'workhorse' of chlorophyll production. Such complimentary roles of the diatom POR isozymes may enable diatoms to fine tune chlorophyll synthesis for both cell growth and expeditious response to shifts in environmental conditions. Although chlorophyll synthesis is expected to be maximal during the light period, the accumulation of POR enzyme in the dark may enable *P. tricornutum* cells to remain poised for chlorophyll synthesis upon illumination. It is also important to note that our

studies do not monitor enzyme activity, but rather enzyme presence. Enzymatic activity will be affected by the availability of light, Pchlide, and the POR enzyme cofactor NADPH [51].

Presently, the functional relationship between POR1 and POR2 enzymes is unknown. Recently developed gene silencing and gene editing techniques for *P. tricornutum* and other diatoms may enable studies probing enzyme function [52–54]. For example, it would be of interest to know if *P. tricornutum* POR1 and POR2 enzymes could catalytically replace one another (i.e., could a cell survive with only one POR isozyme)? The POR enzyme is a member of the short-chain dehydrogenase-reductase (SDR) family of proteins, the majority of which are found as dimers and tetramers. This observation brokers the possibility for both homo- and hetero-complexes when isoenzymes are present [55, 56]. Indeed, oligomerization has been reported for some PORs [57, 58]. The role of the conserved C-terminus of POR2 also is a very interesting target for further exploration.

Discordance between transcriptional and post-translational regulation of light-dependent protochlorophyllide oxidoreductases in *P. tricornutum*:

Although *P. tricornutum por*2 mRNA transcription showed a similar pattern to POR2 cellular protein abundance in our studies, *por*1 transcript abundance did not match observed levels of POR1 protein (Figs. 5,7). This result underscores a potential disjunction between transcriptional and post-translational regulation of the chlorophyll synthesis pathway in diatoms. Furthermore, these observations caution that studying chlorophyll biosynthesis solely at the level of gene transcription might be misleading. Though few reports are available for algae, precedent can be found for a discontinuity between *por* gene expression and POR protein abundance in studies of

terrestrial plants. For example, despite nearly constitutive *por* mRNA transcription, abundance of the sole POR protein of *Pisum sativum* (pea) declines upon exposure of etiolated seedlings to light [59]. Light-induced proteases were found to be responsible for a similar loss of PORA from etiolated seedlings of *H. vulgare* [60, 61]. In addition to targeting the POR enzyme itself, post-translational regulation of chlorophyll synthesis in plants involves a highly reticulated metabolic network that modulates POR substrate availability. In *A. thaliana,* accumulation of the POR enzyme substrate Pchlide triggers the protein FLUORESCENT. This protein interacts with the upstream enzyme glutamyl-tRNA reductase to suppress the formation of aminolevulinic acid - the first committed precursor of tetrapyrrole biosynthesis (reviewed in [62, 63]). Importantly, inhibition of POR precursor accumulation may have evolved to protect plants from high levels of Pchlide and other highly toxic chlorophyll intermediates. These pigments, when unbound in light-harvesting complexes, may otherwise promote photobleaching upon re-illumination via the formation of reactive oxygen species (reviewed in [64, 65]).

Comparison of *por* gene duplicate regulation among diverse taxa:

Although the *por* gene duplicates of many angiosperms and gymnosperms arose from various duplication events in the course of evolution, the regulatory patterns of the resultant *por* transcripts and POR proteins may converge (reviewed in [4]). For example, at least one of the two or more POR enzymes present in *A. thaliana, H. vulgare*, and *Nicotiana tabacum* exhibits diel regulation with daytime peaks in expression (*A. thaliana* PORB [14, 15], *H. vulgare* PORB [16], and *N. tabacum* POR1 and POR2 [66]. However, significant variations among regulatory patterns in *por* gene families have been documented. For example, both *A. thaliana* and *H. vulgare por*A and *por*B are expressed in etiolated seedlings, but solely *por*B mRNA persists in

mature tissues [14, 16]. In contrast, only *N. tabacum* "*por*1" is expressed in seedlings, though both *N. tabacum*"*por*1" and "*por*2" are expressed in mature tissues (unrelated to diatom *por*1 and *por*2; [66]). A third innovation occurs in *Pinus taeda* (loblolly pine), which possess two extended sub-families of *por* genes. Both *por*A and *por*B gene sets are expressed in seedlings and both are up-regulated upon seedling exposure to light, though the protein product of one subfamily appears to be reduced in the light [67].

Although *P. tricornutum* is a marine diatom with duplicate *por* genes obtained by horizontal gene transfer from an ancient green algal lineage, its *por*1 and *por*2 regulatory schemes appear to be similar in some ways to those documented for the land plant *A. thaliana por*C and *por*B, respectively. Both *P. tricornutum por*1 and *por*2 and *A. thaliana por*B and *por*C are under diel regulation [15]. *P. tricornutum por*2 and *A. thaliana por*B are potentially under circadian control and their expression appears unaffected by light intensity. The regulatory responses of *P. tricornutum por*1 and *A. thaliana por*C, on the other hand, are greatly affected by changes in light intensity–though most likely differently. Because patterns of chlorophyll accumulation in response to light intensity contrast in algae and land plants (algae decrease and land plants augment chlorophyll abundance under increased light levels [18, 20, 21]), we anticipate that the *P. tricornutum por*1 and *A. thaliana por*C regulatory responses will be inversed.

As noted above, *por* gene expansions were recently documented in a broad range of algal taxa [3]. The regulatory and functional schemes of these diverse *por* gene sets have not yet been fully characterized. From the minimal information presently available, however, it appears that the regulatory patterns of these *por*s differ. For example, transcriptomic analyses indicate that the

diel regulatory programs observed for *Chrysochromulina tobin* (haptophyte) *por*1 and *por*2 genes differ from those of *P. tricornutum* (Hovde et al. *in prep*) even though these genes stem from the same stramenopile/haptophyte *por* gene duplication event. Further transcriptomic data indicates that *por*2 plays a seminal role in *Heterosigma akashiwo* (stramenopile) life history transitions from a dark-dwelling resting phase to an active, photosynthesizing state [68].

In summary, there exists a diversity of protochlorophyllide reduction strategies within the Stramenopile algae, with some taxa maintaining both gene duplicates (and occasionally an additional copy of one duplicate, as in the diatom *Fragilariopsis cylindrus*) but lacking LIPOR, whereas other taxa possess just one *por* gene plus LIPOR (e.g., *Chattonella subsalsa*, *Ectocarpus siliculosus*, *Pinguiococcus pyrenoidosus*). It will thus also be interesting to compare the conditions promoting POR versus LIPOR expression in these diverse algal representatives.

CONCLUSIONS

-This study represents the first paired analysis of transcriptional and post-translational differences between *por* gene duplicates in an alga. Data suggest that each *P. tricornutum por* gene contributes uniquely to chlorophyll synthesis.

-The regulatory networks governing each *P. tricornutum por* gene appear to be distinct given their different responses to light cues. A lack of concordance between *por*1 mRNA and POR1 protein abundance indicates substantial post-translational regulation.

-This report provides a template for probing regulatory schemes of *por* gene expansions documented in a diverse array of algal taxa. Comparison among algae and between algae and terrestrial plants will lead to interesting insights concerning the fate of homologous gene duplicates across taxa varying in their ecologies, physiologies, and evolutionary histories.

METHODS

Culture growth

Phaeodactylum tricornutum Bohlin CCMP632 stock cultures were grown in 1L f/2 medium [69] contained in 2.8L fernbach flasks that were stoppered with cotton and gauze plugs and shaken at 60rpm. Stock cultures were maintained at 20°C and 100μE m^{-2} s^{-1} on a 12 hour light:12 hour dark photocycle and transferred every 4-5 days to maintain cells in the exponential growth phase. For experimental studies, *P. tricornutum* exponential growth phase cultures were used to inoculate an autoclaved 15L MicroFerm Fermentor (New Brunswick Scientific: Edison, NJ) that contained 12L sterile f/2 medium. Starting cell density was $2.5x10^5$ cells/mL for the $_{200}$L/D-$_{200}$L/D experiment and $5x10^3$ cells/mL for the $_{1200}$L/D-$_{50}$L/D study. The 12L culture was maintained at ~17°C using a VWR 1160 re-circulating water chiller (Radnor, PA). Air provided to the culture (1000 cc/min) was first filtered through a Millex 50mm hydrophobic PFTE 0.2μM in-line filter (EMD Millepore: Billerica, MA) to remove potential bacterial contaminates, then bubbled through sterilized water. The culture was mixed using a custom right-handed (upward mixing) impeller with 4 blades pitched at 45° and overall diameter of 10.6cm (culture vessel of diameter 22.9cm) turning at 50rpm. The photoreactor was illuminated on all lateral surfaces with Xlamp XP-E cool white LEDs (Cree: Durham, NC) operated by a custom, programmable controller. External light was excluded by fitting the culture unit and ~30cm of the bases of its air, media and collection tubes with a cover made of black-out duvetyne fabric (Filmtools, Burbank, CA). Upon reaching experimental density, the cultures were maintained in exponential growth by daily dilution at the beginning of the light period (L0).

Sampling

In addition to a ~3.0 mL sample used for cell counts, 12 tubes (45mL) were collected at each time point for protein studies. These samples were kept on ice until centrifugation at 6000x*g* for 20 min at 4°C. The samples were then decanted, the pellet flash frozen in liquid nitrogen and stored at -80°C. Samples harvested during a 'dark' portion of the light/dark cycle were collected and decanted under dim green light provided by Cree Xlamp XP-E green LEDs in a room protected from external light via a duvetyne drape. The light- and dark-harvested samples were collected into clear 50mL (VWR, Radnor, PA) or black 50mL LiteSafe conical centrifuge tubes (Argos Technologies, Elgin, IL), respectively.

Determination of culture density, pigment abundance and cell size

Culture density was monitored in a BD Accuri C6 flow cytometer (San Jose, CA) using cellular fluorescence to differentiate *P. tricornutum* cells from debris. Cellular fluorescence (FL-3 channel; excitation 488 nm; emission 670nm long pass) was used to estimate *P. tricornutum* cellular pigment content [17]. Flow cytometric measurements of cell size were performed with Life Technologies flow cytometry size calibration kit (Grand Island, NY) using the FSC-H measurement of particle size. Given the elongate (fusiform) shape of *P. tricornutum*, size estimates likely refer to the width of cells as they travel single-file past the electronic detectors, with increased width indicating dividing cells.

Determination of gene expression

cDNA preparation: RNA was extracted from cell pellets thawed on ice using TRIzol reagent (Life Technologies) according to manufacturer's directions with 1mL TRIzol per $2x10^8$ cells.

Recovered RNA was quantitated with a NanoDrop UV-vis spectrophotometer (Thermo Fisher Scientific, Waltham, MA), diluted to less than 20µg per 100µL, and treated with DNase I (12U per 10µg RNA) at 37°C for 30 minutes in the presence of 1X DNase I buffer (Life Technologies). The RNA was then further purified with the RNeasy MinElute Cleanup Kit (Qiagen, Valencia, CA) according to manufacturer's directions. To further ensure RNA quality, RNA was tested for DNA contamination by performing qPCR with RNA extracts using 2-4x the amount of RNA as the amount of cDNA used in normal reactions. A ΔC_t of 5 between the amplification of DNA-contaminated RNAs and the highest experimental C_t for primer TBP (as determined in preliminary experiments) represents only 3.1% contamination (or 0.8-1.6% when accounting for the greater amount of starting nucleic acid), and this level of contamination was considered acceptable for RT-qPCR studies. RNAs demonstrating higher levels of DNA contamination were re-DNAsed, purified, and tested again. RNA was reverse-transcribed to cDNA using the Bio-Rad iScript cDNA Synthesis Kit (Hercules, CA) in 20µL reactions each containing 1µg RNA.

