

RESEARCH ARTICLE

# Evolution of the Tetrapyrrole Biosynthetic Pathway in Secondary Algae: Conservation, Redundancy and Replacement

Jaromír Cihlár<sup>1,2</sup>\*, Zoltán Füßy<sup>1</sup>, Aleš Horák<sup>1,2</sup>, Miroslav Oborník<sup>1,2,3\*</sup>

**1** Biology Centre, Czech Academy of Sciences, Institute of Parasitology, České Budějovice, Czech Republic, **2** University of South Bohemia, Faculty of Science, České Budějovice, Czech Republic, **3** Institute of Microbiology, Czech Academy of Sciences, Třeboň, Czech Republic

\* These authors contributed equally to this work.

\* [obornik@paru.cas.cz](mailto:obornik@paru.cas.cz)



**OPEN ACCESS**

**Citation:** Cihlár J, Füßy Z, Horák A, Oborník M (2016) Evolution of the Tetrapyrrole Biosynthetic Pathway in Secondary Algae: Conservation, Redundancy and Replacement. PLoS ONE 11(11): e0166338. doi:10.1371/journal.pone.0166338

**Editor:** Claude Prigent, Institut de Genetique et Developpement de Rennes, FRANCE

**Received:** September 1, 2016

**Accepted:** October 26, 2016

**Published:** November 18, 2016

**Copyright:** © 2016 Cihlár et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files. Novel sequences of interest were deposited in GenBank under the accession no. KX344033-47.

**Funding:** The Czech Science Foundation (GAP506/12/1522) and the Czech Academy of Sciences provided funding to MO and ZF. Computation resources were provided by CERIT-SC and MetaCentrum, Brno, Czech Republic. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Tetrapyrroles such as chlorophyll and heme are indispensable for life because they are involved in energy fixation and consumption, i.e. photosynthesis and oxidative phosphorylation. In eukaryotes, the tetrapyrrole biosynthetic pathway is shaped by past endosymbioses. We investigated the origins and predicted locations of the enzymes of the heme pathway in the chlorarachniophyte *Bigeloviella natans*, the cryptophyte *Guillardia theta*, the “green” dinoflagellate *Lepidodinium chlorophorum*, and three dinoflagellates with diatom endosymbionts (“dinotoms”): *Durinskia baltica*, *Glenodinium foliaceum* and *Kryptoperidinium foliaceum*. *Bigeloviella natans* appears to contain two separate heme pathways analogous to those found in *Euglena gracilis*; one is predicted to be mitochondrial-cytosolic, while the second is predicted to be plastid-located. In the remaining algae, only plastid-type tetrapyrrole synthesis is present, with a single remnant of the mitochondrial-cytosolic pathway, a ferrochelatase of *G. theta* putatively located in the mitochondrion. The green dinoflagellate contains a single pathway composed of mostly rhodophyte-origin enzymes, and the dinotoms hold two heme pathways of apparently plastidal origin. We suggest that heme pathway enzymes in *B. natans* and *L. chlorophorum* share a predominantly rhodophytic origin. This implies the ancient presence of a rhodophyte-derived plastid in the chlorarachniophyte alga, analogous to the green dinoflagellate, or an exceptionally massive horizontal gene transfer.

## Introduction

Plastid acquisitions are rare evolutionary events that give host cells the metabolic capacities of their new photosynthetic organelles. While there are only two documented primary plastid acquisitions [1,2], involving a eukaryote as host and cyanobacterium as the endosymbiont, the history of higher order eukaryote-to-eukaryote acquisitions is intensely debated [3–8]. Based on available data, it is believed that the red plastids of cryptophytes, alveolates, stramenopiles, and haptophytes (the “CASH taxa”) originate from a single ancient event with a rhodophyte alga as the endosymbiont [9–14]. However, phylogenies of the host organisms are often

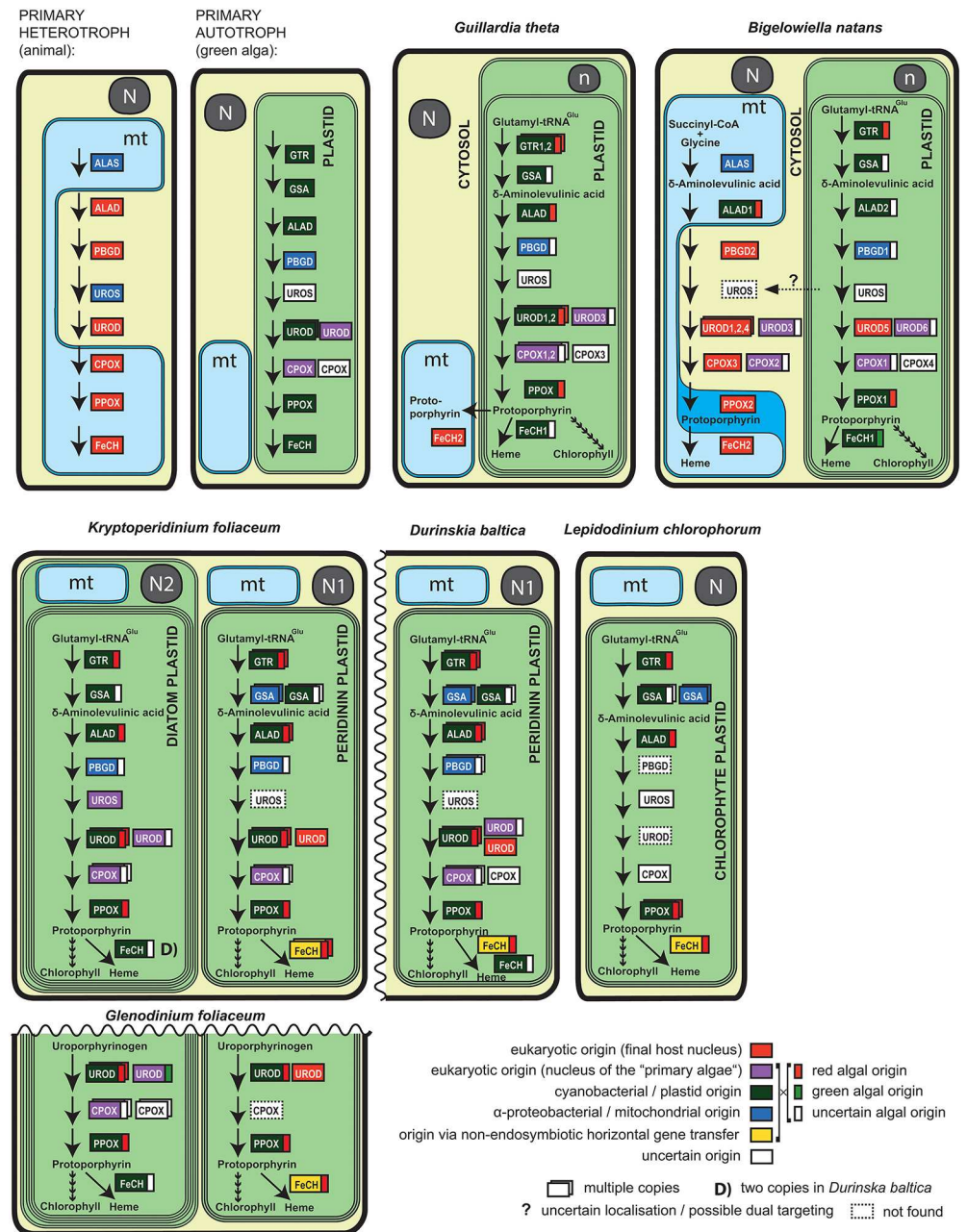
**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** ALA,  $\delta$ -aminolevulinic acid; ALAD, ALA dehydratase; ALAS, ALA synthase; CASH, group of complex algae with red-derived plastid of putatively common origin, comprising cryptophytes, alveolates, stramenopiles and haptophytes; CPOX, coproporphyrinogen oxidase; FeCH, ferrochelatase; GSA-AT, glutamate-1-semialdehyde 2,1-aminotransferase; GTR, glutamate-tRNA reductase; HGT, horizontal gene transfer; PBGD, porphobilinogen deaminase; PPOX, protoporphyrinogen oxidase; UROD, uroporphyrinogen decarboxylase; UROS, uroporphyrinogen-III synthase.

incompatible with this scenario [6,15–18], suggesting that the plastid was transferred horizontally in at least some of these lineages. Furthermore, higher order endosymbioses and horizontal gene transfer (HGT) may be blurring our vision of eukaryotic evolution [7]. For instance, some dinoflagellates replaced their ancestral peridinin-pigmented plastids with plastids originating from serial secondary or tertiary endosymbioses [7,19–23]. The original red plastid was replaced by a secondary green plastid in *Lepidodinium chlorophorum* [24,25]; while in so-called dinotoms (*Glendinium foliaceum*, *Durinskia baltica*), the newly obtained tertiary endosymbiont is an engulfed diatom [26,27]. Serial plastid endosymbioses are sometimes discernible by phylogenetic signal (e.g. if a green plastid replaces a red one). However, ancient events can still be difficult to pinpoint, which might account for the contradictory and peculiar phylogenetic signals observed—for example, the number of green genes in the CASH taxa [28–33] and the proposed independent origin of plastid genes in two main branches of alveolates, the dinoflagellates and apicomplexans [34,35]. In contrast, chlorarachniophytes and phototrophic euglenids acquired green plastids and their extant relatives are heterotrophic, allowing for straightforward evolutionary interpretations of gene origins, based on phylogenetic clustering with their heterotrophic kin or with the chlorophyte plastid donors [31,36,37].

The process of endosymbiosis involves endosymbiont genome reduction via gene transfer to the host nucleus [38,39], allowing for enhanced host control over the organelle [40,41] and reduced functional redundancy of cellular biochemistry [42]. However, the level of reduction differs among algae possessing complex photosynthetic organelles. Most of them, such as diatoms, dinoflagellates or phototrophic euglenids, have a highly reduced algal endosymbiont with multiple membranes surrounding the plastid as the only apparent morphological footprints revealing past complex endosymbioses. Organelle reduction tends to be higher in cases of older symbioses, but also depends on other factors including plastid number and evolutionary constraints [43,44]. For instance, cryptophyte and chlorarachniophyte plastids seem to be evolutionarily frozen and retain a remnant nucleus (nucleomorph) that provides genetic material required, e.g. for the maintenance of protein import mechanisms [31,45,46]. In dinotoms, the diatom endosymbiont still contains a plastid, a mitochondrion and a nucleus and is thought to represent an almost entirely independent cellular compartment [47,48]. Furthermore, it appears that the host dinotom cell holds a metabolically active remnant of the original peridinin-pigmented plastid, presumably the eyespot [47,49].

One of the essential biochemical pathways carried out in plastids is tetrapyrrole synthesis. Tetrapyrroles are cyclic porphyrins coordinated by iron (heme) or magnesium (chlorophyll). They are essential for life, since heme is a substantial component of the respiratory chain and chlorophyll is an indispensable compound in the conversion of light energy to chemical bonds in carbohydrates through photosynthesis. Although the kinetoplastid flagellate *Phytomonas serpens* has been shown to be able to live in the absence of heme, it is an extremely rare metabolic deviation [50]. In phototrophic eukaryotes, tetrapyrroles are required in three cellular compartments: the cytosol, the mitochondrion, and the plastid. Most phototrophs synthesize tetrapyrrole compounds exclusively in the plastid and transport them to other compartments (overview in Fig 1). The biosynthetic process, however, can be more complex; in the excavate alga *Euglena gracilis* two independent tetrapyrrole biosynthesis pathways are present [51–55], likely because of the recent acquisition of the secondary green plastid [36,37]. These parallel tetrapyrrole pathways differ in both evolutionary origin and starting substrate. One pathway derives from the heterotrophic (secondary) host and uses condensation of succinyl-CoA and glycine (via the C4 pathway) to synthesize  $\delta$ -aminolevulinic acid (ALA), the first common precursor, in the mitochondrion. The heterotrophic-type synthesis takes place partly in the mitochondrion and partly in the cytosol as it does in eukaryotic primary heterotrophs [55,56]. The other pathway is located entirely in the plastid and generates ALA from glutamate (via the C5