RT-qPCR: Primer sets of four housekeeping genes identified by Siaut et al. [38] for use in diurnal cycle studies in *P. tricornutum* were tested: *cdkA* (cyclin-dependent kinase A), *H4* (histone), *RPS* (30S ribosomal protein subunit), and *TBP* (TATA box binding protein). Housekeeping genes *cdkA*, *RPS*, and *TBP* were found to be appropriate for the photoperiod experiment, whereas experiment genes *cdkA* and *TBP* performed satisfactorily for the photoacclimation experiment (Supplemental figure 3). The *por1* primer set was derived from Coesel et al. ([70]; their *por*A) and the *por2* primer set was developed in house. Reactions were performed on a Chromo4 Real-Time PCR system with iQ SYBR Green Supermix (Bio-Rad) in

white 96-well plates with optically clear seals (Bio-Rad) using the following program: initial denaturation was at 96°C for 5min, followed by 40 cycles of 30sec denaturation at 96°C; 30sec annealing at 60°C and 30sec extension at 72°C, then a melting curve from 60-95°C with 10sec holds at each 0.5°C interval. Triplicate 20μL reactions were performed with 5ng of cDNA (2μL of 1/20 dilutions) and 0.1μM (*TBP, H4*), 0.2μM (*CDKA, por*2, *RPS*), or 0.3μM (*por1*) primer. A 4-fold dilution series from 100ng to 0.024ng cDNA was amplified simultaneously for the determination of reaction efficiency as per [71]. Negative controls were included to ensure that PCR reagents were not contaminated. Data were analyzed with the Bio-Rad Gene Expression Analysis macro which incorporates primer amplification efficiency and housekeeping gene expression according to the method of Vandesompele et al. [72].

Antibody preparation

3' RACE of *por*1 and *por*2 cDNAs: A subsample of *P. tricornutum* 1L cultures growing exponentially at 100μmol photons m^{-2} s^{-1} and 20°C on a 12 hr light:12 hr dark photoperiod were harvested at ~ L7 by centrifugation at 5,000x*g* for 15 minutes at 4°C. The recovered pellets were flash frozen in liquid nitrogen. RNA was extracted, treated with DNAse, and purified as described above. cDNA synthesis was carried out using the SuperScript III First-Strand Synthesis System (Life Technologies) using 1.2μg RNA per 20μL reaction. The following reaction was used to amplify *por1* and *por2* cDNAs: 1x Phusion HF buffer and 0.02 U/μL Phusion High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA), 200 μM dNTPs (Lamda Biotech, St. Louis, MO), 0.5μM each forward and reverse primers, and 1μL cDNA (from the above reverse transcription) in 25μL reactions. A universal primer that anneals to the polyA tail of the mRNA ([73]; Supplemental table 1) or nested 5' primers that were designed

based on the 5' regions of *por1* and *por2* ESTs from Maheswari et al. ([74]; Supplemental table 1) were used in the cDNA synthesis reactions. Cycling reactions were performed in an Eppendorf Mastercycler gradient thermocycler (Hauppage, NY) as follows: initial denaturation was at 98°C for 30 seconds, followed by 35 cycles of 10 seconds denaturation at 98°C, 20 seconds annealing at 56°C, and a 1 minute extension at 72°C, then 5 minute final elongation at 72°C. Prior to sequencing, reactions were further treated by the addition of Exonuclease I (0.2U)/shrimp alkaline phosphatase (0.08U) and incubated at 37°C for 45min, then 85°C for 15min (Affymetrix, Santa Clara, CA). Sequencing was performed on an ABI 3130xl Genetic Analyzer using the ABI BigDye Terminator v3.1 Cycle Sequencing kit with $1/8^{th}$ the manufacturer's recommended reaction size (Applied BioSystems, Foster City, CA).

P. tricornutum POR1 and POR2 heterologous expression constructs: *por1* and *por2* cDNAs were re-amplified using Phusion High-Fidelity Polymerase and primers designed to enable Gibson Assembly (Gibson et al. 2009; Supplemental table 1), according to PCR protocols described above. The pET-15-HE vector (obtained from Dr. Stoddard, Seattle Children's Hospital, Seattle, WA) was digested by incubation at 37°C for 2hr in a 40μL reaction containing 3μg vector, 7.5U each of NcoI, NotI enzymes and 1X NE Buffer 3 (New England Biolabs). The linearized vector and PCR inserts were each purified with the QIAquick PCR Purification kit (Qiagen). Digested vector (50ng) and either *por1* or *por2* PCR inserts (50ng) at a ratio of ~1M vector:3M insert were incubated with 20μL Gibson Assembly MasterMix (New England Biolabs) at 50°C for 1 hour. A reaction containing 50μL competent *Escherichia coli* DH5α (Life Technologies) and 5μL of the Gibson Assembly product were incubated on ice for 30min, heat shocked at 42°C for 45 seconds, then cooled on ice for 2 minutes. Luria Broth (500μL) was added to the reaction

mixtures prior to incubation at 37°C for 1 hour with shaking at 200rpm. Transformed cells were spread on Luria broth agar plates (1.5% Bacto agar; Becton Dickinson & Co., Franklin Lakes, NJ) containing 100μg/mL carbenicillin (Sigma-Aldrich) and grown overnight at 37°C. Individual colonies were picked and suspended into 5mL of Luria Broth containing 100μg/mL carbenicillin and grown at 37°C overnight with shaking at 200rpm. The QIAprep Spin MiniPrep Kit was used to purify plasmid DNA (Qiagen). GENEWIZ DNA sequencing services (Seattle, WA) were used to sequence plasmid DNAs using an upstream T7 and an internal primer (Supplemental table 1) to verify sequence identity and orientation. The above protocol was then used to transform and plate *Escherichia coli* c2566 (New England Biolabs) with each plasmid.

POR1 and POR2 heterologous expression: Protein expression in *E. coli* c2566 was found to be leaky and growth overnight at 30°C without IPTG induction produced copious quantities of POR1 and POR2 proteins (Fig. 3). Overnight cultures were centrifuged at 2000xg for 20min at 4°C, and re-suspended in 10% the original culture volume of lysis buffer [PBS (50mM sodium phosphate, 300mM sodium chloride, pH 7) containing 0.1% Triton-X and 5mMβ-mercaptoethanol and cOmplete ULTRA EDTA-free protease inhibitors (1 tablet per 10mL; Roche, Nutley, NJ)]. Additionally, 6M guanidine hydrochloride was used to ensure solubility of POR2. Upon addition of 1mg/mL lysozyme (Sigma-Aldrich, St. Louis, MO) and incubation on ice for 30 minutes, the lysates were sonicated at 4°C for 3 minutes in 10-second intervals with 30-second pauses. Supernatants containing soluble POR1 or POR2 enzymes were recovered after centrifugation at 12,000xg for 25 min at 4°C.

Affinity purification: POR1 and POR2 enzymes were subject to affinity purification using the 6X histidine tag encoded in the pET-15-HE vector (Fig. 3). The Clontech TALON Metal Affinity Resin system was used according to manufacturer's protocols using 2mL resin per 25mL lysis buffer containing bacterial cell pellet from 250mL culture (Mountain View, CA). Fractions were evaluated by SDS-PAGE, and those containing purified protein were concentrated in Millipore Amicon Ultra centrifugal filter units per manufacturer's directions.

Antibody preparation: Antibodies were prepared by Yenzym (San Francisco, CA). Anti-POR1 antibody was generated in a chicken and anti-POR2 antibody raised in a pre-screened rabbit. The POR1 antibody was affinity purified using ThermoScientific Pierce NHS-Activated Agarose slurry according to manufacturer's directions, with 6mg POR1 protein coupled to 2mL slurry (Fig. 3). The specificity of the POR2 antibody was verified by comparing the cross-reactivity of un-blocked antibody with that of antibody blocked by incubation at 4°C overnight with 30X (w/w) POR2 affinity purified protein (Fig. 3).

POR1 and POR2 protein quantitation

Protein extraction and quantitation: Triplicate protein extractions were performed for each time point according to the methanol/chloroform/water method of Wessel and Flugge [75]. Unless indicated, all procedures were performed at room temperature. Briefly, each extraction used pellets from two 45mL samples. One mL methanol was added to ~250μL of loose pellet to attain an approximately 80% methanol.

Upon vortexing on high for 60sec, the contents of two tubes were transferred to one 15mL conical centrifuge tube. Chloroform (500uL) was added to the combined solutions, which were vortexed for 60sec at RT. Water was added (1.5mL) and the sample vortexed another 30sec. Phases were separated by centrifugation for 5min at RT. After removal of the upper chloroform phase, the protein containing interface and bottom phase were transferred to a 2mL centrifuge tube. A 1.5mL methanol rinse of the extraction tube was then added to the 2mL tube. The solution was vortexed for 10 sec. Precipitated protein was recovered by centrifugation at 14,600xg for 10min at 4°C. After removal of the supernatant, the pellet was dried by SpeedVac for 5min. Samples were resuspended in 150uL DIGE buffer (7M urea, 2M thiourea, 30mM Tris-base, 4% CHAPS, pH 8.5) that contained Roche cOmplete ULTRA EDTA-free protease inhibitors (1 tablet per 10mL). After the samples were gently vortexed for 1 hr., non-solubilized material was removed by centrifugation at 14,600xg for 12min at 4°C. Protein in the supernatant was quantitated using the Life Technologies EZQ protein extraction kit. Preliminary experiments showed that under these re-solubilization conditions, equal protein extraction efficiency was attained for samples regardless of the upper and lower cell concentration used in these studies.

Western blotting: For each time point, 4μg protein was denatured for 10min at 70°C with 1X NuPage LDS sample buffer and 1X NuPage reducing agent (Life Technologies). Proteins were separated on pre-cast NuPage Novex 4-12% Bis-Tris protein gels (Life Technologies) run in MOPS buffer (50mM MOPS, 50mM Tris-base, 0.1% SDS, 1mM EDTA). SeeBlue Plus2 pre-stained protein standard (Life Technologies) served as molecular markers for protein size determination. Proteins were transferred to nitrocellulose (POR1) or PVDF (POR2) membranes using the iBlot gel transfer device and transfer stacks (Life Technologies). After blocking for 1hr

at 20°C with 5% nonfat milk in TBST buffer (20mM Tris, 150mM NaCl, pH 7.6, 0.1% Tween-20), blots were incubated overnight at 4°C with anti-POR1 (1:600 in TBST with 2% milk) or anti-POR2 (1:7500 in TBST with 2% milk) antibodies. After three 10min washes in TBST, blots were incubated for 1hr at 20°C with HRP-conjugated anti-chicken antibodies (1:20,000 in TBST with 2% milk; Abcam, Cambridge, MA) or HRP-conjugated anti-rabbit antibodies (1:6000 in TBST; GE Amersham, Pittsburgh, PA). Following two 10min washes in TBST and one 10min wash in TBS (TBST without Tween-20), the blots were incubated with SuperSignal West-Pico chemiluminescent substrate (Life Technologies) and visualized for 2min (POR1) or 30sec (POR2) on X-ray film (Phenix Research, Candler, NC). Band intensities were quantitated in ImageJ [76] for each of three to four replicate blots. As each blot contained all time points, data were normalized within each blot such that the total density across time points summed to 100. Membranes were stained with Coomassie Brilliant Blue to check for equal loading.

Statistical analysis

Diel cycling of POR2 abundance per μg protein and per cell was analyzed with paired Welch's unequal variance t-tests. Light interval means were paired with dark interval means from $_{50}$L/D, $_{200}$L/D, and $_{1200}$L/D, and the analysis was repeated when also including the $_{200}$sD$_3$/$_{200}$L$_4$ pair. A two-way ANOVA was performed to compare POR1 abundance per μg protein in the light and dark intervals of $_{1200}$L/D and $_{50}$L/D. A one-way ANOVA followed by a Tukey's HSD test was used to compare POR1 abundance per μg protein among all time points in $_{1200}$L/D and $_{50}$L/D.

ACKNOWLEDGMENTS

We thank John Munch for designing and constructing a fully programmable LED controller and LED panels. HMH was supported by the NSF GRFP (DGE-0718124; DGE-1256082) as well as a University of Washington Department of Biology Plant Biology Fellowship and a Melinda Denton Writing Fellowship. This research was funded 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. RAC dedicates this manuscript to the many outstanding undergraduate students who have done research in her laboratory.

LITERATURE CITED

1. Suzuki JY, Bollivar DW, Bauer CE: **Genetic analysis of chlorophyll biosynthesis.** Annu Rev Genet 1997, **31**:61–89.

2. Fujita Y, Bauer CE: **The light-independent protochlorophyllide reductase: a nitrogenase-like enzyme catalyzing a key reaction for greening in the dark**. In Porphyr Handb. Volume 13. Edited by Kadish K, Smith K, Guilard R. Elsevier Science; 2003:109–156.

3. Hunsperger HM, Randhawa T, Cattolico RA: **Extensive horizontal gene transfer, duplication, and loss of chlorophyll synthesis genes in the algae**. BMC Evol Biol 2015, **15**.

4. 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.

5. 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. Yale University Press; 1970.

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

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

8. Parfrey LW, Lahr DJG, Knoll AH, Katz L a: **Estimating the timing of early eukaryotic diversification with multigene molecular clocks.** Proc Natl Acad Sci U S A 2011, **108**:13624–9.

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

10. Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, De Clerck O: **Phylogeny and Molecular Evolution of the Green Algae**. CRC Crit Rev Plant Sci 2012, **31**:1–46.

11. Yoon HS, Hackett JD, Pinto G, Bhattacharya D: **The single, ancient origin of chromist plastids.** Proc Natl Acad Sci U S A 2002, **99**:15507–15512.

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

13. Zhang J: **Evolution by gene duplication: an update**. Trends Ecol Evol 2003, **18**:292–298.

14. Armstrong G, Runge S, Frick G, Sperling U, Apel K: **Identification of NADPH:protochlorophyllide oxidoreductases A and B: a branched pathway for light-dependent chlorophyll biosynthesis in *Arabidopsis thaliana*.** Plant Physiol 1995, **108**:1505–17.

15. Su Q, Frick G, Armstrong G, Apel K: **POR C of *Arabidopsis thaliana*: a third light- and NADPH-dependent protochlorophyllide oxidoreductase that is differentially regulated by light**. Plant Mol Biol 2001, **47**:805–813.

16. Holtorf H, Reinbothe S, Reinbothe C, Bereza B, Apel K: **Two routes of chlorophyllide synthesis that are differentially regulated by light in barley (*Hordeum vulgare* L.).** Proc Natl Acad Sci U S A 1995, **92**:3254–8.

17. Ragni M, D'Alcalà MR: **Circadian variability in the photobiology of *Phaeodactylum tricornutum*: pigment content**. J Plankton Res 2007, **29**:141–156.

18. 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**:438–448.

19. Anning T, MacIntyre HL, Pratt SM, Sammes PJ, Gibb S, Geider RJ: **Photoacclimation in the marine diatom *Skeletonema costatum***. Limnol Oceanogr 2000, **45**:1807–1817.

20. MacIntyre HL, Kana TM, Anning T, Geider RJ: **Photoacclimation of photosynthesis irradiance response curves and photosynthetic pigments in microalgae and cyanobacteria**. J Phycol 2002, **38**:17–38.

21. Deblois CP, Marchand A, Juneau P: **Comparison of photoacclimation in twelve freshwater photoautotrophs (chlorophyte, bacillaryophyte, cryptophyte and cyanophyte) isolated from a natural community**. PLoS One 2013, **8**.

22. Gschloessl B, Guermeur Y, Cock JM: **HECTAR: a method to predict subcellular targeting in heterokonts.** BMC Bioinformatics 2008, **9**:393.

23. Lang M, Apt KE, Kroth PG: **Protein transport into "complex" diatom plastids utilizes two different targeting signals.** J Biol Chem 1998, **273**:30973–8.

24. Stork S, Lau J, Moog D, Maier U-G: **Three old and one new: protein import into red algal-derived plastids surrounded by four membranes.** Protoplasma 2013, **250**:1013–23.

25. Gibbs SP: **The route of entry of cytoplasmically synthesized proteins into chloroplasts of algae possessing chloroplast ER.** J Cell Sci 1979, **35**:253–266.

26. Kilian O, Kroth PG: **Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids.** Plant J 2005, **41**:175–83.

27. Birve S, Selstam E, Johansson B: **Secondary structure of NADPH: protochlorophyllide oxidoreductase examined by circular dichroism and prediction methods**. Biochem J 1996, **317**:549–555.

28. 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**:724–728.

29. Lebedev N, Karginova O, McIvor 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**:12562–12574.

30. 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.

31. 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**:18160–18166.

32. 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**:2113–2119.

33. 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**:161–176.

34. Belyaeva OB, Litvin FF: **Photoactive pigment—enzyme complexes of chlorophyll precursor in plant leaves**. Biochem 2007, **72**:1458–1477.

35. **Joint Genome Institute: Genome Portal** [http://genomeportal.jgi.doe.gov/]

36. Aronsson H, Sundqvist C, Timko MP, Dahlin C: **The importance of the C-terminal region and Cys residues for the membrane association of the NADPH:protochlorophyllide oxidoreductase in pea.** FEBS Lett 2001, **502**:11–5.

37. Martino A De, Meichenin A, Shi J, Pan K, Bowler C: **Genetic and phenotypic characterization of *Phaeodactylum tricornutum* (Bacillariophyceae) accessions**. J Phycol 2007, **43**:992–1009.

38. Siaut M, Heijde M, Mangogna M, Montsant A, Coesel S, Allen A, Manfredonia A, Falciatore A, Bowler C: **Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum.** Gene 2007, **406**:23–35.

39. Valle KC, Nymark M, Aamot I, Hancke K, Winge P, Andresen K, Johnsen G, Brembu T, Bones AM: **System responses to equal doses of photosynthetically usable radiation of blue, green, and red light in the marine diatom *Phaeodactylum tricornutum***. PLoS One 2014, **9**:e114211.

40. Bell-Pedersen D, Cassone VM, Earnest DJ, Golden SS, Hardin PE, Thomas TL, Zoran MJ: **Circadian rhythms from multiple oscillators: lesson from diverse organisms**. Nat Rev Genet 2005, **6**:544–556.

41. 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.

42. Putt M, Prezelin BB: **Diel periodicity of photosynthesis and cell division compared in *Thalassiosira weissflogii* (Bacillariophyceae)**. J Phycol 1988, **24**:315–324.

43. Wilhelm C, Jungandreas A, Jakob T, Goss R: **Light acclimation in diatoms: from phenomenology to mechanisms.** Mar Genomics 2014, **16**:5–15.

44. Nymark M, Valle KC, Brembu T, Hancke K, Winge P, Andresen K, Johnsen G, Bones AM: **An integrated analysis of molecular acclimation to high light in the marine diatom *Phaeodactylum tricornutum*.** PLoS One 2009, **4**:e7743.

45. Nymark M, Valle KC, Hancke K, Winge P, Andresen K, Johnsen G, Bones AM, Brembu T: **Molecular and photosynthetic responses to prolonged darkness and subsequent acclimation to re-illumination in the diatom *Phaeodactylum tricornutum*.** PLoS One 2013, **8**:e58722.

46. McQuoid MR, Hobson LA: **Diatom resting stages**. J Phycol 1996, **32**:889–902.

47. Falciatore A, Bowler C: **The evolution and function of blue and red light photoreceptors**. Curr Top Dev Biol 2005, **68**:317–350.

48. Oliveri P, Fortunato AE, Petrone L, Ishikawa-Fujiwara T, Kobayashi Y, Todo T, Antonova O, Arboleda E, Zantke J, Tessmar-Raible K, Falciatore A: **The cryptochrome/photolyase family in aquatic organisms.** Mar Genomics 2014:1–15.

49. Takahashi F, Yamagata D, Ishikawa M, Fukamatsu Y, Ogura Y, Kasahara M, Kiyosue T, Kikuyama M, Wada M, Kataoka H: **AUREOCHROME, a photoreceptor required for photomorphogenesis in stramenopiles.** Proc Natl Acad Sci U S A 2007, **104**:19625–30.

50. Schellenberger Costa B, Sachse M, Jungandreas A, Bartulos CR, Gruber A, Jakob T, Kroth PG, Wilhelm C: **Aureochrome 1a is involved in the photoacclimation of the diatom *Phaeodactylum tricornutum*.** PLoS One 2013, **8**:e74451.

51. Heyes DJ, Kruk J, Hunter CN: **Spectroscopic and kinetic characterization of the light-dependent enzyme protochlorophyllide oxidoreductase (POR) using monovinyl and divinyl substrates.** Biochem J 2006, **394**:243–8.

52. De Riso V, Raniello R, Maumus F, Rogato A, Falciatore A: **Gene silencing in the marine diatom *Phaeodactylum tricornutum***. Nucleic Acids Res 2009, **37**.

53. Trentacoste EM, Shrestha RP, Smith SR, Glé C, Hartmann AC, Hildebrand M, Gerwick WH: **Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth.** Proc Natl Acad Sci U S A 2013, **110**:19748–53.

54. Daboussi F, Leduc S, Maréchal A, Dubois G, Guyot V, Perez-Michaut C, Amato A, Falciatore A, Juillerat A, Beurdeley M, Voytas DF, Cavarec L, Duchateau P: **Genome engineering empowers the diatom *Phaeodactylum tricornutum* for biotechnology.** Nat Commun 2014, **5**(May):3831.

55. Jörnvall H, Persson B, Krook M, Atrian S, Gonzalez-Duarte R, Jeffery J, Ghosh D: **Short-chain dehydrogenases/reductases (SDR)**. Biochemistry 1995, **34**:6003–6013.

56. 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**:3895–3906.

57. Martin GE, Timko MP, Wilks HM: **Purification and kinetic analysis of pea (*Pisum sativum* L.) NADPH:protochlorophyllide oxidoreductase expressed as a fusion with maltose-binding protein in Escherichia coli.** Biochem J 1997, **325**:139–45.

58. Yuan M, Zhang D-W, Zhang Z-W, Chen Y-E, Yuan S, Guo Y-R, Lin H-H: **Assembly of NADPH: protochlorophyllide oxidoreductase complex is needed for effective greening of barley seedlings.** J Plant Physiol 2012, **169**:1311–6.

59. Spano A, He Z, Michel H, Hunt D, Timko M: **Molecular cloning, nuclear gene structure, and developmental expression of NADPH: protochlorophyllide oxidoreductase in pea (*Pisum sativum* L.)**. Plant Mol Biol 1992, **18**:967–972.

60. Dehesh K, Apel K: **The function of proteases during the light-dependent transformation of etioplasts to chloroplasts in barley (*Hordeum vulgare* L.)**. Planta 1983, **157**:381–383.

61. Reinbothe C, Apel K, Reinbothe S: **A light-induced protease from barley plastids degrades NADPH:protochlorophyllide oxidoreductase complexed with chlorophyllide.** Mol Cell Biol 1995, **15**:6206–12.

62. Tanaka R, Tanaka A: **Tetrapyrrole biosynthesis in higher plants.** Annu Rev Plant Biol 2007, **58**:321–346.

63. Wang P, Grimm B: **Organization of chlorophyll biosynthesis and insertion of chlorophyll into the chlorophyll-binding proteins in chloroplasts**. Photosynth Res 2015.

64. Reinbothe S, Reinbothe C, Apel K, Lebedev N: **Evolution of chlorophyll biosynthesis—the challenge to survive photooxidation**. Cell 1996, **86**:703–705.