**Fig 1. Arrangement of the heme biosynthetic pathway in algae with complex plastids.** Inferred origins of enzymes are represented by colored boxes with flags where applicable. Localizations of *B. natans* and *G. theta* enzymes were predicted by SignalP and TargetP (see [Material and Methods](#)). Dashed arrows indicate a possible dual localization of UROS in both the cytosol and the plastid of *B. natans*. Only key metabolites are shown for clarity, for example the starting substrates for mitochondrial-cytosolic C4 (succinyl-CoA and glycine) and plastid C5 pathways (glutamyl-tRNA<sup>Glu</sup>). Parts of the pathway identical to *K. foliaceum* are not depicted in the *D. baltica* and *G. foliaceum* scheme. Schematic representation of organelles: N, N1—nucleus of the host; N2—nucleus of the endosymbiont diatom; n—nucleomorph of the endosymbiont; mt—mitochondrion. Enzymes: ALAS—delta-aminolevulinic acid synthase; GTR—glutamate-tRNA reductase; GSA—glutamate-1-semialdehyde 2,1-aminotransferase; ALAD—aminolevulinic acid dehydratase; PBGD—porphobilinogen deaminase; UROS—uroporphyrinogen-III synthase; UROD—uroporphyrinogen decarboxylase; CPOX—coproporphyrinogen oxidase; PPOX—protoporphyrinogen oxidase; FeCH—ferrochelatase. A typical pathway in a primary heterotroph and a primary autotroph are shown for comparison (Košný and Oborník 2011).

doi:10.1371/journal.pone.0166338.g001

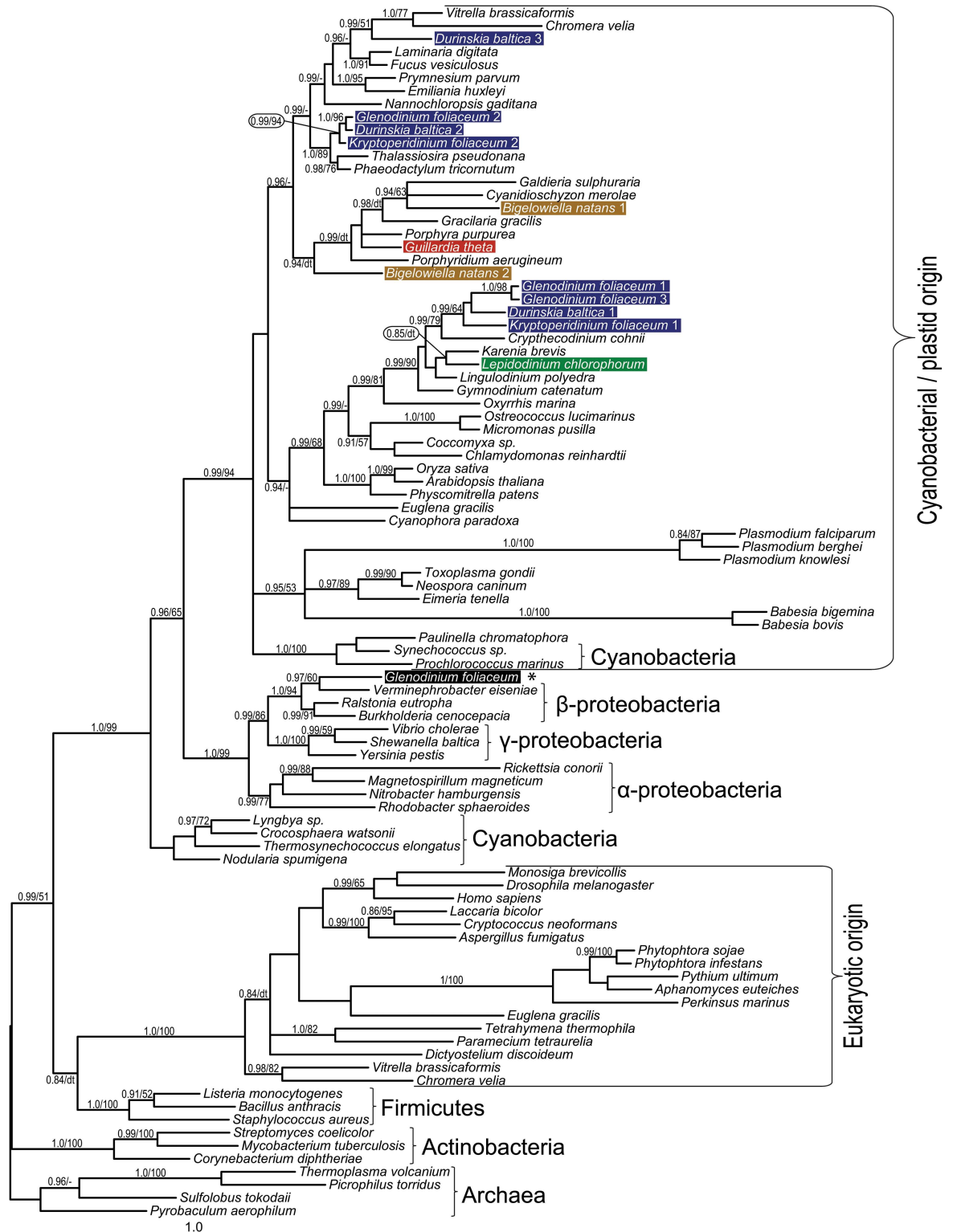
pathway). Evolutionarily, the plastid pathway originates from the green algal endosymbiont and, ultimately, from the cyanobacterium engulfed during primary endosymbiosis. In *E. gracilis*, the two tetrapyrrole pathways do not overlap, and thus produce tetrapyrroles separately for the cytosol and mitochondrion and for the plastid [51,54]. The presence of two redundant tetrapyrrole pathways is interpreted as an intermediate state in endosymbiosis [55] and would allow for the loss of one of the pathways in order to streamline cellular biochemistry. Usually, the mitochondrial-cytosolic pathway is lost in the course of evolution in eukaryotic phototrophs [55,57] and it is rare to see the plastid pathway disappear in exchange for the cytosolic counterpart, but the parasite *Perkinsus marinus* [58], for example, retained the heterotrophic heme synthesis pathway despite the continued presence of a relict plastid (reviewed in [59]). The alveolate alga *Chromera velia*, on the other hand, employs a hybrid tetrapyrrole pathway. Synthesis is initiated in the mitochondrion via the C4 pathway and is predicted to continue in the plastid. This hybrid synthesis qualifies *Chromera* as the only known phototroph able to synthesize tetrapyrroles from glycine [57], similar to heme biosynthetic processes in apicomplexan parasites that still possess a remnant, non-photosynthetic plastid [60].

The organization of heme synthesis is currently uncharacterized in most algae with complex plastids. In order to map the level of pathway conservation, reduction or replacement in further phototrophic lineages, we investigated phylogenetic relationships and predicted the cellular locations of enzymes involved in tetrapyrrole biosynthesis in the cryptophyte *Guillardia theta*, the chlorarachniophyte *Bigelowiella natans*, the green dinoflagellate *Lepidodinium chlorophorum* and the dinotoms *Durinskia baltica*, *Glenodinium foliaceum* and *Kryptoperidinium foliaceum*.

## Results and Discussion

### *Bigelowiella natans* possesses two heme pathways

In *B. natans*, we identified thirteen and nine sequences, respectively, of enzymes belonging to the algal endosymbiont heme pathway (autotrophic pathway) and to the heterotrophic (mitochondrial-cytosolic) pathway. The chlorarachniophyte host pathway is typical for eukaryotic heterotrophs, with ALA being synthesized by the mitochondrial C4 pathway; the enzymes involved are predicted to localize to the mitochondrion (aminolevulinic acid synthase, or BnALAS; protoporphyrinogen oxidase, BnPPOX2; and ferrochelatase, BnFeCH2) and cytosol (porphobilinogen deaminase, BnPBGD2; uroporphyrinogen decarboxylase BnUROD1-4, and coproporphyrinogen oxidase, BnCPOX2, -3) (see Fig 1, Material and Methods and S1 Table for details). An N-terminal mitochondrial transit peptide was not predicted in BnALAS, but mitochondrial transit peptides have not been detected in any eukaryotic ALAS [57] examined so far, in spite of the fact that an N-terminal extension is apparent when the eukaryotic protein is aligned to bacterial homologs (not shown) and its amino acid composition resembles that of mitochondrial transit peptides. Moreover, ALAS has never been experimentally found outside of the mitochondrion, likely because it uses succinyl-CoA, a product of the mitochondrial citrate cycle, as the initial substrate [57,61]. The host gene for aminolevulinic acid dehydratase (ALAD) was lost and was likely replaced by a cyanobacterial (plastid) homolog (Fig 2); a predicted mitochondrial transit peptide at its N-terminus further supports a mitochondrial location (see S1 Table for details). BnPBGD2, responsible for the next step of the mitochondrial-cytosolic pathway, forms an unsupported but stable clade with other eukaryotic sequences, branching as sister to *E. gracilis* and oomycetes (non-photosynthetic stramenopiles). This clade, composed of animal, fungal, (phototrophic) excavate, heterotrophic stramenopile and chlorarachniophyte sequences, very likely represents the only PBGD enzymes originating in the eukaryotic nucleus because all phototrophic eukaryotes utilize PBGD of  $\alpha$ -proteobacterial origin (S1 Fig) [50,56,57].



**Fig 2. Bayesian phylogenetic tree as inferred from ALAD amino acid sequences.** Taxa of interest in this study are highlighted by colored bars: blue for dinotoms, green for *Lepidodinium chlorophorum*, ochre for *Bigelowiella natans* and red for *Guillardia theta*. The tree shows red algal origin for *B. natans* and *G. theta* enzymes. For *L. chlorophorum*, we suggest a gene duplication / loss of paralogs scenario (see text); despite branching as sister to green algae, other dinoflagellates contained in the same clade do not possess a green algal plastid. Numbers near branches indicate Bayesian posterior probabilities followed

by the bootstrap of respective clades from the likelihood analysis. Only support values greater than 0.85 (Bayesian) and 50 (likelihood) are shown. dt—different topology in the ML tree, see [S2 Fig](#); a dash indicates an unsupported topology. An asterisk marks inferred bacterial contamination in *G. foliaceum* data.

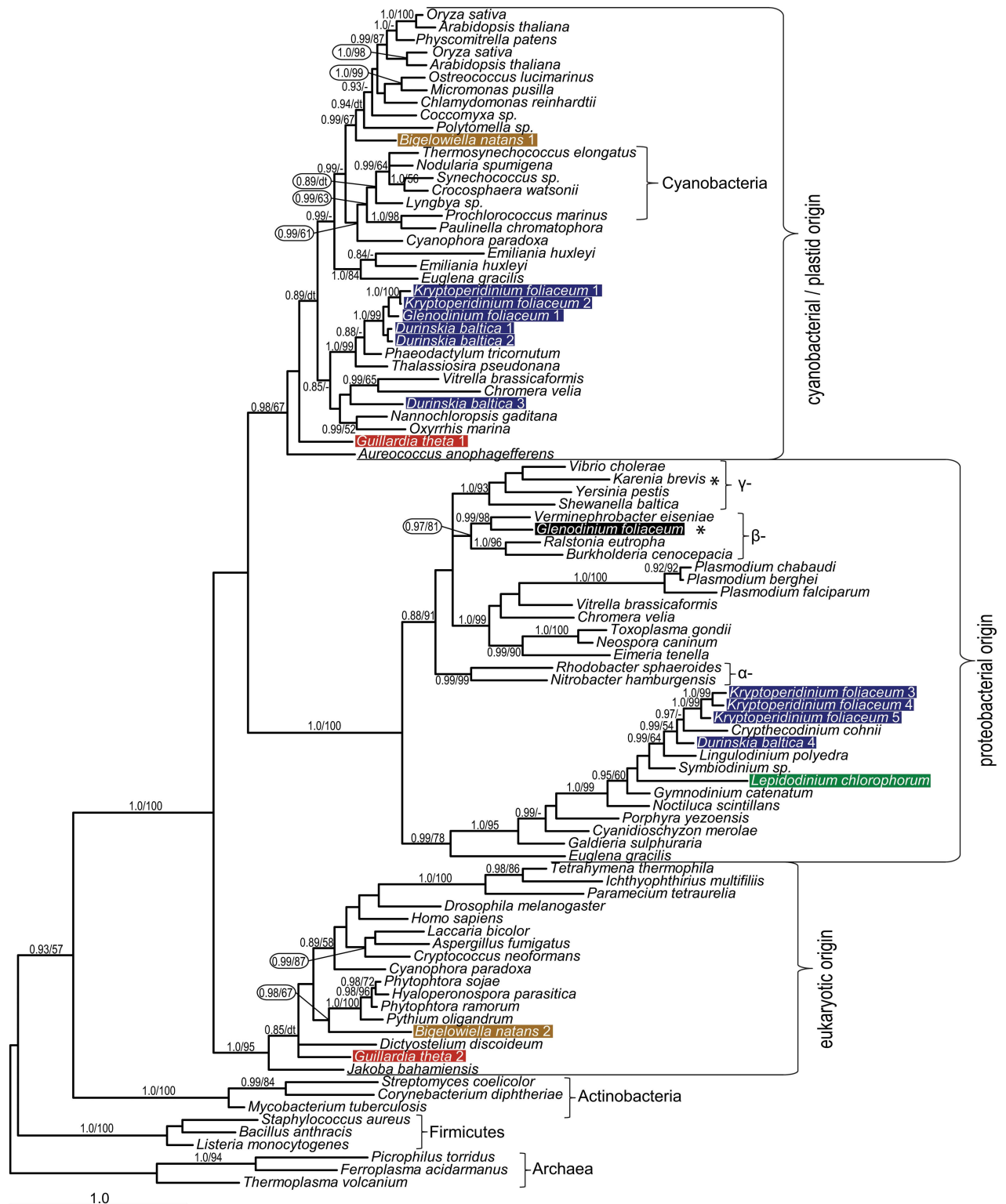
doi:10.1371/journal.pone.0166338.g002

In spite of the presence of two tetrapyrrole pathways in the chlorarachniophyte, only a single gene coding for a putatively plastid-targeted uroporphyrinogen-III synthase (UROS) was found in the genome database. The enzyme possesses a bipartite targeting sequence at the N-terminus, necessary for delivering the protein into the secondary plastid ([Fig 1](#), [S1 Table](#)). Notably, all UROS genes in eukaryotic phototrophs form a single compact cluster, although the origin of this phototrophic clade is not clear ([S1 Fig](#)). Since we failed to find the cytosolic (heterotrophic) UROS in the genome, we can speculate that it was either not covered by the sequencing or annotation methods used or the recovered *B. natans* protein is dual-targeted to the cytosol and the plastid.

Most of the enzymes involved in the mitochondrial-cytosolic pathway are encoded by a single gene in *B. natans*, with the exception of UROD. All six genes coding for UROD display eukaryotic origin in *B. natans*; however, four of the UROD sequences (BnUROD1, -2, -4 and -5) are obvious multiple paralogs likely originating from the secondary host nucleus, while two paralogs originate from the endosymbiont (algal) nucleus (BnUROD3 and -6) ([Fig 1](#) and [S1 Fig](#)). According to predictions, three UROD enzymes (BnUROD1, -2, -3) are cytosolic, while three other URODs appear to be plastid-located, and heterotrophic enzyme BnUROD5 may have functionally replaced the cyanobacterial counterparts in the autotrophic pathway ([S1 Table](#)).

There are two genes coding for CPOX (BnCPOX2, -3) predicted to function within the heterotrophic pathway in *B. natans*; the former appeared within the clade composed of red-derived secondary algae, and the latter has a nuclear origin with an unsupported sister position to ciliate sequences ([S1 Fig](#)). A eukaryotic origin is also suggested for the putatively mitochondrion-located BnPPOX2 in *B. natans*. The eukaryotic clade is supported by Bayesian analysis in this case but its internal structure is not resolved, forming numerous polytomies ([S1 Fig](#)), with *B. natans* PPOX2 appearing as the earliest eukaryotic branch. The mitochondrial ferrochelatase 2 is derived from the chlorarachniophyte host and is phylogenetically affiliated with heterotrophic stramenopiles (oomycetes), the only representatives of the SAR group in this particular clade ([Fig 3](#)).

The autotrophic tetrapyrrole pathway displays mosaic origins in *B. natans*, similar to that in other eukaryotic phototrophs. It is mostly composed of cyanobacterial-derived enzymes (glutamate-tRNA reductase BnGTR, glutamate-1-semialdehyde 2,1-aminotransferase BnGSA-AT, BnALAD1, -2, BnPPOX1 and BnFeCH1), but also of enzymes likely originating from the endosymbiont (primary host) nucleus (BnUROD6, BnCPOX1), one enzyme displaying an  $\alpha$ -proteobacterial (likely mitochondrial) origin (BnPBGD1), and an additional CPOX (BnCPOX4) of uncertain origin (see [Fig 1](#) for summary). Importantly, many of the aforementioned enzymes show unexpected phylogenetic affiliations: GTR, GSA-AT, ALAD1, CPOX1, -2 and PPOX1 in *B. natans* cluster with rhodophytes or algae with red secondary plastids (see [Figs 1, 2, 4](#) and [S1](#) for details), in spite of the chlorophyte origin of the current chlorarachniophyte plastid [[62](#)]. Only a single *B. natans* enzyme, one of the two ferrochelatases, displays the expected and supported chlorophyte origin (BnFeCH1, [Fig 3](#)). Although bootstrap values for the ML analyses are moderately supported, the hypothetical scenarios for red origin of many of these proteins are highly supported by Bayesian inference, which is more robust in analyses of data with high variability across sites. Furthermore, considering the good support for plastid genome relationship with green algae [[62](#)], one would expect the clear and well-supported association of *B. natans* sequences with either chlorophytes or heterotrophic eukaryotes; this is not observed.



**Fig 3. Bayesian phylogenetic tree as inferred from FeCH amino acid sequences.** Taxa of interest in this study are highlighted by colored bars: blue for dinotoms, green for *Lepidodinium chlorophorum*, ochre for *Bigelowiella natans* and red for *Guillardia theta*. We document two orthologs, one of unresolved cyanobacterial origin and the other of eukaryotic origin, for *B. natans* and *G. theta* enzymes. The *L. chlorophorum* sequence branches together with other dinoflagellates, suggesting its origin lies in the peridinin plastid repertoire. Numbers near branches indicate Bayesian posterior probabilities followed by the bootstrap of respective clades from the likelihood

analysis. Only support values greater than 0.85 (Bayesian) and 50 (likelihood) are shown. dt—different topology in the ML tree, see [S2 Fig](#); a dash indicates unsupported topology. An asterisk marks inferred bacterial contamination in *G. foliaceum* and *Karenia brevis* data.

doi:10.1371/journal.pone.0166338.g003

Several enzymes of the autotrophic pathway are present in multiple copies in *B. natans*, namely ALAD ([Fig 2](#)), UROD and CPOX ([S1 Fig](#)). The genome of *B. natans* contains two, likely paralagous, genes encoding ALAD, at least one of them affiliated with the red lineage ([Fig 2](#)). Three genes coding for autotrophic CPOX have been found in the *B. natans* genome, and all three of them appear to be plastid-targeted ([S1 Table](#)): one gene (BnCPOX4) is recovered with sequences from other phototrophs with no support; the origin of this clade is unclear. BnCPOX1 appeared within the clade composed of red-derived secondary algae and appeared to be related neither to chlorophytes nor to rhodophytes ([S2 Fig](#)). We propose that these genes might originate from the endosymbiont (primary host) nucleus ([Figs 1 and S1](#)).

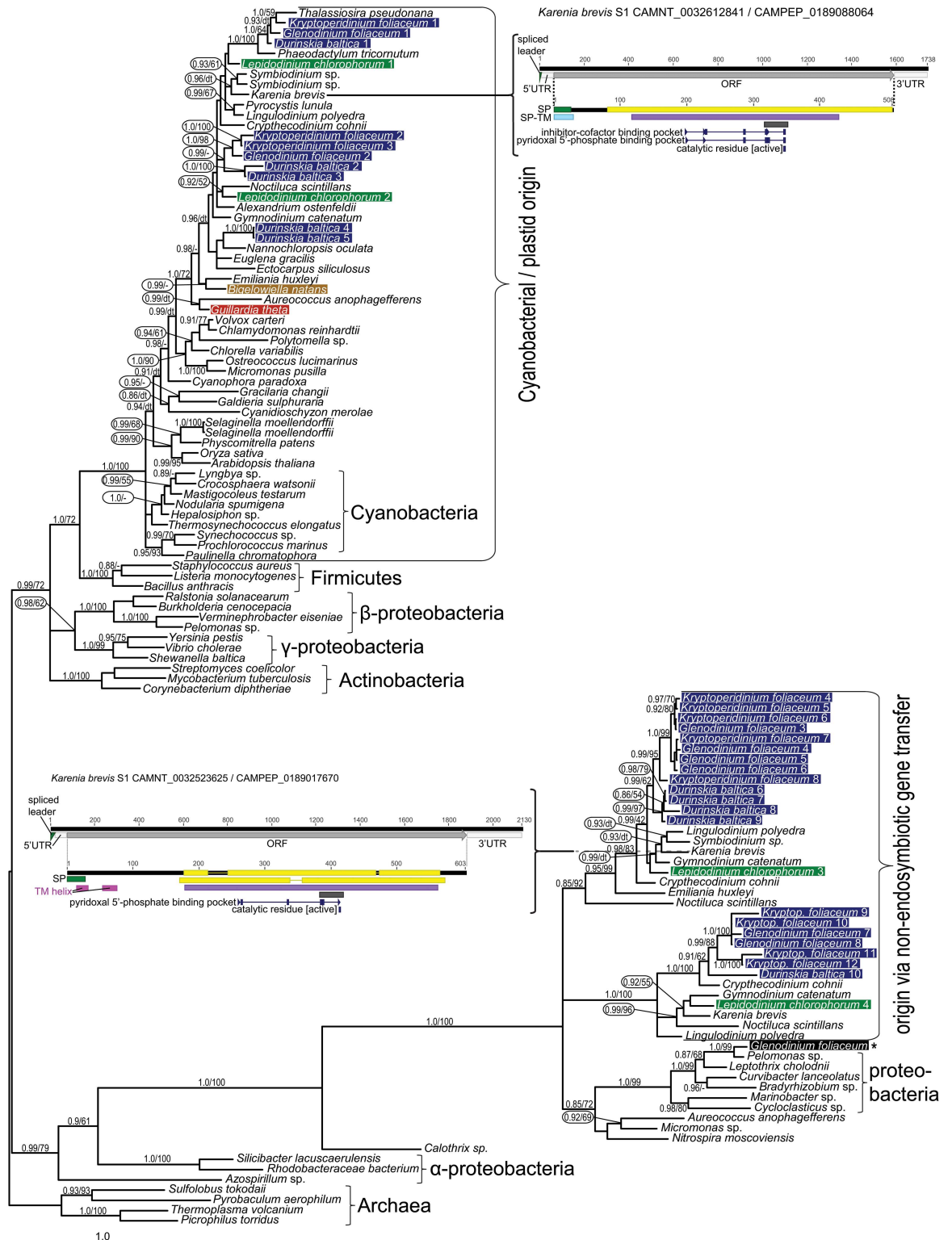
The putative cellular locations corresponding to all enzymes involved in tetrapyrrole biosynthesis in *B. natans* are shown in [S1 Table](#), as inferred using SignalP [63] and TargetP [64]. The entire autotrophic tetrapyrrole pathway is likely located in the plastid stroma of the chlorarachniophytes; no enzyme seems to be targeted to the periplastidal space with the possible exception of BnPBGD2, which fulfills some of the criteria described by Curtis et al. [31], namely high number of introns, D/K amino acids at the C terminus, and transit peptide net charge -1. The biological function of an isolated enzyme in this compartment, however, would be unclear and this targeting is unlikely. The plastid-origin BnALAD1 is putatively retargeted to the chlorarachniophyte mitochondrion where it is involved in the mitochondrial-cytosolic pathway. Conversely, paralogs of heterotrophic UROD4 and -5 appear to be retargeted to the plastid compartment ([Fig 1 and S1 Table](#)).

*B. natans* shares the analogous biparallel architecture of tetrapyrrole biosynthesis with photosynthetic euglenids. Plastid acquisition occurred relatively recently in euglenids, since phototrophic euglenids constitute a monophyletic group [36] and the previous presence of plastids was never confirmed in phagotrophic euglenids or their osmotrophic relatives [37]. In spite of the independent origins of chlorarachniophytes (Rhizaria) and euglenophytes (Excavata) [62], they display a similar pattern of tetrapyrrole synthesis [55]. Both algae possess two nearly complete pathways, one originating from the primary heterotrophic host, the other from the engulfed algal endosymbiont ([Fig 1](#)). In both algae the reduction of the redundant pathway has already begun and the mitochondrial-cytosolic pathway is partially reduced. In *Euglena*, two enzymes of plastid origin functionally replaced the original heterotrophic pathway genes for UROS and UROD, either via dual targeting or by sharing the products of the reactions they catalyze between compartments [55]. Similarly, in *Bigeloviella*, the original ALAD and UROS from the heterotrophic pathway were likely replaced via dual localization of the plastid-derived enzymes or through the exchange of pathway intermediates ([Fig 1](#)). Furthermore, one of the PBGD enzymes (possibly BnPBGD2) must be dually targeted for the heterotrophic pathway to function. As in other phototrophs, both *B. natans* and *E. gracilis* contain multiple copies (orthologs) of UROD and CPOX. In summary, the level of metabolic reduction is comparable in *Bigeloviella* and *Euglena*, which suggests that they acquired their green plastids at approximately the same time (assuming similar rates of evolution) or a constraint imposed on cellular metabolism that prevents a loss of redundancy in chlorarachniophytes.