65. Schoefs B, Franck F: **Protochlorophyllide reduction: mechanisms and evolution**. Photochem Photobiol 2003, **78**:543–557.

66. Masuda T, Fusada N, Shiraishi T, Kuroda H, Awai K, Shimada H, Ohta H, Takamiya K-I: **Identification of two differentially regulated isoforms of protochlorophyllide oxidoreductase (POR) from tobacco revealed a wide variety of light- and development-dependent regulations of POR gene expression among angiosperms**. Photosynth Res 2002, **74**:165–172.

67. Skinner JS, Timko MP: **Differential expression of genes encoding the light-dependent and light-independent enzymes for protochlorophyllide reduction during development in loblolly pine**. Plant Mol Biol 1999, **39**:577–592.

68. **Marine Microbial Eukaryote Transcriptome Sequencing Project** [http://marinemicroeukaryotes.org/]

69. Guillard R, Rhyther J: **Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (cleve) Gran.** Can J Microbiol 1962, **8**:229–239.

70. Coesel S, Mangogna M, Ishikawa T, Heijde M, Rogato A, Finazzi G, Todo T, Bowler C, Falciatore A: **Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity**. EMBO Rep 2009, **10**:655–661.

71. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR.** Nucleic Acids Res 2001, **29**:2002–7.

72. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F: **Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.** Genome Biol 2002, **3**:RESEARCH0034.

73. Scotto-Lavino E, Du G, Frohman M a: **5' end cDNA amplification using classic RACE.** Nat Protoc 2006, **1**:2555–62.

74. Maheswari U, Jabbari K, Petit J-L, Porcel BM, Allen AE, Cadoret J-P, De Martino A, Heijde M, Kaas R, La Roche J, Lopez PJ, Martin-Jézéquel V, Meichenin A, Mock T, Schnitzler Parker M, Vardi A, Armbrust EV, Weissenbach J, Katinka M, Bowler C: **Digital expression profiling of novel diatom transcripts provides insight into their biological functions.** Genome Biol 2010, **11**:R85.

75. Wessel D, Flügge UI: **A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids.** Anal Biochem 1984, **138**:141–143.

76. **ImageJ** [http://imagej.nih.gov/ij/]

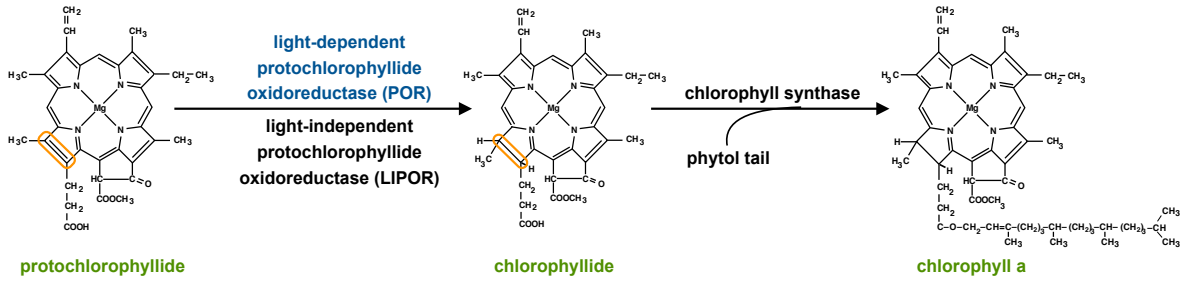


Figure 1: The transformation of the chlorophyll pigment precursor protochlorophyllide into chlorophyllide can be catalyzed by either the light-dependent (POR) or light-independent (LIPOR) protochlorophyllide oxidoreductases. Upon the addition of a phytol tail to chlorophyllide by chlorophyll synthase, chorophyll *a* synthesis is completed.

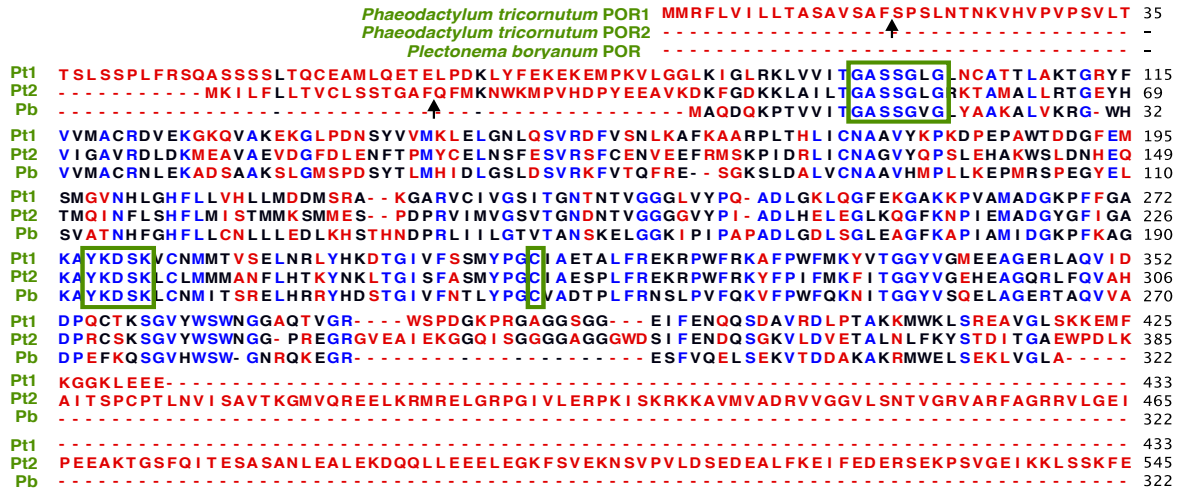


Figure 2. Multiple sequence alignment of diatom *P. tricornutum* and cyanobacterium *Plectonema boryanum* POR proteins. Color indicates sequence amino acid conservation across the three PORs: highly conserved (blue), semi-conserved (black), and poorly conserved (red). Green boxes indicate the N-terminal Rossman fold (GxxxGxG) for NADPH-binding [27], YxxxK active-site whose tryptophan (Y) donates a protein to Pchlide during catalysis [28–31], as well as the universally conserved cysteine (C) residue indispensable to proper POR enzyme conformation [32]. Arrows indicate signal peptide cleavage sites. The C-terminal extension of POR2 is not present in the mature protein.

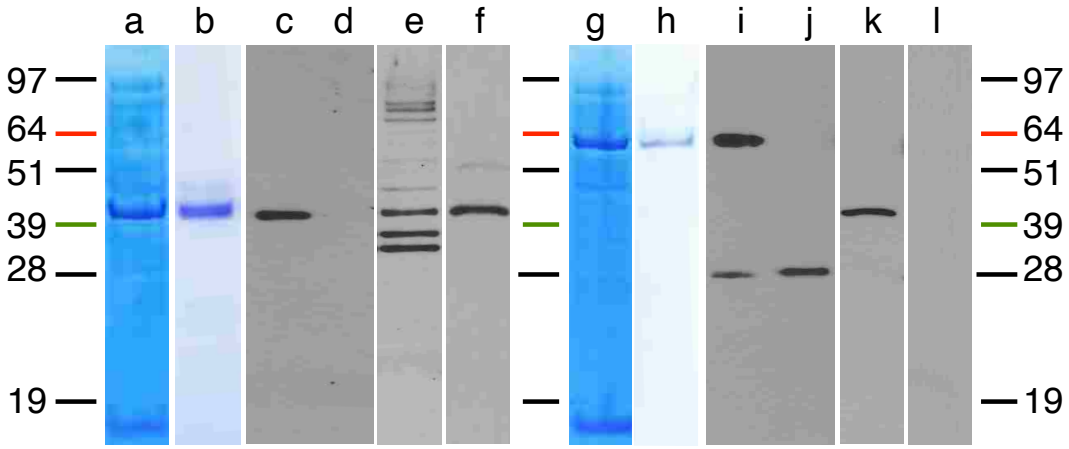


Figure 3: Preparation of anti-*P. tricornutum* POR1 and anti-*P. tricornutum* POR2 antibodies. Heterologous expression (a) and affinity-purification (b) of POR1 protein for antibody production. The anti-POR1 antibody cross-reacted with affinity-purified POR1 protein (c), but not POR2 protein (d). Reactivity of the anti-POR1 antibody to *P. tricornutum* extracts prior to (e) and after (f) affinity-purification of the antibody against POR1 protein. Heterologous expression (g) and affinity-purification (h) of POR2 protein for antibody production. The anti-POR2 antibody cross-reacted with affinity-purified POR2 protein (i), but not POR1 protein (j). Reactivity of the anti-POR2 antibody to *P. tricornutum* extracts without (k) and with (l) 30X (w/w) blocking of the anti-POR2 antibody with affinity-purified POR2 antigen.

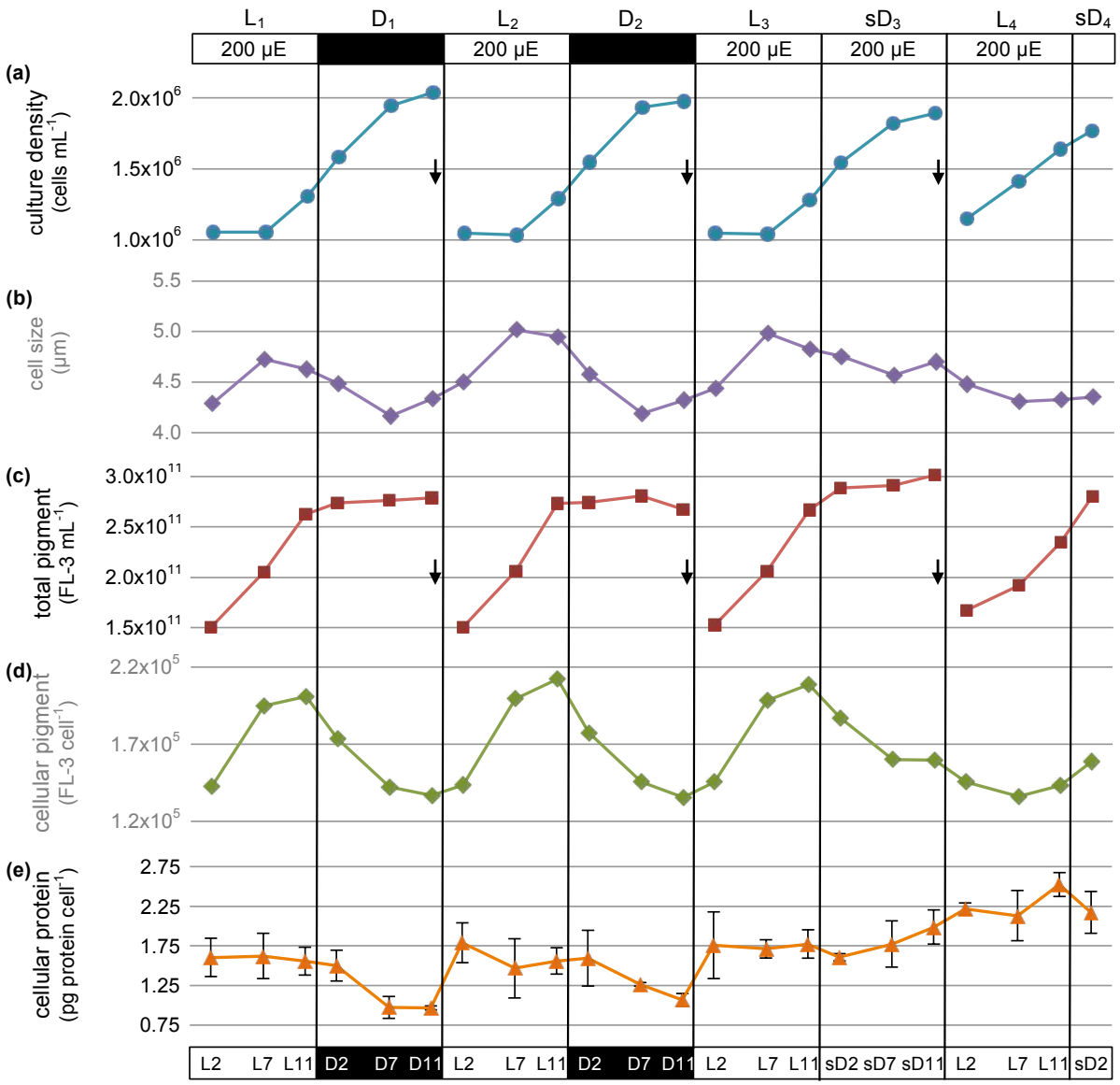


Figure 4: *P. tricornutum* growth and physiology under 200µE $m^{-2}s^{-1}$ 12L:12D ($_{200}$L/D) and transition to constant illumination ($_{200}$L/L). Culture acclimated to $_{200}$L/D was sampled for two days prior to and 1.6 days after a transition to $_{200}$L/L. Arrows indicate daily dilution with fresh medium at L0 to ~1.0 x10^6 cells/mL. (a) Culture density (cells/mL), (b) size (µm), (c) total pigment (FL-3 fluorescence/mL), (d) cellular pigment (FL-3 fluorescence/cell), and (e) cellular protein (pg/cell; error bars show standard deviation). The culture was acclimated for 4.2 generations (4 days) under $_{200}$L/D before sampling began.

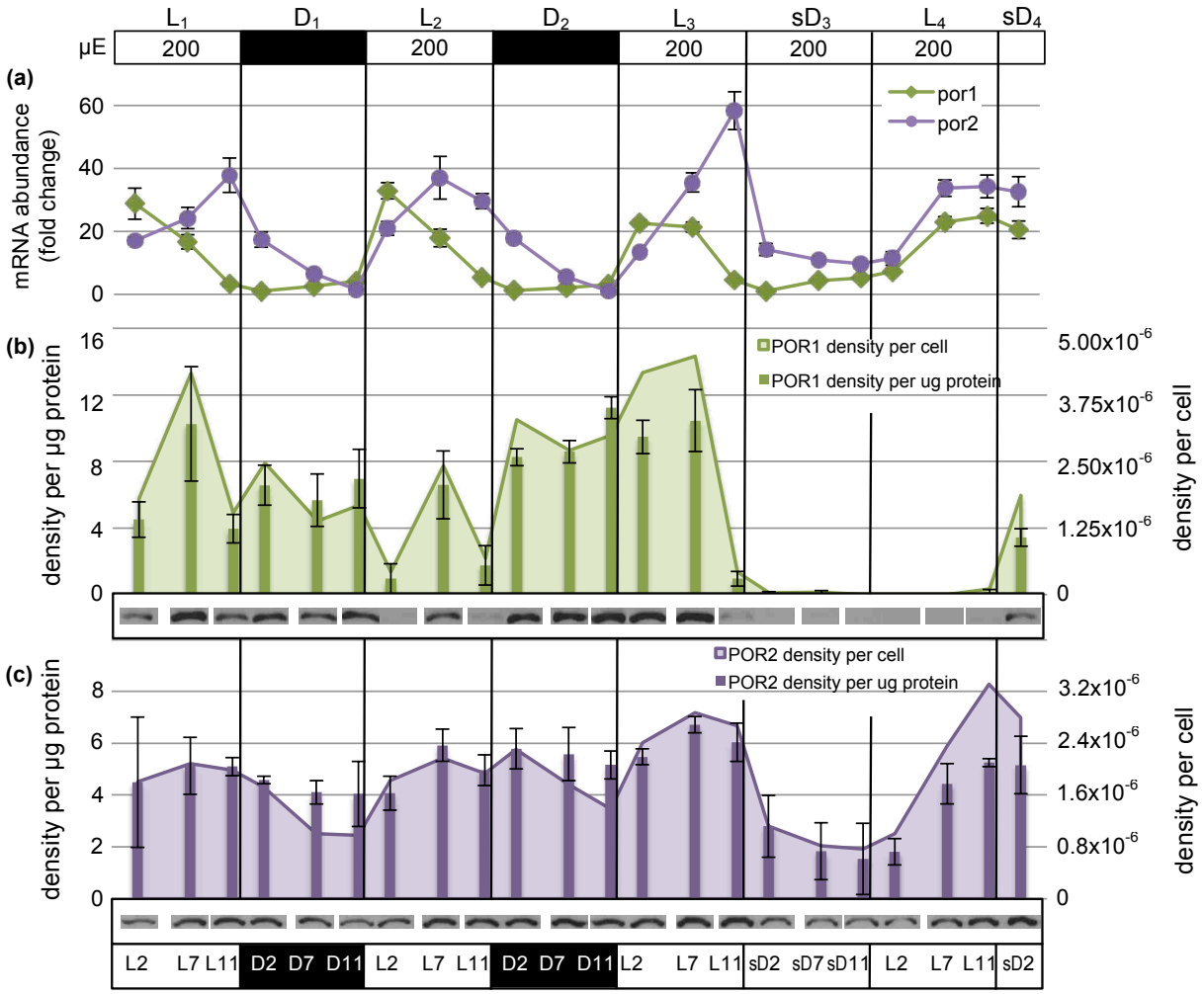


Figure 5: *Por* mRNA transcript and POR protein abundance in *P. tricornutum* grown under $_{200}$L/D and transitioned to a $_{200}$L/L light regime. (a) Fold-change in *por*1 and *por*2 mRNA transcript abundance measured by RT-qPCR. (b) POR1 density per µg protein measured by Western blotting (bars; error bars show standard deviation) and normalized to cellular proteins levels to attain POR1 density per cell (area chart). Representative blot data is shown below each sampled time point. (c) As in (b) but for POR2. Light and dark interval POR2 abundances *per cell* were significantly different when analyzed for the $_{200}$L/D, $_{1200}$L/D, and $_{50}$L/D regimes ($p=0.003$), whereas no difference was detected at the level of POR2 abundance *per µg* protein ($p=0.250$) (all three light regimes showed a $_{200}$L/D-like pattern for POR2 per µg protein as well as POR2 per cell; Fig. 7). Incorporation of $_{200}L_3$ and $_{200}sD_3$ into the analysis of *cellular* POR2 levels under $_{50}$L/D, $_{200}$L/D and $_{1200}$L/D increased statistical significance of the diel oscillatory response to $p=0.00087$. Data obtained from studies using the 12L photobioreactor and preliminary data obtained using small, 1L batch-grown (i.e., non-diluted) cultures showed similar photoperiodic cell responses, suggesting that observed results were light/dark cycle induced rather than cued by an L0 nutrient influx in the photobioreactor.

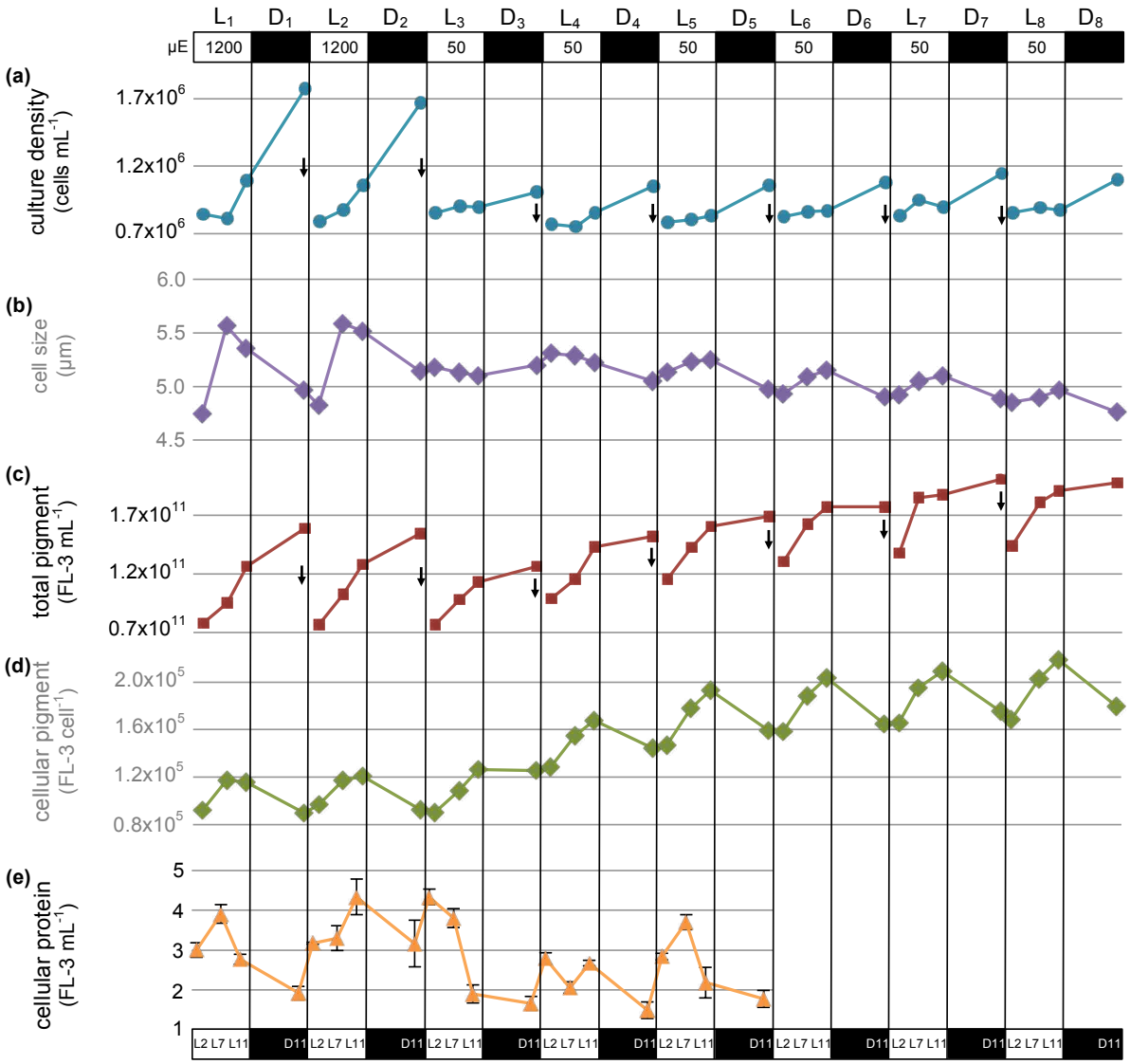


Figure 6: *P. tricornutum* growth and physiology under 1200µE $m^{-2}s^{-1}$ 12L:12D ($_{200}$L/D) and transition to 50µE $m^{-2}s^{-1}$ ($_{50}$L/D). Culture acclimated to $_{1200}$L/D were sampled two days prior to and three days after a transition to $_{50}$L/D, and then monitored for an additional three days. Arrows indicate daily dilution with fresh medium at L0 to ~0.8 x10^7 cells/mL. (a) Culture density (cells/mL), (b) size (µm), (c) total pigment (FL-3 fluorescence/mL), (d) cellular pigment (FL-3 fluorescence/cell), and (e) cellular protein (pg/cell; error bars show standard deviation). The culture was acclimated for 8.5 generations (5 days) under $_{1200}$L/D before beginning sampling.