### The ultimate step of the cryptophyte pathway is bifurcated

With the exception of the ferredoxin predicted to localize in the mitochondrion (see below), only the set of enzymes originating from the algal endosymbiont and putatively targeted to the





**Fig 4. Bayesian phylogenetic tree as inferred from GSA-AT amino acid sequences.** Taxa of interest in this study are highlighted by colored bars, blue for dinotoms, green for *Lepidodinium chlorophorum*, ochre for *Bigelowiella natans* and red for *Guillardia theta*. Numbers near branches indicate Bayesian posterior probabilities followed by the bootstrap of respective clades from the maximum likelihood (ML) analysis. Only support values greater than 0.85 (Bayesian) and 50 (ML) are shown. dt—different topology in the ML tree, see S2 Fig; a dash indicates unsupported topology. The tree demonstrates the

cyanobacterial origin of canonical GSA-AT, while the non-canonical GSA-AT originates in proteobacteria. Schematics of *Karenia brevis* transcripts and respective proteins are shown for complete representatives of canonical and non-canonical GSA-AT. The presence of a spliced-leader sequence at the 5' end suggests nuclear encoding and transcription of these genes. An N-terminal presequence of the resulting protein putatively targets both enzymes into the plastid. The canonical and non-canonical enzymes share motifs for pyridoxal 5'-phosphate binding and a catalytical residue. UTR—untranslated region; ORF—open reading frame; TM—transmembrane domain; SP—signal peptide; SP-TM—signal peptide predicted by the SignalP-TM networks; yellow bar—Panther Class III aminotransferase / glutamate-1-semialdehyde 2,1-aminomutase hit; violet bar—Pfam Class III aminotransferase hit; grey bar—PROSITE Class III aminotransferase hit; numbers represent scale in nt or aa.

doi:10.1371/journal.pone.0166338.g004

plastid compartment was found in *Guillardia theta* (Fig 1). We found three copies of UROD (one duplicated cyanobacterial gene and one gene originating from the endosymbiont nucleus) and CPOX (one duplicated gene from the endosymbiont nucleus and one gene of uncertain origin, GtCPOX2; see Figs 1 and S1 and S1 Table); both multiplied sets of enzymes are consequently combinations of orthologs originating from the primary endosymbiotic event (see S1 Fig for details), with one of them duplicated a second time. Gene duplication and functional specialization are also seen in other algae and plants [65]. While the duplication of the cyanobacterial UROD seems to be deeply branching, paralogs of CPOX arose more recently. Most of the enzymes involved originated from a cyanobacterial plastid ancestor related to rhodophytes and the CASH lineage (GtGTR1 and -2, GtGSA-AT, GtALAD, GtUROD1 and -2, GtPPOX; see Figs 2–4 and S1), supplemented by enzymes originating from the endosymbiont nucleus (UROD3, CPOX1 and -3),  $\alpha$ -proteobacteria (mitochondria; GtPBGD) and enzymes encoded by algae-affiliated genes of unknown origins (GtUROS, GtCPOX2) (see overview in Fig 1). This is in line with the presence of the rhodophyte-derived secondary plastid in cryptophytes [66,67]. Clustering of GtUROD3 with homologs from green algae, *E. gracilis* and sequences from the CASH group (S1 Fig) may be the result of gene duplication followed by lineage-specific gene loss (discussed below).

The mitochondrion-located ferrochelatase GtFeCH2 is eukaryotic in origin, as are its homologs in heterotrophic eukaryotes and the glaucophyte *Cyanophora paradoxa* (Figs 1 and 4), and putatively targeted to the mitochondrion (S1 Table). The retention of mitochondrial ferrochelatase may be the result of slower rates of evolution in the cryptophyte plastid when compared to other red-derived secondary plastids [55–57,68], an evolutionary constraint placed on its role in the biology of the organism, or an independent and more recent plastid acquisition in cryptophytes. The latter view is consistent with the presence of a nucleomorph in cryptophytes and with the growing body of evidence showing that phototrophic cryptophytes emerged through an independent endosymbiosis event [18,69,70].

## Novel type of GSA-AT and a proteobacterial FeCH in dinotoms

We analyzed transcriptomes from the dinotoms (dinoflagellates with tertiary diatom endosymbionts) *Glenodinium foliaceum* CCAP 1116/3, *Kryptoperidinium foliaceum* CCMP 1326 and *Durinskia baltica* (available via the MMETSP initiative, see S3 Table). Two redundant heme pathways are present in dinotoms: one is located in the diatom endosymbiont, while the second remains in the putative remnant of the original peridinin plastid, the eye spot [47]. These separate pathways seem to supply tetrapyrroles to the two independent symbiotic partners. Our inferred trees display similar topologies and evolutionary relationships as those published previously (see Fig 1 for summary; [47]); the endosymbiont pathway enzymes are related to sequences from free-living diatoms, while the host enzymes cluster together with other dinoflagellates. Furthermore, we identified a new biochemically

uncharacterized subfamily of putative GSA-AT, which is found in other dinoflagellates as well as in *Aureococcus anophagefferens*, *Emiliania huxleyi* and the chlorophyte *Micromonas*. Since the non-canonical putative GSA-AT clade contains a smaller clade composed of proteobacteria, the possibility of bacterial contamination has to be taken into account. However, spliced-leader sequences in the GSA-AT transcripts from *Karenia brevis* (CAMNT\_0032523625 and CAMNT\_0032609079, see Fig 4 for details) indicate that the genes are located in the dinoflagellate nucleus. The respective translated sequences contain N-terminal presequences putatively targeting the protein product to the peridinin plastid [71] (Figs 1 and 4). There is also a single sequence from *G. foliaceum* closely related to *Pelomonas* sp. in the bacterial cluster; this particular gene might be a bacterial contaminant (Fig 4). The origin of the novel GSA-AT clade remains unclear.

We also analyzed the last enzyme of the pathway, ferrochelatase (FeCH), not included in previously published analyses [47]. Ferrochelatase in particular shows complex origins in eukaryotic phototrophs, which includes non-endosymbiotic gene transfer from proteobacteria to the ancestor of chromerids and apicomplexans (Fig 3). Two clades in the ferrochelatase tree contain dinotoms: the genes of cyanobacterial origin came from the diatom endosymbiont; this clade also contains the heterotrophic dinoflagellate *Oxyrrhis marina*, numerous cyanobacteria, a glaucophyte, the rhizarian *Paulinella chromatophora*, chlorophytes, plants, heterokonts, eustigmatophytes, a haptophyte, *Euglena gracilis* and chromerids. The other clade is sister to apicomplexans, chromerids, and *E. gracilis*, and contains rhodophytes, peridinin-pigmented dinoflagellates, and a heterotrophic dinoflagellate (*Cryptheconidium cohni*). Its origin is unclear but might be proteobacterial (Fig 3). The tree topology is consistent with the presence of two ferrochelatases in chromerids [57], phototrophic euglenids [55], and dinotoms. While in apicomplexan parasites the complex origin of ferrochelatase is a result of non-endosymbiotic gene transfer from proteobacteria to Apicomplexa, the ferrochelatases in dinotoms arose from endosymbiotic association with the ancestor of the peridinin plastid and later tertiary endosymbiosis with the diatom endosymbiont.

The heme pathway is redundant in dinotoms; however, there are putatively necessary genes missing from their transcriptomes (Fig 1). The most striking absence is that of dinoflagellate-like UROS. We are unable to unambiguously discriminate between endosymbiont and host nuclear-encoded enzymes based solely on their sequences as spliced leaders are often missing; still, it appears that the diatom-like enzymes are exclusively used by the diatom endosymbiont, mainly due to retained characteristics required for diatom-like protein transport (the ASAFAP motif [47,72]) as well as difficulties in the hypothetical transport of proteins from the host cytoplasm over 6 membranes (endomembrane system of the host + putative plasma membrane of endosymbiont + four membranes of the diatom plastid). Conversely, any transport of UROS from inside the diatom endosymbiont compartment to the eyespot (remnant of the original dinoflagellate peridinin plastid) is hard to imagine. The absence of the original dinoflagellate UROS could be explained by transport of pre-uroporphyrinogen (hydroxymethylbilane), however, pre-uroporphyrinogen is highly unstable [73]. Furthermore, all the antecedent enzymes in the eyespot pathway (GTR, GSA-AT, ALAD, PBGD) would become redundant and therefore should have been lost from the genome. Insufficient sequencing and high divergence of UROS may explain the total absence of transcripts of the dinoflagellate-like UROS. On the other hand, the absence of *Glenodinium* orthologs of KfUROD1, KfUROD4 and the KfCPOX3+4+5 cluster and the *Kryptoperidinium* ortholog of the GfCPOX3+4 cluster (S1 Fig) may also be a result of gene loss, as other functional copies remained. In several cases, sequencing and assembly errors may interfere with determining the exact number of closely related paralogs, GSA-AT being an example of high gene copy number in dinoflagellates (Fig 4).

## The rhodophyte pathway is conserved in algae with green plastids

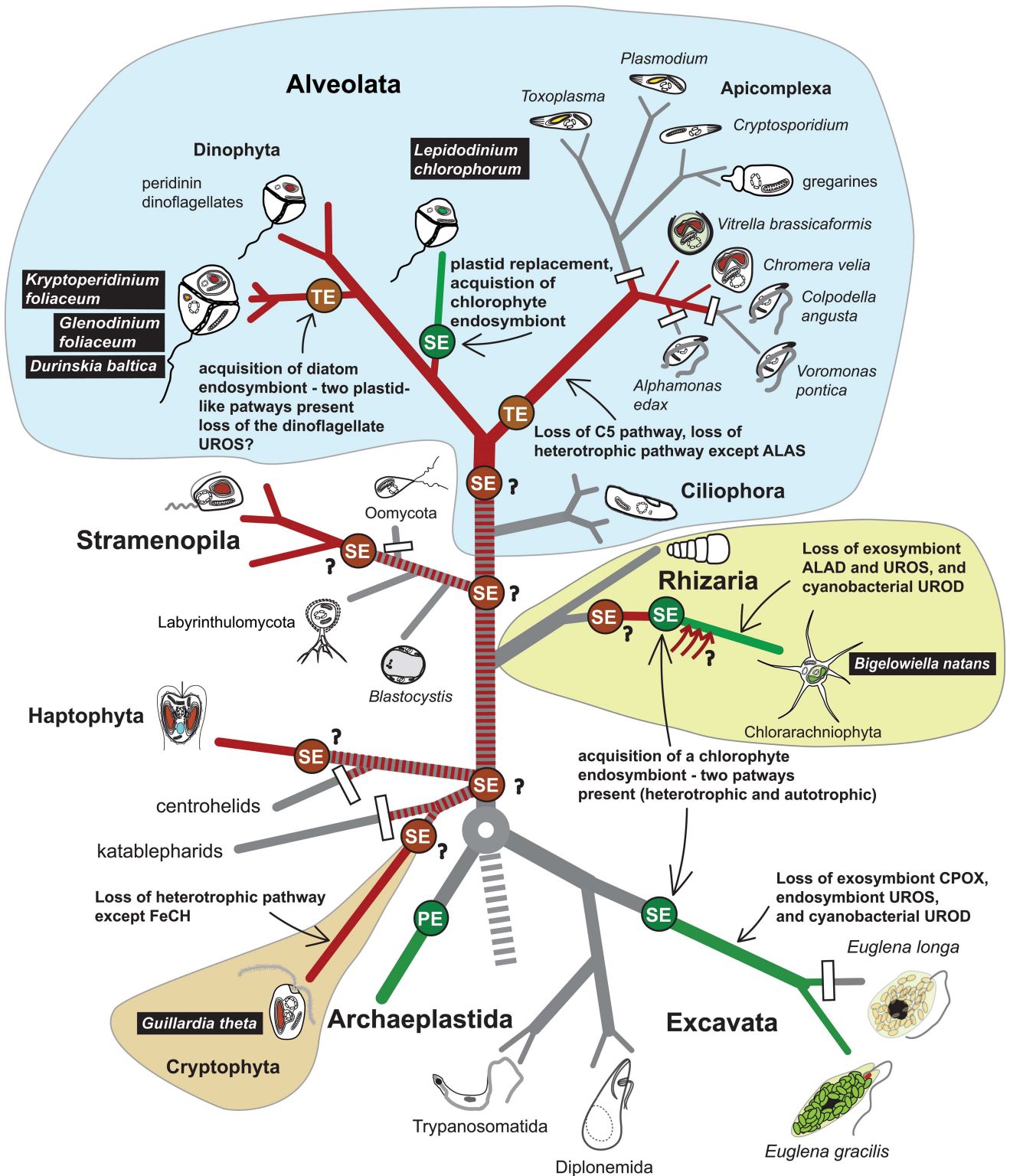
Using our transcriptomic data (S2 Table), we mapped the tetrapyrrole pathway in the green dinoflagellate *Lepidodinium chlorophorum* [24,74]. We found most of the genes of the plastid tetrapyrrole pathway and no traces of the mitochondrial-cytosolic pathway (see Fig 1 for summaries), consistent with the autotrophic history of the species. Some enzymes (LcGTR, LcPPOX2 and LcFeCH) cluster with the red lineage (S1 Fig). Other sequences branch with green algae and plants but always also branch together with other dinoflagellate sequences (LcALAD, LcGSA-AT1, -2, LcPPOX1). For example, sequences of canonical GSA-AT from *L. chlorophorum* form a cluster with sequences from the red lineage (Fig 4); however streptophytes, chlorophytes and rhodophytes are unresolved and placed in the ancestral position (see Figs 4 and S1 for details).

Interpreting most phylogenetic trees is very complicated, mainly due to the existence of multiple genes originating in the host nucleus, the endosymbiont nucleus, cyanobacteria (plastid), and proteobacteria. Together these factors make the evolution of heme pathway enzymes difficult to follow, particularly in eukaryotic phototrophs. However, the phylogenetic placement of *L. chlorophorum* GSA-AT and other enzymes together with their orthologs from dinoflagellates with rhodophyte-derived plastids suggests a red origin. In LcALAD and LcPPOX, with affinities to green algae, we presume the topology could be the result of the duplication of cyanobacteria-derived genes and subsequent lineage-specific gene loss (Fig 2), similar to GtUROD3. In other words, the hypothetical ancestor of the plastid contained two paralogs, for clarity here denoted as red and green, and one these paralogs was later lost in each lineage (the green one in rhodophytes and stramenopiles, the red one in the green lineage and dinoflagellates), masking the true origin of the gene. A clear example is seen in PPOX: the gene passed through a duplication event (S1 Fig) and *L. chlorophorum* genes are present in both cyanobacterial PPOX clades, with either rhodophytes or chlorophytes at the root. Again, they group together with other algae possessing rhodophyte-derived plastids. This suggests the green paralog was inherited vertically, not via endosymbiotic gene transfer from the green endosymbiont. It is noteworthy that the paralogs of ALAD and PPOX retained in rhodophytes also remained in stramenopiles, haptophytes and dinotoms, while the genes found in green algae and plants are present only in dinoflagellates. Consequently, most of the “red-related” genes in dinotoms apparently originate from, and reside in, the diatom endosymbiont (Figs 1 and 2, S1) and the ancestral dinoflagellate genes appear “green-related”. An alternative hypothesis would imply horizontal (eukaryote-to-eukaryote) gene transfer of ALAD and PPOX from green algal prey or from a putative green plastid to the ancestor of dinoflagellates, to the exclusion of chromerids and apicomplexans that possess a red-related gene. This putative green plastid would appear cryptic from today’s perspective, as it must have been later functionally replaced and partially genetically masked by the current red-derived peridinin plastid.

Therefore, it appears that most of the enzymes considered here originate from a rhodophyte source, in spite of the dinoflagellate’s chlorophyte-derived plastid. The chlorophyte-derived plastid is thought to have replaced the original peridinin pigmented dinoflagellate plastid through serial secondary endosymbiosis in this species [23]. Regardless, we observed that the original rhodophyte-derived pathway introduced with the peridinin plastid is highly conserved in this dinoflagellate, in agreement with the “shopping bag” or plastid promiscuity hypothesis, resulting in a mosaic evolution of the plastid proteome [75]. This functional conservation of tetrapyrrole biosynthesis genes might result from a predisposition of red genes to be targeted to the new plastid (as they were already successfully targeted to the old one)—an advantage the newcomer green genes presumably lacked.

The predominantly rhodophyte origins of the heme pathway in *L. chlorophorum* likely represent a set from the previously acquired peridinin plastid, but the “purchased” old genes were put into a newer, better shopping bag. The presence of red-derived heme pathway enzymes in *B. natans* plastids could be a result of non-endosymbiotic HGT from (red) algal prey according to the “you are what you eat” hypothesis [76]. Indeed, a rich fraction of red-related genes in *B. natans*, including photosynthesis-related genes, has already been reported [77,78]. Additional photosynthesis-related proteins were recruited from bacteria, indicating that extensive HGT does take place in chlorarachniophytes [77]. One explanation for the ease of HGT in this case comes from the origin of heme pathway genes; *B. natans* sequences often cluster with the CASH taxa. If these genes were horizontally transferred from one of the CASH lineages, they already coded for some plastid targeting signal as heme synthesis is presumably plastid-located in these lineages. *Bigeloviella natans* does not use the same protein import complex (SELMA) as algae from the CASH taxa [79], but plastid proteins in these groups have some structural similarities (e.g. the presence of a bipartite N terminal extension comprising a signal peptide plus a transit peptide) suggesting similarities also in the protein transport mechanisms. On the other hand, the abovementioned HGT events from the red lineage must have taken place after the green plastid acquisition but before the divergence of two basal chlorarachniophyte lineages comprising *Lotharella amoebiformis* and *B. natans* [78]. Altogether, it seems less probable that massive gene replacement via non-endosymbiotic HGT would occur in enzymes forming an essential and compartmentalized metabolic pathway. Taking into account similarities with *L. chlorophorum*, we can speculate that the green dinoflagellate “heritage” scenario also applies to *B. natans*. The rhodophyte origin of the tetrapyrrole pathway in *B. natans* may therefore reflect the previous presence of a hypothetical red-derived plastid in the ancestor of chlorarachniophytes (Fig 5). Rhizarians exhibit predatory heterotrophic or parasitic lifestyles, and a cryptic plastid could be held initially as a kleptoplast [80]. Similarly in dinoflagellates, the origins of “green” ALAD and PPOX may trace back to a cryptic endosymbiosis or gene transfer from the green lineage in the common ancestor of extant dinoflagellates including *Oxyrrhis* (Fig 2), mirroring the gene flow from the red lineage observed in *B. natans* [78]. However, the number of genes significantly related to the green lineage is extremely limited in studied dinoflagellates and *Chromera velia* [33,81,82] and does not necessarily imply the cryptic introduction of a green plastid. Indeed, the observed topologies might be artifacts showing false phylogenetic affinities. However, considering balanced sampling of each higher taxon (CASH taxa, green algae plus plants, red algae), this is less likely to happen in all cases.

In order to determine how endosymbiotic events occurred, a robust reconstruction of the gene repertoire of photosynthetic algae is needed. Not all genes diverged at the same time and gene multiplication and lineage-specific losses may hinder phylogenetic signal resolution. Indeed, genes with conserved or ancient evolutionary histories display different topologies than those acquired more recently via HGT [34] or those possessing less conserved functions [83]. Curtis et al. [31] reported a high number of algal-related genes in *G. theta* that acquired new functions and putatively also cellular localizations through endosymbiotic gene transfer to the host nucleus, regardless of their evolutionary origin. This is in conflict with the conserved origins implicated in this study; we suggest that enzymes of essential plastid pathways, such as tetrapyrrole biosynthesis, resist functional replacement due to conserved localization to a specialized compartment. The protein transport mechanism (SELMA), present in all investigated CASH taxa, is strong molecular evidence for the monophyletic origin of the CASH plastid [5,6,14]. During the course of evolution, proteins transported into the plastid via this mechanism have acquired an N-terminal transport signal. This potentially enables their lateral movement to new eukaryote hosts and allows them to maintain their original functions inside the organelle, provided the same transport mechanism is employed in the new host.



**Fig 5. A simplified scheme of the evolution of the tetrapyrrole biosynthesis pathway with an emphasis on the models from this study (black bars).** Primary endosymbiosis (PE) gave rise to the Archaeplastida comprising red algae, green algae and glaucophytes. Following the divergence of main eukaryotic lineages, secondary (SE) or tertiary endosymbiosis (TE) events equipped the ancestors of CASH taxa (cryptophytes, alveolates, stramenopiles and haptophytes) with photosynthetic capabilities. Contradictory evidence has been debated over the last years as for the history of CASH plastid acquisitions (e.g. [5,8,34]). A plastid-early scenario (the chromalveolate hypothesis) posits that all

CASH taxa are monophyletic and the CASH plastid was vertically transferred (dashed red line) and lost in extant non-photosynthetic descendants (such as ciliates and most rhizarians). Plastid-late scenarios require multiple lateral acquisitions of the CASH plastid (question marks) but better reflect some current phylogenomic analyses of the plastid recipients (e.g. [18,69]). Loss of photosynthesis/plastids have been documented in many sister lineages, such as oomycetes or apicomplexans, however these are in line with plastid-late scenarios as well. A cryptic SE with a CASH alga or numerous HGT (red arrows) events are inferred before the divergence of extant chlorarachniophytes (this work, [78]), which was masked by the acquisition of the current green algal endosymbiont. A similar situation in *L. chlorophorum* led to the peridinin plastid replacement with a green plastid, however the majority of red-related heme pathway enzymes remained functionally conserved in the successor plastid. The loss of the heterotrophic pathway possibly occurred several times independently in the stramenopile and dinoflagellate lineages, as *Perkinsus marinus*, sister to dinoflagellates, still contains a functional mitochondrial-cytosolic pathway [8].

doi:10.1371/journal.pone.0166338.g005

This is consistent with theories of lateral CASH plastid transfer into unrelated branches of the eukaryote tree, giving rise to the CASH taxa [34]. Conversely in cases of serial plastid replacement, a potential new plastid would encounter a pre-existing, functional set of proteins ensuring the function of the original organelle. The new plastid might also be inhabited by the original plastid's protein compendium including the protein transport machinery, rather than continuing to use its own proteome, which would be incompatible with SELMA. With this in mind, we presume that proteins having the ability to be transported to the CASH plastid could also be transported into a successor plastid, enabling the detection of serial endosymbiotic histories with higher confidence than cytoplasmic proteins.

## Conclusions

The tetrapyrrole biosynthesis pathway in phototrophic eukaryotes is an evolutionary mosaic originating in proteobacteria, cyanobacteria and eukaryotes. It represents a shopping bag of enzymes collected during the history of plastid endosymbiosis retained, due to its essential role in metabolism, even after photosynthetic capabilities have been lost. Here we confirm that the tertiary plastids of dinotoms represent largely independent compartments with tetrapyrrole biosynthesis occurring parallel to biosynthesis in the peridinin plastid. The enzymes putatively localized to the former plastid branch sister to dinoflagellate enzymes, while the tertiary plastid contains enzymes branching sister to those of diatoms, mirroring the origin of the respective organelles. In *G. theta*, the pathway is located almost entirely in the plastid, with the exception of a eukaryotic ferrochelatase apparently localized to the mitochondrion, indicating either a slow evolutionary rate or an evolutionary constraint. Furthermore, we observed that the majority of the pathway is evolutionarily conserved and related to the red lineage even in organisms that currently possess a plastid of green algal provenance, i.e. the dinoflagellate *Lepidodinium chlorophorum* and the chlorarachniophyte *B. natans*. Hence, if the protein targeting machinery is compatible with the new plastid compartment, the tetrapyrrole synthesis pathway can be relocated "as is", which is illustrated in the case of *L. chlorophorum*. Intriguingly, such a scenario may imply the existence of a cryptic red-derived plastid earlier in the history of chlorarachniophytes. While the evolution of eukaryotes is becoming clearer with increasing data from deeper lineages, the history of plastid acquisitions resists revealing an unequivocal scenario due to massive gene transfer and phylogenetic bias. We suggest that a targeted approach directed at conserved processes could result in new, relevant hypotheses even in the genomic era.