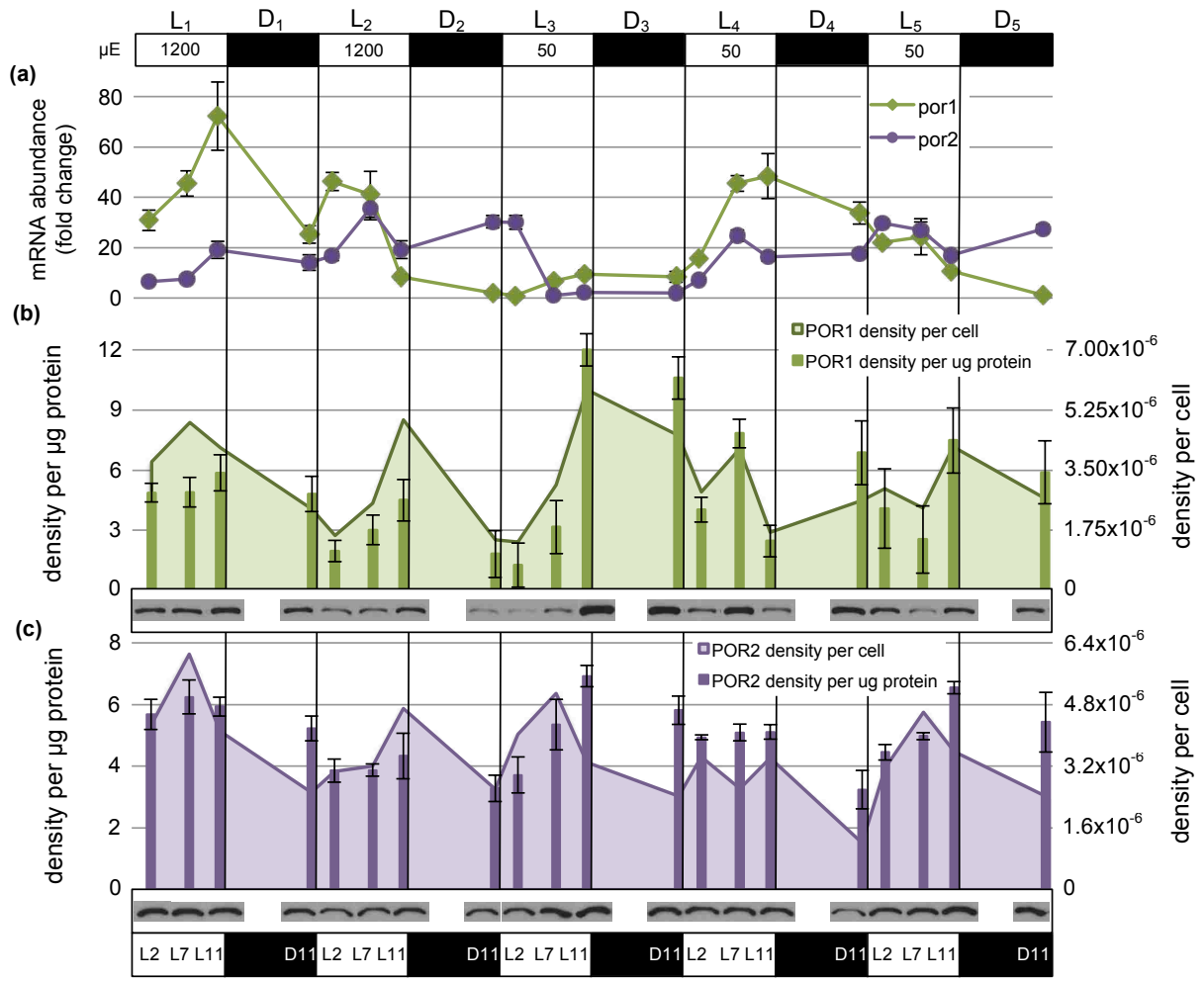
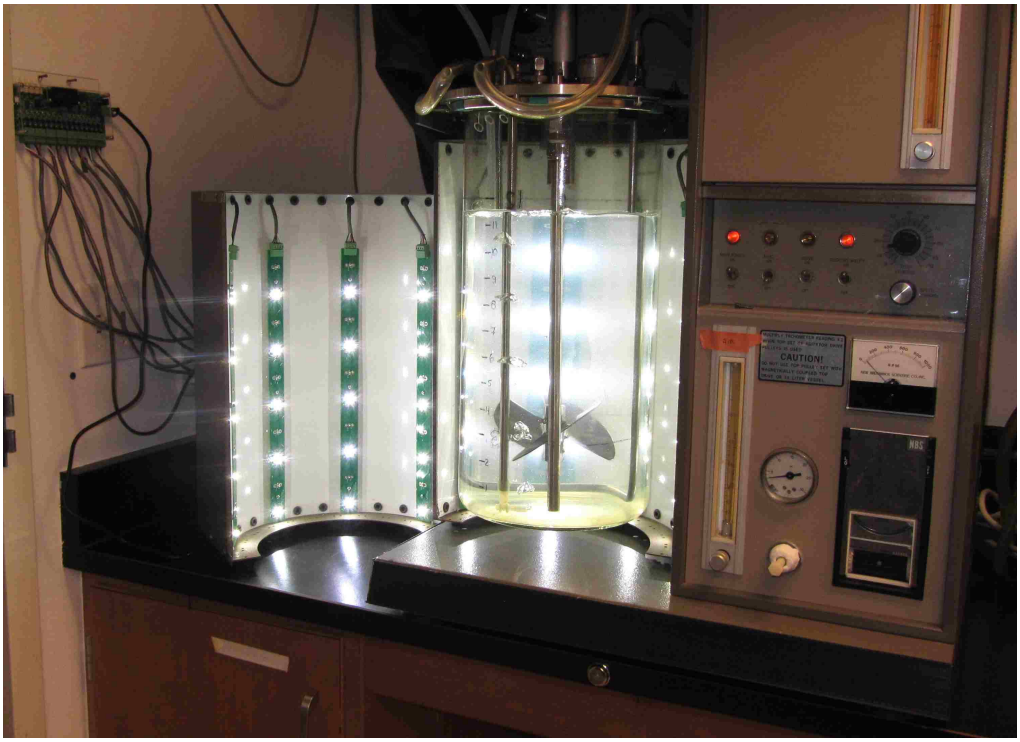
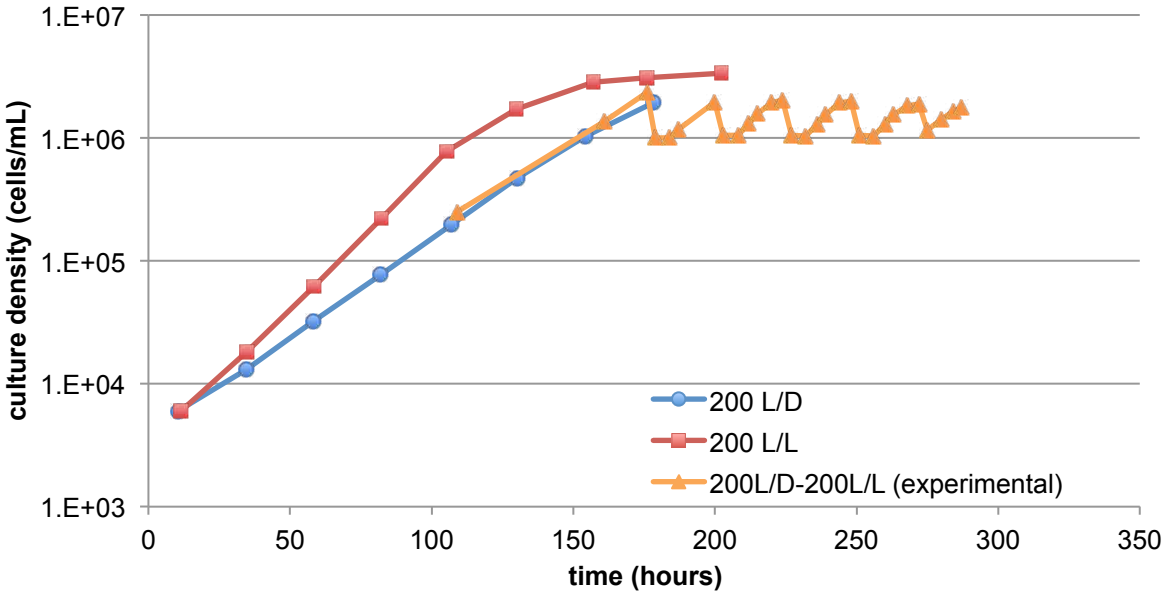


Figure 7: *Por* mRNA transcript and POR protein abundance in *P. tricornutum* grown under $_{1200}$L/D and transitioned to $_{50}$L/D. (a) Fold-change in *por*1 and *por*2 mRNA transcript abundance as measured by RT-qPCR. (b) POR1 density per µg protein as measured by Western blotting (bars; error bars show standard deviation) and normalized to cellular proteins levels to attain POR1 density per cell (area chart). Representative blot data is shown below each sampled time point. (c) As in (b) but for POR2.

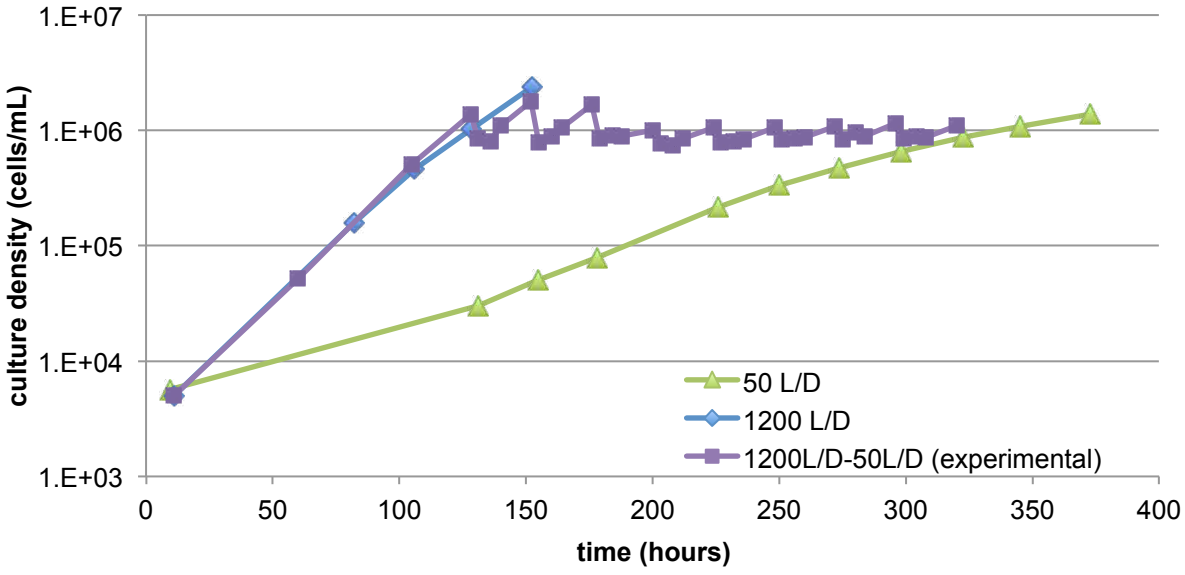
Supplemental figure 1: 15L photobioreactor with LED lighting.



Supplemental figure 2: Light intensity-dependent growth of *P. tricornutum*

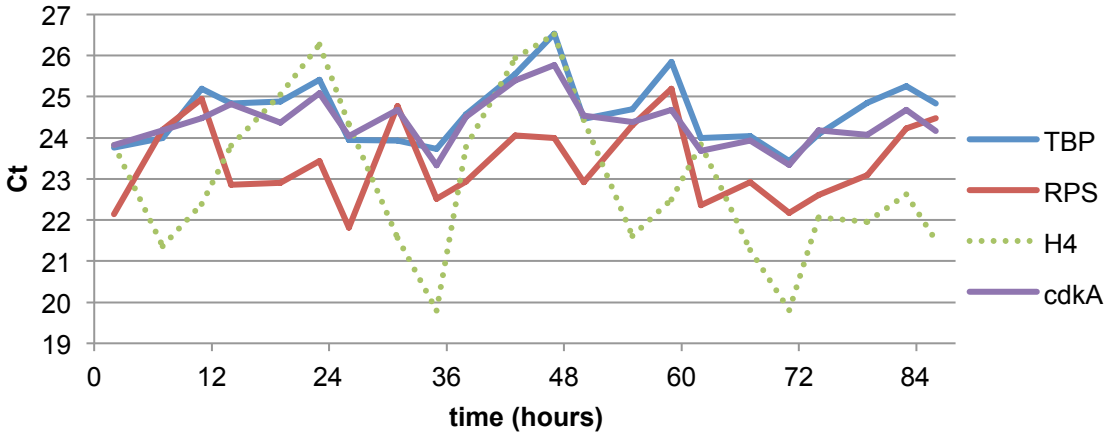


a. Comparison of culture growth for *P. tricornutum* at 200µE m^{-2} s^{-1} on a 12hr light:12hr dark regime ($_{200}$L/D), under constant illumination at 200µE m^{-2} s^{-1} ($_{200}$L/L), and under the experimental culture conditions of Figs. 4,5 ($_{200}$L/D-$_{200}$L/L).