## Material and Methods

The complete genomic sequences of the cryptophyte alga *Guillardia theta* (<http://genome.jgi.doe.gov/Guith1/Guith1.home.html>) and the chlorarachniophyte *Bigeloviella natans* (<http://genome.jgi.doe.gov/Big1a1/Big1a1.home.html>) were searched using BLAST [84] for genes encoding enzymes involved in the synthesis of tetrapyrroles (ALAS, GTR, GSA-AT, ALAD,

PBGD, UROS, UROD, CPOX, PPOX, and FeCH). Homologous sequences were those used in Kořený *et al.* [57]; newly added sequences are listed in S3 Table. All alignments were made using MUSCLE [85] and ambiguous regions were removed in SeaView [86]. Phylogenetic trees were constructed using Maximum Likelihood (RAxML v8.2.4; [87]), Bayesian inference (PHYLOBAYES v3.3b; [88]) and a method designed to deal with amino acid saturation (AsaturA v18.10.2002; [89]). ML trees were computed using the LG model with gamma distribution in 4 categories and 1000 replicates. Bayesian inferences were calculated with the following parameters: 2 chains, 15,000 generations under the C20 model with Poisson exchange rate, sampling every 100 generations, and a maximum divergence of 0.1.

Sequences from *G. theta* and *B. natans* were inspected for the presence of N-terminal leader sequences by SignalP 3.0 [63] and TargetP [64], predicting localization to either the mitochondrion (mitochondrial transit peptide) or the plastid (bipartite leader composed of ER signal peptide followed by a transit peptide). GSA-AT of *Karenia brevis* (Fig 4) were automatically annotated using the InterProScan feature of Geneious 8.1 [90].

The transcriptome library of *Lepidodinium chlorophorum* (Roscoff collection no. RCC1488) was generated using the NEBNext Ultra Directional RNA Library kit (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's instructions. Quality assessment and sequencing were performed in a specialized facility, using the Illumina MiSeq (2×250 bp) platform. The generated reads were quality-trimmed using the FASTQ Toolkit (v1.0) of the Illumina BaseSpace platform and then assembled using Trinity v2.1.1 [91] and SOAPdenovo2 v2.0 r240 [92] and clustered using CAP3 [93]. Gene assembly completion was assessed with BUSCO software using the complete eukaryotic gene profile [94] on protein models generated by the TransDecoder script of the Trinity package. Some characteristics of the obtained transcriptome are listed in S2 Table. Novel sequences of interest were deposited in GenBank under the accession no. KX344033-47.

## Supporting Information

**S1 Fig. Bayesian phylogenetic trees as inferred from the amino acid sequences.** A) ALAS, B) GTR, C) PBGD, D) UROS, E) UROD, F) CPOX and G) PPOX. Taxa of interest of this study are highlighted by colored bars: blue for dinotoms, green for *Lepidodinium chlorophorum*, ochre for *Bigelowiella natans* and red for *Guillardia theta*. The tree demonstrates the mitochondrial origin of ALAS. Numbers near branches indicate Bayesian posterior probabilities followed by the bootstrap of respective clades from the likelihood analysis. Only support values greater than 0.85 (Bayesian) and 50 (likelihood) are shown. dt—different topology in the likelihood tree, see S2 Fig; a dash indicates unsupported topology. Asterisks mark possible contaminations. LcPPOXa, -b, -c; LcUROSa, -b = non-overlapping protein models, putatively fragments of LcPPOX1 and LcUROS.  
(PDF)

**S2 Fig. Maximum likelihood trees as inferred from amino acid sequences.** Numbers near branches indicate bootstrap values; only support values greater than 50 are shown. A, ALAS—delta-aminolevulinic acid synthase; B, GTR—glutamate-tRNA reductase; C, GSA—glutamate-1-semialdehyde aminotransferase; D, ALAD—aminolevulinic acid dehydratase; E, PBGD—porphobilinogen deaminase; F, UROS—uroporphyrinogen synthase; G, UROD—uroporphyrinogen decarboxylase; H, CPOX—coproporphyrinogen oxidase; I, PPOX—protoporphyrinogen oxidase; J, FeCH—ferrochelatase. LcPPOXa, -b, -c; LcUROSa, -b = non-overlapping protein models, putatively fragments of LcPPOX1 and LcUROS.  
(PDF)



**S1 Table. Targeting presequences in *Bigeloviella natans* and *Guillardia theta*.** Targeting probabilities were determined using SignalP and TargetP as described in Materials and Methods. Respective targeting peptide sequences are listed. The presence of a signal peptide followed by a chloroplast targeting peptide (cTP) implies localization to the plastid; mitochondrial enzymes encode mitochondrial targeting peptide presequences, while cytoplasmic enzymes lack presequences. If available, models with longer N-termini (e.g. Pasa, Fgenesh) were included in pre-sequence analysis and are listed in the table.  
(PDF)

**S2 Table. Characteristics of the obtained transcriptome libraries of *Lepidodinium chlorophorum*.** Number of reads and bases of two libraries are listed after quality trimming (two reads per library, see [Material and Methods](#)). The resulting number of contigs and coding sequences were analyzed using the BUSCO pipeline with a set of 429 BUSCO groups of orthologs. Ortholog counts: C = complete; D = duplicated; F = fragments; M = missing.  
(PDF)

**S3 Table. List of sequences added to the original datasets of Kořený *et al.* [26, 28].** Gene copy designation for species of interest in this study is shown in brackets according to their designation in respective trees (an asterisk marks putative contaminant sequences). Databases: CGP = Cyanophora Genome Project; CryptoDB = Cryptosporidium Genomic Resource; gb = GenBank; Gruber *et al.* 2015 Plant J, 10.1111/tpj.12734; jgi = DOE Joint Genome Institute; MMETSP = Marine Microbial Eukaryote Transcriptome Sequencing Project; Nori = NoriBLAST, Porphyra Genome Project; psb = bioinformatics.psb.ugent.be; VH = Courtesy of Vladimír Hampl, unpublished; Cmb = combined samples.  
(PDF)

## Acknowledgments

We acknowledge computation resources provided by CERIT-SC and MetaCentrum, Brno, Czech Republic. We thank Heather J. Esson for language correction, Luděk Kořený for helpful discussions and John M. Archibald for providing original sequence data before their public release via JGI.

## Author Contributions

**Conceptualization:** MO.

**Investigation:** JC ZF AH.

**Methodology:** MO.

**Visualization:** JC ZF MO.

**Writing – original draft:** JC ZF MO.

## References

1. McFadden GI, Guy L, Saw JH, Ettema TJG, Eme L, Sharpe SC, et al. Origin and evolution of plastids and photosynthesis in eukaryotes. *Cold Spring Harb Perspect Biol.* 2014; 6. doi: [10.1101/cshperspect.a016105](https://doi.org/10.1101/cshperspect.a016105) PMID: [24691960](https://pubmed.ncbi.nlm.nih.gov/24691960/)
2. Nowack ECM, Melkonian M, Glöckner G. Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Curr Biol.* 2008; 18: 410–418. doi: [10.1016/j.cub.2008.02.051](https://doi.org/10.1016/j.cub.2008.02.051) PMID: [18356055](https://pubmed.ncbi.nlm.nih.gov/18356055/)

3. Keeling PJ. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu Rev Plant Biol.* 2013; 64: 583–607. doi: [10.1146/annurev-arplant-050312-120144](https://doi.org/10.1146/annurev-arplant-050312-120144) PMID: [23451781](https://pubmed.ncbi.nlm.nih.gov/23451781/)
4. 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. doi: [10.1038/ncomms6764](https://doi.org/10.1038/ncomms6764) PMID: [25493338](https://pubmed.ncbi.nlm.nih.gov/25493338/)
5. Zimorski V, Ku C, Martin WF, Gould SB. Endosymbiotic theory for organelle origins. *Curr Opin Microbiol.* 2014; 22: 38–48. doi: [10.1016/j.mib.2014.09.008](https://doi.org/10.1016/j.mib.2014.09.008) PMID: [25306530](https://pubmed.ncbi.nlm.nih.gov/25306530/)
6. Gould SB, Maier U-G, Martin WF. Protein import and the origin of red complex plastids. *Curr Biol.* 2015; 25: R515–R521. doi: [10.1016/j.cub.2015.04.033](https://doi.org/10.1016/j.cub.2015.04.033) PMID: [26079086](https://pubmed.ncbi.nlm.nih.gov/26079086/)
7. Archibald JM. Genomic perspectives on the birth and spread of plastids. *Proc Natl Acad Sci U S A.* 2015; 112: 10147–53. doi: [10.1073/pnas.1421374112](https://doi.org/10.1073/pnas.1421374112) PMID: [25902528](https://pubmed.ncbi.nlm.nih.gov/25902528/)
8. Waller RF, Gornik SG, Kořený L, Pain A. Metabolic pathway redundancy within the apicomplexan-dinoflagellate radiation argues against an ancient chromalveolate plastid. *Commun Integr Biol.* 2016; 9. doi: [10.1080/19420889.2015.1116653](https://doi.org/10.1080/19420889.2015.1116653) PMID: [27066182](https://pubmed.ncbi.nlm.nih.gov/27066182/)
9. Cavalier-Smith T. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Eukaryot Microbiol.* 1999; 46: 347–366. doi: [10.1111/j.1550-7408.1999.tb04614.x](https://doi.org/10.1111/j.1550-7408.1999.tb04614.x) PMID: [18092388](https://pubmed.ncbi.nlm.nih.gov/18092388/)
10. Lane CE, Archibald JM. The eukaryotic tree of life: endosymbiosis takes its TOL. *Trends Ecol Evol.* 2008; 23: 268–275. doi: [10.1016/j.tree.2008.02.004](https://doi.org/10.1016/j.tree.2008.02.004) PMID: [18378040](https://pubmed.ncbi.nlm.nih.gov/18378040/)
11. Sanchez-Puerta MV, Delwiche CF. A hypothesis for plastid evolution in chromalveolates. *J Phycol.* 2008; 44: 1097–1107. doi: [10.1111/j.1529-8817.2008.00559.x](https://doi.org/10.1111/j.1529-8817.2008.00559.x) PMID: [27041706](https://pubmed.ncbi.nlm.nih.gov/27041706/)
12. Bodyl A, Stiller JW, Mackiewicz P. Chromalveolate plastids: direct descent or multiple endosymbioses? *Trends Ecol Evol.* 2009; 24: 119–121. doi: [10.1016/j.tree.2008.11.003](https://doi.org/10.1016/j.tree.2008.11.003) PMID: [19200617](https://pubmed.ncbi.nlm.nih.gov/19200617/)
13. Janouškovec J, Horák A, Oborník M, Lukeš J, Keeling PJ. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc Natl Acad Sci U S A.* 2010; 107: 10949–54. doi: [10.1073/pnas.1003335107](https://doi.org/10.1073/pnas.1003335107) PMID: [20534454](https://pubmed.ncbi.nlm.nih.gov/20534454/)
14. Felsner G, Sommer MS, Gruenheit N, Hempel F, Moog D, Zauner S, et al. ERAD components in organisms with complex red plastids suggest recruitment of a preexisting protein transport pathway for the periplastid membrane. *Genome Biol Evol.* 2011; 3: 140–150. doi: [10.1093/gbe/evq074](https://doi.org/10.1093/gbe/evq074) PMID: [21081314](https://pubmed.ncbi.nlm.nih.gov/21081314/)
15. Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, et al. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic “supergroups”. *Proc Natl Acad Sci U S A.* 2009; 106: 3859–64. doi: [10.1073/pnas.0807880106](https://doi.org/10.1073/pnas.0807880106) PMID: [19237557](https://pubmed.ncbi.nlm.nih.gov/19237557/)
16. Brown MW, Sharpe SC, Silberman JD, Heiss A a, Lang BF, Simpson AGB, et al. Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proc Biol Sci.* 2013; 280: 20131755. doi: [10.1098/rspb.2013.1755](https://doi.org/10.1098/rspb.2013.1755) PMID: [23986111](https://pubmed.ncbi.nlm.nih.gov/23986111/)
17. Yabuki A, Kamikawa R, Ishikawa S a, Kolisko M, Kim E, Tanabe AS, et al. *Palpitomonas bilix* represents a basal cryptist lineage: insight into the character evolution in Cryptista. *Sci Rep.* 2014; 4: 4641. doi: [10.1038/srep04641](https://doi.org/10.1038/srep04641) PMID: [24717814](https://pubmed.ncbi.nlm.nih.gov/24717814/)
18. Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, et al. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc R Soc B Biol Sci.* 2016; 283: 20152802. doi: [10.1098/rspb.2015.2802](https://doi.org/10.1098/rspb.2015.2802) PMID: [26817772](https://pubmed.ncbi.nlm.nih.gov/26817772/)
19. Delwiche CF. Tracing the thread of plastid—Diversity through the tapestry of life. *Am Nat.* 1999; 154: S164–S177. Available: <http://www.jstor.org/stable/10.2307/2463984> doi: [10.1086/303291](https://doi.org/10.1086/303291) PMID: [10527925](https://pubmed.ncbi.nlm.nih.gov/10527925/)
20. Inagaki Y, Dacks JB, Ford Doolittle W, Watanabe KI, Ohama T. Evolutionary relationship between dinoflagellates bearing obligate diatom endosymbionts: Insight into tertiary endosymbiosis. *Int J Syst Evol Microbiol.* 2000; 50: 2075–2081. doi: [10.1099/00207713-50-6-2075](https://doi.org/10.1099/00207713-50-6-2075) PMID: [11155982](https://pubmed.ncbi.nlm.nih.gov/11155982/)
21. Saldarriaga JF, Taylor FJR, Keeling PJ, Cavalier-smith T. Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses and replacements. *J Mol Evol.* 2001; 53: 204–213. doi: [10.1007/s002390010210](https://doi.org/10.1007/s002390010210) PMID: [11523007](https://pubmed.ncbi.nlm.nih.gov/11523007/)
22. Ishida K, Green BR. Second- and third-hand chloroplasts in dinoflagellates: Phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *Proc Natl Acad Sci U S A.* 2002; 99: 9294–9299. doi: [10.1073/pnas.142091799](https://doi.org/10.1073/pnas.142091799) PMID: [12089328](https://pubmed.ncbi.nlm.nih.gov/12089328/)
23. Keeling PJ. The endosymbiotic origin, diversification and fate of plastids. *Philos Trans R Soc Lond B Biol Sci.* 2010; 365: 729–48. doi: [10.1098/rstb.2009.0103](https://doi.org/10.1098/rstb.2009.0103) PMID: [20124341](https://pubmed.ncbi.nlm.nih.gov/20124341/)

24. Elbrachter M, Schnepf E. *Gymnodinium chlorophorum*, a new, green, bloom-forming dinoflagellate (Gymnodiniales, Dinophyceae) with a vestigial prasinophyte endosymbiont. *Phycologia*. 1996; 35: 381–393. doi: [10.2216/I0031-8884-35-5-381.1](https://doi.org/10.2216/I0031-8884-35-5-381.1)
25. Takishita K, Kawachi M, Noel MH, Matsumoto T, Kakizoe N, Watanabe MM, et al. Origins of plastids and glyceraldehyde-3-phosphate dehydrogenase genes in the green-colored dinoflagellate *Lepidodinium chlorophorum*. *Gene*. 2008; 410: 26–36. doi: [10.1016/j.gene.2007.11.008](https://doi.org/10.1016/j.gene.2007.11.008) PMID: [18191504](https://pubmed.ncbi.nlm.nih.gov/18191504/)
26. Chesnick JM, Morden CW, Schmiege AM. Identity of the endosymbiont of *Peridinium foliaceum* (Pyrrophyta): Analysis of the rbcLS operon. *J Phycol*. 1996; 32: 850–857. doi: [10.1111/J.0022-3646.1996.00850.X](https://doi.org/10.1111/J.0022-3646.1996.00850.X)
27. Imanian B, Keeling PJ. The dinoflagellates *Durinskia baltica* and *Kryptoperidinium foliaceum* retain functionally overlapping mitochondria from two evolutionarily distinct lineages. *BMC Evol Biol*. 2007; 7: 172. doi: [10.1186/1471-2148-7-172](https://doi.org/10.1186/1471-2148-7-172) PMID: [17892581](https://pubmed.ncbi.nlm.nih.gov/17892581/)
28. Moustafa A, Beszteri BB, Maier UG, Bowler C, Valentin K, Bhattacharya D. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science*. 2009; 324: 1724–6. doi: [10.1126/science.1172983](https://doi.org/10.1126/science.1172983) PMID: [19556510](https://pubmed.ncbi.nlm.nih.gov/19556510/)
29. Dorrell RG, Smith AG. Do red and green make brown?: Perspectives on plastid acquisitions within chromalveolates. *Eukaryot Cell*. 2011; 10: 856–868. doi: [10.1128/EC.00326-10](https://doi.org/10.1128/EC.00326-10) PMID: [21622904](https://pubmed.ncbi.nlm.nih.gov/21622904/)
30. 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: 1220–1230. doi: [10.1093/gbe/evr100](https://doi.org/10.1093/gbe/evr100) PMID: [21965651](https://pubmed.ncbi.nlm.nih.gov/21965651/)
31. Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, et al. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature*. 2012; 492: 59–65. doi: [10.1038/nature11681](https://doi.org/10.1038/nature11681) PMID: [23201678](https://pubmed.ncbi.nlm.nih.gov/23201678/)
32. Deschamps P, Moreira D. Reevaluating the green contribution to diatom genomes. *Genome Biol Evol*. 2012; 4: 683–688. doi: [10.1093/gbe/evs053](https://doi.org/10.1093/gbe/evs053) PMID: [22684208](https://pubmed.ncbi.nlm.nih.gov/22684208/)
33. Burki F, Flegontov P, Oborník M, Cihlář J, Pain A, Lukeš J, et al. Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. *Genome Biol Evol*. 2012; 4: 626–635. doi: [10.1093/gbe/evs049](https://doi.org/10.1093/gbe/evs049) PMID: [22593553](https://pubmed.ncbi.nlm.nih.gov/22593553/)
34. Petersen J, Ludewig AK, Michael V, Bunk B, Jarek M, Baurain D, et al. *Chromera velia*, endosymbioses and the rhodoplex hypothesis—Plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). *Genome Biol Evol*. 2014; 6: 666–684. doi: [10.1093/gbe/evu043](https://doi.org/10.1093/gbe/evu043) PMID: [24572015](https://pubmed.ncbi.nlm.nih.gov/24572015/)
35. Ševčíková T, Horák A, Klimeš V, Zbránková V, Demir-Hilton E, Sudek S, et al. Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Sci Rep*. 2015; 5: 10134. doi: [10.1038/srep10134](https://doi.org/10.1038/srep10134) PMID: [26017773](https://pubmed.ncbi.nlm.nih.gov/26017773/)
36. Yamaguchi A, Yubuki N, Leander BS. Morphostasis in a novel eukaryote illuminates the evolutionary transition from phagotrophy to phototrophy: description of *Rapaza viridis* n. gen. et sp. (Euglenozoa, Euglenida). *BMC Evol Biol*. BioMed Central Ltd; 2012; 12: 29. doi: [10.1186/1471-2148-12-29](https://doi.org/10.1186/1471-2148-12-29) PMID: [22401606](https://pubmed.ncbi.nlm.nih.gov/22401606/)
37. Hrdá Š, Fousek J, Szabová J, Hampel V, Vlček Č. The plastid genome of *Eutreptiella* provides a window into the process of secondary endosymbiosis of plastid in euglenids. *PLoS One*. 2012; 7: e33746. doi: [10.1371/journal.pone.0033746](https://doi.org/10.1371/journal.pone.0033746) PMID: [22448269](https://pubmed.ncbi.nlm.nih.gov/22448269/)
38. Deusch O, Landan G, Roettger M, Gruenheit N, Kowallik K V., Allen JF, et al. Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Mol Biol Evol*. 2008; 25: 748–761. doi: [10.1093/molbev/msn022](https://doi.org/10.1093/molbev/msn022) PMID: [18222943](https://pubmed.ncbi.nlm.nih.gov/18222943/)
39. Huang S, Shingaki-Wells RN, Taylor NL, Millar AH. The rice mitochondria proteome and its response during development and to the environment. *Front Plant Sci*. 2013; 4: 16. doi: [10.3389/fpls.2013.00016](https://doi.org/10.3389/fpls.2013.00016) PMID: [23403394](https://pubmed.ncbi.nlm.nih.gov/23403394/)
40. Gross J, Bhattacharya D. Mitochondrial and plastid evolution in eukaryotes: an outsiders' perspective. *Nat Rev Genet*. 2009; 10: 495–505. doi: [10.1038/nrg2649](https://doi.org/10.1038/nrg2649) PMID: [19506574](https://pubmed.ncbi.nlm.nih.gov/19506574/)
41. Basak I, Moeller SG. Emerging facets of plastid division regulation. *Planta*. 2013. pp. 389–398. doi: [10.1007/s00425-012-1743-6](https://doi.org/10.1007/s00425-012-1743-6) PMID: [22965912](https://pubmed.ncbi.nlm.nih.gov/22965912/)
42. Gould SB, Waller RF, McFadden GI. Plastid evolution. *Annu Rev Plant Biol*. 2008; 59: 491–517. doi: [10.1146/annurev.arplant.59.032607.092915](https://doi.org/10.1146/annurev.arplant.59.032607.092915) PMID: [18315522](https://pubmed.ncbi.nlm.nih.gov/18315522/)
43. Barbrook AC, Howe CJ, Purton S. Why are plastid genomes retained in non-photosynthetic organisms? *Trends Plant Sci*. 2006; 11: 101–108. doi: [10.1016/j.tplants.2005.12.004](https://doi.org/10.1016/j.tplants.2005.12.004) PMID: [16406301](https://pubmed.ncbi.nlm.nih.gov/16406301/)
44. Smith DR, Crosby K, Lee RW. Correlation between nuclear plastid DNA abundance and plastid number supports the limited transfer window hypothesis. *Genome Biol Evol*. 2011; 3: 365–371. doi: [10.1093/gbe/evr001](https://doi.org/10.1093/gbe/evr001) PMID: [21292629](https://pubmed.ncbi.nlm.nih.gov/21292629/)