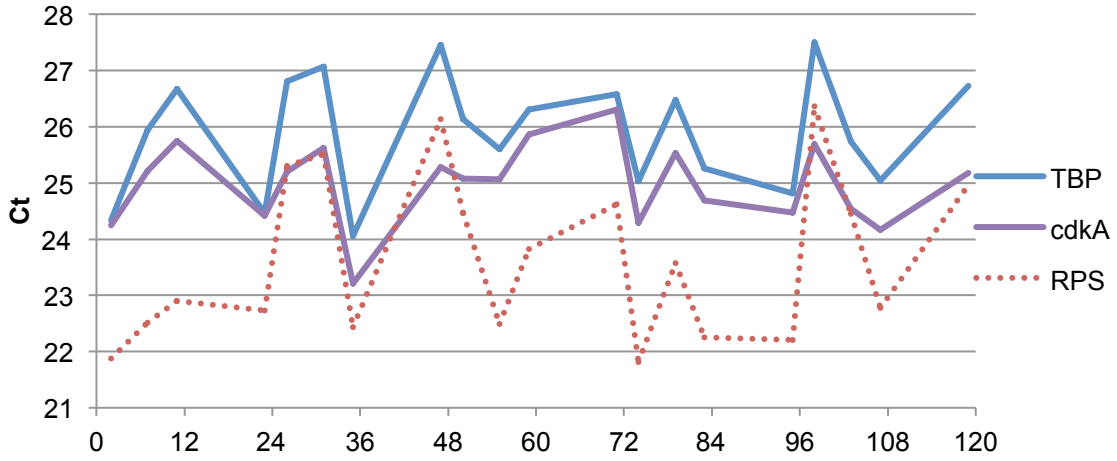


b. Comparison of culture growth for *P. tricornutum* at 50µE m^{-2} s^{-1} on a 12hr light:12hr dark regime ($_{50}$L/D), at 1200µE m^{-2} s^{-1} on a 12hr light:12hr dark regime ($_{1200}$L/D), and under the experimental culture conditions of Figs. 6,7 ($_{1200}$L/D-$_{50}$L/D).

Supplemental figure 3: Housekeeping gene expression



a. Housekeeping gene expression over the course of the $_{200}$**L/D:**$_{200}$**L/L** experiment.
ΔC_t TBP=3.47; ΔC_t RPS=3.38; ΔC_t H4=6.72; ΔC_t cdkA=2.93.
Data normalized to: TBP, RPS, and cdkA only. H4 not used.



b. Housekeeping gene expression over the course of the $_{1200}$**L/D:**$_{50}$**L/D** experiment.
ΔC_t TBP=3.45; ΔC_t RPS=4.55; ΔC_t cdkA=3.10.
Data normalized to: TBP and cdkA only. RPS not used.

Supplemental table 1: Primers used for RT-qPCR, 3' RACE and amplicon preparation for cloning.

RT-qPCR	**housekeeping and experimental genes**				
primer ID	*direction*	*sequence*	*anneal*	*description*	*citation*
por1_fw1	forward	5' AATGTCCATGGGTGTCAATCA 3'	60°C	*por*1	Coesel et al. 2009
por1_rv1	reverse	5' GTGGGTAGACGAGACCTCCAC 3'	60°C	*por*1	Coesel et al. 2009
por2_fw4	forward	5' ACAATCACGAACAGACCATGC 3'	60°C	*por*2	present study
por2_rv4	reverse	5' AGATCGGCAATAGGGTACACG 3'	60°C	*por*2	present study
cdkA_fw	forward	5' CGAAGTCGTTACCCTGTGGT 3'	60°C	cyclin-dependent kinase A	Siaut et al. 2007
cdkA_rv	reverse	5' CCAATTGATCGGCTTCAGAT 3'	60°C	cyclin-dependent kinase A	Siaut et al. 2007
H4_fw	forward	5' AGGTCCTTCGCGACAATATC 3'	60°C	histone	Siaut et al. 2007
H4_rv	reverse	5' ACGGAATCACGAATGACGTT 3'	60°C	histone	Siaut et al. 2007
RPS_fw	forward	5' CGAAGTCAACCAGGAAACCAA 3'	60°C	30S ribosomal protein subunit	Siaut et al. 2007
RPS_rv	reverse	5' GTGCAAGAGACCGGACATACC 3'	60°C	30S ribosomal protein subunit	Siaut et al. 2007
TBP_fw	forward	5' ACCGGAGTCAAGAGCACACAC 3'	60°C	TATA box binding protein	Siaut et al. 2007
TBP_rv	reverse	5' CGGAATGCGCGTATACCAGT 3'	60°C	TATA box binding protein	Siaut et al. 2007

3' RACE	**first strand synthesis**				
primer ID	*direction*	*sequence*	*anneal*	*description*	*citation*
rtp	reverse	5'-GCTCGCGAGCGCGTTTAAACGCGCACGC GTTTTTTTTTTTTTTTTTVN-3'	65°C	anneals to poly-A tail for reverse transcription	Scotto-Lavino et al. 2006
np1	reverse	5'-GCTCGCGAGCGCGTTTAAAC-3'	56°C	outer nested primer	Scotto-Lavino et al. 2006
np2	reverse	5'-GCGTTTAAACGCGCACGCGT-3'	56°C	inner nested primer	Scotto-Lavino et al. 2006
por1outA	forward	5' TCGGATTTAGTCATGATGCG 3'	56°C	outer nested primer	present study

por1inA	forward	5’ CGTTTCCTAGTGATTTTGCTG 3’	56°C	inner nested primer	present study
por2outA	forward	5’ TACTCACCGTGTGCTTGTCA 3’	56°C	outer nested primer	present study
por2inA	forward	5’ CAAGCACCGGGGCTTTCCA 3’	56°C	inner nested primer	present study

pET-15-HE amplicon preparation for cloning

primer ID	*direction*	*sequence*	*anneal*	*description*	*citation*
por1_heFWDs	forward	5' CTGGTGCCGCGCGGCAGCTCCATGCCCA AGGTATTGGGAGGT 3'	53-59°C	*por*1 amplicon	present study
por1_heREV	reverse	5' GCCGGATCCTCGAGCTAGCGCTACTATT CCTCTTCAAGCTTTCCACC 3'	53-59°C	*por*1 amplicon	present study
por2_heFWDs	forward	5' CTGGTGCCGCGCGGCAGCTCCATGCCGG TTCATGATCCTTAT 3'	53-59°C	*por*2 amplicon	present study
por2_heREV	reverse	5' GCCGGATCCTCGAGCTAGCGCTACTATT CAAACTTGCTGCTCAACTT 3'	53-59°C	*por*2 amplicon	present study

CONCLUSIONS

The evolution of LIPOR and POR genes in eukaryotic algae via endosymbiotic gene transfer, horizontal gene transfer, gene duplication, and gene loss is a powerful reminder that genome construction is on-going, multi-layered, and complex. Particularly, the trend in angiosperms, haptophytes, and many stramenopiles [1–3] towards maintenance of multiple POR enzymes in the absence of LIPOR is striking. Potential accumulation of chlorophyll precursor pigments in the dark in organisms lacking LIPOR may necessitate strict control of chlorophyll synthesis [4]. This obligation is perhaps met in part by subdivision of chlorophyll synthesis needs among various POR enzymes each optimized to a particular regulatory or catalytic role.

Data presented herein evidence alternative roles of dual POR proteins in a stramenopile alga, the diatom *P. tricornutum*. One of its POR enzymes, POR1, accumulates in a manner consistent with photoacclimation to lower irradiances with increased chlorophyll levels. Abundance of the other POR enzyme, POR2, increases in the light period and decreases in the dark period and this diurnal rhythmicity is maintained despite changes in photoperiod and light intensity. Future work is necessary to determine whether the catalytic capacities of either of these *P. tricornutum* enzymes differ and whether they form hetero- or homocomplexes. Furthermore, the finding that a mRNA encoded C-terminal extension of POR2 is conserved across diatoms and yet cleaved from the mature protein opens various lines of investigation into the role of this peptide—in the regulation of chlorophyll synthesis, POR protein transit into the chloroplast, or another potential function.

Characterization of the roles of dual POR enzymes in chlorophyll synthesis in a diatom is a first step in elucidating the various ways that expanded POR families may be utilized across the algae. Diverse *por* gene duplication events among the land plants have led to diverse regulatory schema [2, 5]. Given the various gene duplication events giving rise to the stramenopile/haptophyte, dinoflagellate and chlorarachniophyte *por* gene sets, it will be interesting to explore regulatory and functional differences and similarities among their POR isozymes. Even among taxa whose *por* paralogs can be traced to the same duplication event, such as the various classes of stramenopiles and haptophytes, the diversification of these algal lineages over hundred of millions of years [6–8] and their sheer diversity of physiologies and ecologies will likely lead to differentiation of their *por* regulatory programs.

Lastly, an unanticipated but exciting finding in these studies was that stramenopile and haptophyte algae share *por* gene duplicates of chlorophytic origin that have replaced their ancestral rhodophytic *por* genes [3]. The derived position of haptophyte POR proteins in both the POR1 and POR2 branches of stramenopile PORs suggests that haptophytes obtained these *por* genes from the stramenopiles, whether via horizontal gene transfer early in the establishment of the haptophyte lineage or by endosymbiotic gene transfer. The latter possibility is especially intriguing given that it is now believed that cryptophyte, stramenopile and haptophyte algae trace their chloroplasts to a single secondary endosymbiotic uptake of a rhodophytic alga, and that this chloroplast was then passed between lineages in higher order endosymbioses in an as-yet-unknown order [9, 10]. These data would then support the emergent hypothesis that the order of endosymbiosis is: cryptophyte → stramenopile → haptophyte [11]. Resolution of haptophyte clades of POR proteins within rather than sister to stramenopile POR proteins was likely facilitated by extensive taxonomic sampling accomplished in this study, enabled by the recent

availability of newly sequenced algal genomes and transcriptomes [e.g., 12]. Further phylogenetic examination of nuclear-encoded, chloroplast-localized proteins shared by cryptophyte, stramenopile and haptophyte algae may permit a definitive understanding of the relationships of these ecologically and economically important algal taxa.

Works Cited

1. Fujita Y, Bauer CE: **The light-independent protochlorophyllide reductase: a nitrogenase-like enzyme catalyzing a key reaction for greening in the dark**. In Porphyr Handb. Volume 13. Edited by Kadish K, Smith K, Guilard R. Elsevier Science; 2003:109–156.

2. 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.

3. Hunsperger HM, Randhawa T, Cattolico RA: **Extensive horizontal gene transfer, duplication, and loss of chlorophyll synthesis genes in the algae**. BMC Evol Biol 2015, **15**.

4. Reinbothe S, Reinbothe C, Apel K, Lebedev N: **Evolution of chlorophyll biosynthesis—the challenge to survive photooxidation**. Cell 1996, **86**:703–705.