45. Gillott MA, Gibbs SP. Cryptomonad nucleomorph: its ultrastructure and evolutionary significance. *J Phycol.* 1980; 16: 558–568. doi: [10.1111/j.1529-8817.1980.tb03074.x](https://doi.org/10.1111/j.1529-8817.1980.tb03074.x)
46. Gilson PR, McFadden GI. Good things in small packages: the tiny genomes of chlorarachniophyte endosymbionts. *BioEssays.* 1997; 19: 167–173. doi: [10.1002/bies.950190212](https://doi.org/10.1002/bies.950190212) PMID: [9046247](https://pubmed.ncbi.nlm.nih.gov/9046247/)
47. Hehenberger E, Imanian B, Burki F, Keeling PJ. Evidence for the retention of two evolutionary distinct plastids in dinoflagellates with diatom endosymbionts. *Genome Biol Evol.* 2014; 6: 2321–2334. doi: [10.1093/gbe/evu182](https://doi.org/10.1093/gbe/evu182) PMID: [25172904](https://pubmed.ncbi.nlm.nih.gov/25172904/)
48. Imanian B, Keeling PJ. Horizontal gene transfer and redundancy of tryptophan biosynthetic enzymes in dinotoms. *Genome Biol Evol.* 2014; 6: 333–343. doi: [10.1093/gbe/evu014](https://doi.org/10.1093/gbe/evu014) PMID: [24448981](https://pubmed.ncbi.nlm.nih.gov/24448981/)
49. Dodge JD. The functional and phylogenetic significance of dinoflagellate eyespots. *BioSystems.* 1983; 16: 259–267. doi: [10.1016/0303-2647\(83\)90009-6](https://doi.org/10.1016/0303-2647(83)90009-6) PMID: [6687045](https://pubmed.ncbi.nlm.nih.gov/6687045/)
50. Kořený L, Sobotka R, Kovářová J, Gnipová A, Flegontov P, Horváth A, et al. Aerobic kinetoplastid flagellate *Phytomonas* does not require heme for viability. *Proc Natl Acad Sci U S A.* 2012; 109: 3808–3813. doi: [10.1073/pnas.1201089109](https://doi.org/10.1073/pnas.1201089109) PMID: [22355128](https://pubmed.ncbi.nlm.nih.gov/22355128/)
51. Weinstein JD, Beale SI. Separate Physiological roles and subcellular compartments for two tetrapyrrole biosynthesis pathway in *Euglena gracilis*. *J Biol Chem.* 1983; 258: 6799–6807. PMID: [6133868](https://pubmed.ncbi.nlm.nih.gov/6133868/)
52. Mayer SM, Beale SI, Weinstein JD. Enzymatic conversion of glutamate to  $\delta$ -aminolevulinic acid in soluble extract of *Euglena gracilis*. *J Biol Chem.* 1987; 262: 12541–12549. PMID: [2442164](https://pubmed.ncbi.nlm.nih.gov/2442164/)
53. Mayer SM, Beale S.  $\delta$ -Aminolevulinic acid biosynthesis from glutamate in *Euglena gracilis*. *Plant Physiol.* 1991; 97: 1094–1102. PMID: [16668494](https://pubmed.ncbi.nlm.nih.gov/16668494/)
54. Iida K, Mimura I, Kajiwara M. Evaluation of two biosynthetic pathways to  $\delta$ -aminolevulinic acid in *Euglena gracilis*. *Eur J Biochem.* 2002; 269: 291–7. Available: <http://www.ncbi.nlm.nih.gov/pubmed/11784323> PMID: [11784323](https://pubmed.ncbi.nlm.nih.gov/11784323/)
55. Kořený L, Oborník M. Sequence evidence for the presence of two tetrapyrrole pathways in *Euglena gracilis*. *Genome Biol Evol.* 2011; 3: 359–364. doi: [10.1093/gbe/evr029](https://doi.org/10.1093/gbe/evr029) PMID: [21444293](https://pubmed.ncbi.nlm.nih.gov/21444293/)
56. Oborník M, Green BR. Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Mol Biol Evol.* 2005; 22: 2343–2353. doi: [10.1093/molbev/msi230](https://doi.org/10.1093/molbev/msi230) PMID: [16093570](https://pubmed.ncbi.nlm.nih.gov/16093570/)
57. Kořený L, Sobotka R, Janouškovec J, Keeling PJ, Oborník M. Tetrapyrrole synthesis of photosynthetic chromerids is likely homologous to the unusual pathway of apicomplexan parasites. *Plant Cell.* 2011; 23: 3454–62. doi: [10.1105/tpc.111.089102](https://doi.org/10.1105/tpc.111.089102) PMID: [21963666](https://pubmed.ncbi.nlm.nih.gov/21963666/)
58. Janouškovec J, Tikhonenkov D V, Burki F, Howe AT, Kolísko M, Mylnikov AP, et al. Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *Proc Natl Acad Sci U S A.* 2015; 112: 10200–7. doi: [10.1073/pnas.1423790112](https://doi.org/10.1073/pnas.1423790112) PMID: [25717057](https://pubmed.ncbi.nlm.nih.gov/25717057/)
59. Fernández Robledo JA, Caler E, Matsuzaki M, Keeling PJ, Shanmugam D, Roos DS, et al. The search for the missing link: A relic plastid in *Perkinsus*? [Internet]. *International Journal for Parasitology.* 2011. pp. 1217–1229. doi: [10.1016/j.ijpara.2011.07.008](https://doi.org/10.1016/j.ijpara.2011.07.008) PMID: [21889509](https://pubmed.ncbi.nlm.nih.gov/21889509/)
60. Ralph SA, van Dooren GG, Waller RF, Crawford MJ, Fraunholz MJ, Foth BJ, et al. Metabolic maps and functions of the *Plasmodium falciparum* apicoplast. *Nat Rev Microbiol.* 2004; 2: 203–216. doi: [10.1038/nrmicro843](https://doi.org/10.1038/nrmicro843) PMID: [15083156](https://pubmed.ncbi.nlm.nih.gov/15083156/)
61. Kořený L, Lukeš J, Oborník M. Evolution of the haem synthetic pathway in kinetoplastid flagellates: An essential pathway that is not essential after all? *Int J Parasitol.* 2010; 40: 149–156. doi: [10.1016/j.ijpara.2009.11.007](https://doi.org/10.1016/j.ijpara.2009.11.007) PMID: [19968994](https://pubmed.ncbi.nlm.nih.gov/19968994/)
62. Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. The complete chloroplast genome of the chlorarachniophyte *Bigeloviella natans*: Evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Mol Biol Evol.* 2007; 24: 54–62. doi: [10.1093/molbev/msl129](https://doi.org/10.1093/molbev/msl129) PMID: [16990439](https://pubmed.ncbi.nlm.nih.gov/16990439/)
63. Bendtsen JD, Nielsen H, Von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol.* 2004; 340: 783–795. doi: [10.1016/j.jmb.2004.05.028](https://doi.org/10.1016/j.jmb.2004.05.028) PMID: [15223320](https://pubmed.ncbi.nlm.nih.gov/15223320/)
64. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc.* 2007; 2: 953–971. doi: [10.1038/nprot.2007.131](https://doi.org/10.1038/nprot.2007.131) PMID: [17446895](https://pubmed.ncbi.nlm.nih.gov/17446895/)
65. Masuda T, Suzuki T, Shimada H, Ohta H, Takamiya K. Subcellular localization of two types of ferrochelatase in cucumber. *Planta.* 2003; 217: 602–609. doi: [10.1007/s00425-003-1019-2](https://doi.org/10.1007/s00425-003-1019-2) PMID: [12905021](https://pubmed.ncbi.nlm.nih.gov/12905021/)
66. Maier UG, Rensing SA, Igloi GL, Maerz M. Twintrons are not unique to the *Euglena* chloroplast genome: structure and evolution of a plastome cpn60 gene from a cryptomonad. *Mol Gen Genet.* 1995; 246: 128–131. doi: [10.1007/BF00290141](https://doi.org/10.1007/BF00290141) PMID: [7823908](https://pubmed.ncbi.nlm.nih.gov/7823908/)
67. Douglas SE, Penny SL. The plastid genome of the cryptophyte alga, *Guillardia theta*: Complete sequence and conserved syntenic groups confirm its common ancestry with red algae. *J Mol Evol.* 1999; 48: 236–244. doi: [10.1007/PL00006462](https://doi.org/10.1007/PL00006462) PMID: [9929392](https://pubmed.ncbi.nlm.nih.gov/9929392/)

68. Woo YH, Ansari H, Otto TD, Klinger CM, Kolisko M, Michálek J, et al. Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *Elife*. 2015; 4: 1–41. doi: [10.7554/eLife.06974](https://doi.org/10.7554/eLife.06974) PMID: [26175406](https://pubmed.ncbi.nlm.nih.gov/26175406/)
69. Burki F, Okamoto N, Pombert J-F, Keeling PJ. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc Biol Sci*. 2012; 279: 2246–54. doi: [10.1098/rspb.2011.2301](https://doi.org/10.1098/rspb.2011.2301) PMID: [22298847](https://pubmed.ncbi.nlm.nih.gov/22298847/)
70. 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. doi: [10.1073/pnas.1110633108](https://doi.org/10.1073/pnas.1110633108) PMID: [21810989](https://pubmed.ncbi.nlm.nih.gov/21810989/)
71. Zhang H, Hou Y, Miranda L, Campbell DA, Sturm NR, Gaasterland T, et al. Spliced leader RNA trans-splicing in dinoflagellates. *Proc Natl Acad Sci U S A*. 2007; 104: 4618–4623. doi: [10.1073/pnas.0700258104](https://doi.org/10.1073/pnas.0700258104) PMID: [17360573](https://pubmed.ncbi.nlm.nih.gov/17360573/)
72. Gruber A, Vugrinec S, Hempel F, Gould SB, Maier UG, Kroth PG. Protein targeting into complex diatom plastids: Functional characterisation of a specific targeting motif. *Plant Mol Biol*. 2007; 64: 519–530. doi: [10.1007/s11103-007-9171-x](https://doi.org/10.1007/s11103-007-9171-x) PMID: [17484021](https://pubmed.ncbi.nlm.nih.gov/17484021/)
73. Jordan PM. The biosynthesis of uroporphyrinogen III: mechanism of action of porphobilinogen deaminase. *Ciba Found Symp*. NETHERLANDS; 1994; 180: 70–96.
74. Hansen G, Botes L, De Salas M. Ultrastructure and large subunit rDNA sequences of *Lepidodinium viride* reveal a close relationship to *Lepidodinium chlorophorum* comb. nov. (= *Gymnodinium chlorophorum*). *Phycol Res*. 2007; 55: 25–41. doi: [10.1111/j.1440-1835.2006.00442.x](https://doi.org/10.1111/j.1440-1835.2006.00442.x)
75. Larkum AWD, Lockhart PJ, Howe CJ. Shopping for plastids. *Trends Plant Sci*. 2007; 12: 189–195. doi: [10.1016/j.tplants.2007.03.011](https://doi.org/10.1016/j.tplants.2007.03.011) PMID: [17416546](https://pubmed.ncbi.nlm.nih.gov/17416546/)
76. 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. doi: [10.1016/S0168-9525\(98\)01494-2](https://doi.org/10.1016/S0168-9525(98)01494-2) PMID: [9724962](https://pubmed.ncbi.nlm.nih.gov/9724962/)
77. Archibald JM, Rogers MB, Toop M, Ishida K-I, Keeling PJ. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigeloviella natans*. *Proc Natl Acad Sci U S A*. 2003; 100: 7678–7683. doi: [10.1073/pnas.1230951100](https://doi.org/10.1073/pnas.1230951100) PMID: [12777624](https://pubmed.ncbi.nlm.nih.gov/12777624/)
78. 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. doi: [10.1371/journal.pone.0101158](https://doi.org/10.1371/journal.pone.0101158) PMID: [24972019](https://pubmed.ncbi.nlm.nih.gov/24972019/)
79. Hiraoka Y, Burki F, Keeling PJ. Genome-based reconstruction of the protein import machinery in the secondary plastid of a chlorarachniophyte alga. *Eukaryot Cell*. 2012; 11: 324–333. doi: [10.1128/EC.05264-11](https://doi.org/10.1128/EC.05264-11) PMID: [22267775](https://pubmed.ncbi.nlm.nih.gov/22267775/)
80. Stoecker DK, Johnson MD, De Vargas C, Not F. Acquired phototrophy in aquatic protists. *Aquat Microb Ecol*. 2009; 57: 279–310. doi: [10.3354/ame01340](https://doi.org/10.3354/ame01340)
81. Burki F, Imanian B, Hehenberger E, Hiraoka Y, Maruyama S, Keeling PJ. Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. *Eukaryot Cell*. 2014; 13: 246–255. doi: [10.1128/EC.00299-13](https://doi.org/10.1128/EC.00299-13) PMID: [24297445](https://pubmed.ncbi.nlm.nih.gov/24297445/)
82. Moreira D, Deschamps P. What was the real contribution of endosymbionts to the eukaryotic nucleus? insights from photosynthetic eukaryotes. *Cold Spring Harb Perspect Biol*. 2014; 6: 1–9. doi: [10.1101/cshperspect.a016014](https://doi.org/10.1101/cshperspect.a016014) PMID: [24984774](https://pubmed.ncbi.nlm.nih.gov/24984774/)
83. Pittis AA, Gabaldón T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature*. 2016; 531: 101–4. doi: [10.1038/nature16941](https://doi.org/10.1038/nature16941) PMID: [26840490](https://pubmed.ncbi.nlm.nih.gov/26840490/)
84. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215: 403–10. doi: [10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/)
85. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004; 32: 1792–1797. doi: [10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340) PMID: [15034147](https://pubmed.ncbi.nlm.nih.gov/15034147/)
86. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 2010; 27: 221–4. doi: [10.1093/molbev/msp259](https://doi.org/10.1093/molbev/msp259) PMID: [19854763](https://pubmed.ncbi.nlm.nih.gov/19854763/)
87. Stamatakis A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006; 22: 2688–2690. doi: [10.1093/bioinformatics/btl446](https://doi.org/10.1093/bioinformatics/btl446) PMID: [16928733](https://pubmed.ncbi.nlm.nih.gov/16928733/)
88. Lartillot N, Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol*. 2004; 21: 1095–1109. doi: [10.1093/molbev/msh112](https://doi.org/10.1093/molbev/msh112) PMID: [15014145](https://pubmed.ncbi.nlm.nih.gov/15014145/)

89. Van de Peer Y, Frickey T, Taylor JS, Meyer A. Dealing with saturation at the amino acid level: A case study based on anciently duplicated zebrafish genes. *Gene*. 2002; 295: 205–211. PMID: [12354655](#)
90. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012; 28: 1647–1649. doi: [10.1093/bioinformatics/bts199](#) PMID: [22543367](#)
91. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011; 29: 644–52. doi: [10.1038/nbt.1883](#) PMID: [21572440](#)
92. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*. 2012; 1: 18. doi: [10.1186/2047-217X-1-18](#) PMID: [23587118](#)
93. Huang X, Madan a. CAP 3: A DNA sequence assembly program. *Genome Res*. 1999; 9: 868–877. doi: [10.1101/gr.9.9.868](#) PMID: [10508846](#)
94. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015; 31: 3210–3212. doi: [10.1093/bioinformatics/btv351](#) PMID: [26059717](#)