5. Masuda T, Fusada N, Shiraishi T, Kuroda H, Awai K, Shimada H, Ohta H, Takamiya K-I: **Identification of two differentially regulated isoforms of protochlorophyllide oxidoreductase (POR) from tobacco revealed a wide variety of light- and development-dependent regulations of POR gene expression among angiosperms**. Photosynth Res 2002, **74**:165–172.

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

7. 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**:161–176.

8. Parfrey LW, Lahr DJG, Knoll AH, Katz LA: **Estimating the timing of early eukaryotic diversification with multigene molecular clocks.** Proc Natl Acad Sci U S A 2011, **108**:13624–9.

9. Bodył A, Stiller JW, Mackiewicz P: **Chromalveolate plastids: direct descent or multiple endosymbioses?** Trends Ecol Evol 2009, **24**:119–21; author reply 121–2.

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

11. 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.

12. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler L a, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, Beszteri B, Bidle KD, Cameron CT, Campbell L, Caron D a, Cattolico RA, Collier JL, Coyne K, Davy SK, Deschamps P, Dyhrman ST, Edvardsen B, Gates RD, Gobler CJ, Greenwood SJ, Guida SM, Jacobi JL, Jakobsen KS, James ER, Jenkins B, 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.

APPENDIX A

The mitochondrial and chloroplast genomes of the haptophyte *Chrysochromulina tobin* contain unique repeat structures and gene profiles

Blake T Hovde[1], Shawn R Starkenburg[2], Heather M Hunsperger[3], Laina D Mercer[4], Chloe R Deodato[3], Ramesh K Jha[2], Olga Chertkov[2], Raymond J Monnat[1,5] and Rose Ann Cattolico[3§]

[§]corresponding author

[1]Department of Genome Sciences, University of Washington, Seattle, WA, USA
[2]Los Alamos National Laboratory, Bioscience Division, Los Alamos, NM, USA
[3]Department of Biology, University of Washington, Seattle, WA, USA
[4]Department of Statistics, University of Washington, Seattle, WA, USA
[5]Department of Pathology, University of Washington, Seattle, WA, USA

Abstract

Background: Haptophytes are widely and abundantly distributed in both marine and freshwater ecosystems. Few genomic analyses of representatives within this taxon have been reported, despite their early evolutionary origins and their prominent role in global carbon fixation.

Results: The complete mitochondrial and chloroplast genome sequences of the haptophyte *Chrysochromulina tobin* (Prymnesiales) provide insight into the architecture and gene content of haptophyte organellar genomes. The mitochondrial genome (~34 kb) encodes 21 protein coding genes and contains a complex, 9 kb tandem repeat region. Similar to other haptophytes and rhodophytes, but not cryptophytes or stramenopiles, the mitochondrial genome has lost the *nad*7, *nad*9 and *nad*11 genes. The ~105 kb chloroplast genome encodes 112 protein

coding genes, including *ycf*39 which has strong structural homology to NADP-binding nitrate transcriptional regulators; a divergent 'CheY-like' two-component response regulator (*ycf*55) and Tic/Toc (*ycf*60 and *ycf*80) membrane transporters. Notably, a zinc finger domain has been identified in the *rpl*36 ribosomal protein gene of all chloroplasts sequenced to date with the exception of haptophytes and cryptophytes - algae that have gained (via lateral gene transfer) an alternative *rpl*36 lacking the zinc finger motif. The two *C. tobin*chloroplast ribosomal RNA operon spacer regions differ in tRNA content. Additionally, each ribosomal operon contains multiple single nucleotide polymorphisms (SNPs) - a pattern observed in rhodophytes and cryptophytes, but few stramenopiles. Analysis of small (<200 bp) chloroplast encoded tandem and inverted repeats in *C. tobin* and 78 other algal chloroplast genomes show that repeat type, size and location are correlated with gene identity and taxonomic clade.

Conclusion: The *Chrysochromulina tobin* organellar genomes provide new insight into organellar function and evolution. These are the first organellar genomes to be determined for the prymnesiales, a taxon that is present in both oceanic and freshwater systems and represents major primary photosynthetic producers and contributors to global ecosystem stability.

Status: Published in BMC Genomics 2014, **15**:604.

APPENDIX B

Genome sequence and transcriptome analyses of *Chrysochromulina tobin*: metabolic tools for enhanced algal fitness in the prominent order Prymnesiales (Haptophyceae)

Blake T. Hovde[1§*], Chloe R. Deodato[2], Heather M. Hunsperger[2], Scott A. Ryken[2], Will Yost[2], Ramesh K. Jha[3], Johnathan Patterson[2], Raymond J. Monnat Jr.[1,4], Steven B. Barlow[5], Shawn R. Starkenburg[3], and Rose Ann Cattolico[2§*]

[§]corresponding authors, *co-first authors

[1]University of Washington, Department of Genome Sciences, Seattle, WA, USA
[2]University of Washington, Department of Biology, Seattle, WA, USA
[3]Los Alamos National Laboratory, Los Alamos, NM, USA
[4]University of Washington, Department of Pathology, Seattle, WA, USA
[5]San Diego State University, Electron Microscope Facility, San Diego, CA, USA

Abstract

Haptophytes are recognized as seminal players in aquatic ecosystem function. These algae are important in global carbon sequestration, form destructive harmful blooms, and given their rich fatty acid content, serve as a highly nutritive food source to a broad range of eco-cohorts. Haptophyte dominance in both fresh and marine waters is supported by the mixotrophic nature of many taxa. Despite their importance, the nuclear genome sequence of only one haptophyte, *Emiliania huxleyi* (Isochrysidales), is available. Here we report the draft genome sequence of *Chrysochromulina tobin* (Prymnesiales), and transcriptome data collected at seven time points over a 24-hour light/dark cycle. The nuclear genome of *C. tobin* is small (59 Mb), compact (~40% of the genome is protein coding) and encodes an estimated number of 16,777 genes.

Genes important to fatty acid synthesis, modification, and catabolism show distinct patterns of expression when monitored over the circadian photoperiod. The *C. tobin* genome harbors the first hybrid polyketide synthase/non-ribosomal peptide synthase gene complex reported for an algal species, and encodes potential anti-microbial peptides and proteins involved in multidrug and toxic compound extrusion. A new haptophyte xanthorhodopsin was also identified, together with two "red" RuBisCO activases that are shared across many algal lineages. Data suggests that several identified genes may be products of lateral gene transfer. The *Chrysochromulina tobin* genome sequence provides new information on the evolutionary history, ecology and economic importance of haptophytes.

Status: In review.

APPENDIX C

Cell morphology, mixotrophy and genetics of *Chrysochromulina tobin* sp. nov. (Haptophyta) - a new model system for analyzing oleaginous algae

Chloe R. Deodato[1]*, Steven Barlow[2]*, Robert A. Andersen[3], Blake Hovde[4], Johnathan Patterson[1], William Yost[1], Heather M. Hunsperger[1], Shawn Starkenburg[5], Rose Ann Cattolico[1§]

[§]corresponding author, *co-first authors

[1]Department of Biology, University of Washington, Seattle, WA, USA
[2]Department of Biology, San Diego State University, San Diego, CA, USA
[3]Friday Harbor Laboratories, University of Washington, Friday Harbor, WA, USA
[4]Department of Genome Sciences, University of Washington, Seattle, WA, USA
[5]Los Alamos National Laboratories, Bioscience Division, Los Alamos, NM

Abstract

Background: Until recent studies documented their extensive contribution to primary productivity and carbon sequestration, haptophytes remained underappreciated players in global ecosystem processes. The ability of many haptophyte species to grow mixotrophically contributes to their dominance in both marine and freshwater environments, while their high fatty acid complement provides a calorically rich nutritional source for eco-cohorts. Sexual reproduction has been observed for many species.

Chrysochromulina parva Lackey, the type species for the prymnesiophyte clade of the haptophyte taxon, is a small saddle-shaped cell embellished with scales. We propose that a scale-less *Chrysochromulina* culture (CCMP291$_{RAC}$) was misidentified and should be taxonomically reassigned. Given the morphological, genetic and physiological characteristics of the organism, it

will serve as an excellent model organism for a systems biology assessment of haptophyte biology.

Results: Scanning and transmission electron microscopy show the naked, ~6µm CCMP291$_{RAC}$ cell is saddle-shaped to globose. The complex rootlet system anchors two flagella that are sub-apically inserted near the long filiform haptonema. Two lipid bodies, associated with numerous mitochondria, are each nestled close to a chloroplast. Plastoglobuli in the chloroplast are frequently associated with the thylakoidal membranes that encircle an internally located pyrenoid. The Golgi apparatus is robust with large, club-shaped cisternae. Mixotrophic cultures with their full bacterial biome (~10 spp.) grow faster and produce ~30% more lipids than those associated with a single bacterial species.

Sequence and phylogenetic analyses of chloroplast (*psb*A, *rbc*L), mitochondrial (*cox*1, *nad*5) and nuclear genes (rDNA) show CCMP291$_{RAC}$ to genetically differ from the type species and to be nested within the prymnesiophyte B2 clade. Whole genome sequencing reveals the organism to be haploid, though genes associated with meiosis and DNA repair are present.

Conclusions: *Chrysochromulina tobin* sp. nov. should be recognized as a new species. *C. tobin* is proposed as a model system for lipid biogenesis studies in haptophytes due to the wall-less nature of the organism, the presence of two lipid bodies, its copious fatty acid production, the regulation of lipid synthesis by physiological cues and the augmentation of lipid synthesis by the presence of bacterial cohorts. The potential sexual cycle of this organism adds to its candidacy as an experimental system.

Status: In preparation.

APPENDIX D

Evolution and maintenance of haploid-diploid life cycles in natural populations: the case of the marine brown alga *Ectocarpus*

Lucía Couceiro[1§], Mickael Le Gac[1,4], Heather M. Hunsperger[1,5], Stéphane Mauger[1], Christophe Destombe[1], J. Mark Cock[2], Sophia Ahmed[2,6], Susana M. Coelho[2], Myriam Valero[1], and Akira F. Peters[3]

[§]corresponding author

[1]CNRS, Sorbonne Universités, UPMC, UMI 3614, Evolutionary Biology and Ecology of Algae, Station Biologique de Roscoff, Roscoff, France
[2]CNRS, Sorbonne Universités, UPMC, UMR 7139, Algal Genetics, Station Biologique de Roscoff, Roscoff, France
[3]Bezhin Rosko, Santec, France
[4]Present address: Ifremer, DYNECO PELAGOS, Plouzané, France
[5]Present address: Department of Biology, University of Washington, Seattle, Washington
[6]Present address: Department for Plant Sciences, University of Leeds, Leeds, UK

Abstract

The evolutionary stability of haploid-diploid life cycles is still controversial. Mathematical models indicate that niche differences between ploidy phases may be a necessary condition for the evolution and maintenance of these life cycles. Nevertheless, experimental support for this prediction remains elusive. In the present work, we explored this hypothesis in natural populations of the brown alga *Ectocarpus*. Consistent with the life cycle described in culture, *E. crouaniorum* in NW France and *E. siliculosus* in SW Italy exhibited an alternation between haploid gametophytes and diploid sporophytes. Our field data invalidated, however, the long-standing view of an isomorphic alternation of generations. Gametophytes and sporophytes

displayed marked differences in size and, conforming to theoretical predictions, occupied different spatio-temporal niches. Gametophytes were found almost exclusively on the alga *Scytosiphon lomentaria* during spring while sporophytes were present year-round on abiotic substrata. Paradoxically, *E. siliculosus* in NW France exhibited similar habitat usage despite the absence of alternation of ploidy phases. Diploid sporophytes grew both epilithically and epiphytically, and this mainly-asexual population gained the same ecological advantage postulated for haploid-diploid populations. Consequently, an ecological interpretation of the niche differences between haploid and diploid individuals does not seem to satisfactorily explain the evolution of the *Ectocarpus* life cycle.

Status: Accepted in Evolution.