# Evolution of Nuclear and Plastid Genomes in Dinoflagellates Experiencing Plastid Replacements

August 2018

Eriko MATSUO

# Evolution of Nuclear and Plastid Genomes in Dinoflagellates Experiencing Plastid Replacements

A Dissertation Submitted to
the Graduate School of Life and Environmental Sciences,
the University of Tsukuba
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Science
(Doctoral Program in Biological Sciences )

Eriko MATSUO

## Abstract

The ancestral dinoflagellate most likely established a peridinin-containing plastid, which have been inherited to the extant photosynthetic descendants. However, some species such as kareniacean dinoflagellates and *Lepidodinium* are known to have discarded ancestral peridinin-containing plastids and bear "non-canonical" plastids lacking peridinin, which were established through haptophyte and a green algal endosymbioses, respectively. In general the transformation of the endosymbiotic algae into the plastid should have associated with severe reduction of the endosymbiont genome and reorganization of host nuclear genome. Nevertheless, the whole picture of these genome evolutions still remain unclear in dinoflagellates bearing non-canonical plastids.

In this work, firstly I evaluated the impact of the endosymbiosis in a dinoflagellate cell on the plastid genome (pl-genome) of endosymbiont. I compared the pl-genomes between free-living pedinophytes (green algae), *Lepidodinium chlorophorum* and two undescribed dinoflagellate strains TRD-132 and MRD-151, of which current 'green-colored' plastids were separately acquired from pedinophyte endosymbionts. The size of the three pedinophyte-derived pl-genomes varies from ~66 to ~102 Kb and numbers of functionally assignable open reading frames (ORFs) they harbored ranges from 57 to 67 . I here propose that the pl-genomes of pedinophyte endosymbionts in the separate dinoflagellate cells commonly underwent reductive processes for two reasons. Firstly, the ORF repertories in the three pedinophyte-derived pl-genomes appeared to be subsets of those in the pedinophyte pl-genomes. Secondly, the pedinophyte-derived pl-genomes commonly but separately lost the inverted repeats (IRs), as all the three pedinophyte pl-genomes share the particular feature.

Next, I reported reorganization of nucleus-encoded genes in two kareniacean species (*Karenia* (*Ke.*) *brevis* and *Karlodinium* (*Kl.*) *veneficum*) and *L. chlorophorum*. For plastid function and maintenance, the aforementioned dinoflagellates were known to use nucleus-encoded proteins vertically inherited from the ancestral dinoflagellates (vertically inherited- or VI-type), and those acquired from non-dinoflagellate organisms (including the endosymbiont). These observations indicated that the proteomes of the non-canonical plastids derived from a haptophyte and a green alga were modified by "exogenous" genes acquired from non-dinoflagellate organisms. However, there was no systematic evaluation addressing

1

how "exogenous" genes reshaped individual metabolic pathways localized in a non-canonical plastid. In this study, I surveyed transcriptomic data from *Ke. brevis*, *Kl. veneficum* and *L. chlorophorum*, and identified proteins involved in three plastid metabolic pathways synthesizing chlorophyll *a* (Chl *a*), heme and isoprene. The origins of the individual proteins of my interest were investigated, and assessed how the three pathways were modified before and after the algal endosymbioses, which gave rise to the current non-canonical plastids. I found a clear difference in the contribution of VI-type proteins across the three pathways. In both *Ke. brevis*/*Kl. veneficum* and *L. chlorophorum*, I observed a substantial contribution of VI-type proteins to the isoprene and heme biosynthesises. In sharp contrast, VI-type protein was barely detected in the Chl *a* biosynthesis in the three dinoflagellates. In this study, I will propose a hypothesis to explain why acquisition of non-canonical plastid reorganized Chl *a* pathway intensely in both kareniacean and *L. chlorophorum*, while it gave little/weak impact on isoprene and heme biosynthesises was observed in any of the three dinoflagellates.

## Abbreviations

| | |
|---|---|
| ALA | aminolevulinic acid |
| ALAD | aminolevulinic acid dehydratase |
| BPP | Bayesian posterior probability |
| CDP-ME | 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol |
| CDP-MEP | 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol |
| Chl *a* | chlorophyll *a* |
| CPOX | coproporphyrinogen oxidase |
| CS | chlorophyll synthase |
| DVR | divinyl protochlorophyllide reductase |
| DXP | 1-deoxy-D-xylulose-5-phosphate |
| DXR | 1-deoxy-D-xylulose-5-phosphate reductoisomerase |
| DXS | 1-deoxy-D-xylulose-5-phosphate synthase |
| EA-type | endosymbiotically acquired-type |
| EGT | endosymbiotic gene transfer |
| FeCH | ferrochelatase |
| GSAT | glutamate-1-semialdehyde 2,1-aminomutase |
| GTR | glutamyl-tRNA reductase |
| HMB-PP | 1-hydroxy-2-methyl-2-butenyl 4-diphosphate |
| IPP | isopentenyl diphosphate |
| IRs | inverted repeats |
| *Ke. brevis* | *Karenia brevis* |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| *Kl. veneficum* | *Karlodinium veneficum* |
| KO | KEGG Orthology |
| KOID | KEGG Orthology identifier |
| LA-type | laterally acquired-type |
| MCMC | Markov Chain Monte Carlo |
| MEcPP | 2-C-methyl-D-erythritol 2,4-cyclodiphosphate |
| MEP | 2-C-methyl-D-erythritol 4-phosphate |

| | |
|---|---|
| MgCH | Mg chelatase |
| MgPME | Mg-protoporphyrin IX monomethyl ester |
| MgPMT | S-adenosylmethionine:Mg-protoporphyrin O-methyltransferase |
| ML | maximum-likelihood |
| MLBP | ML bootstrap value |
| MMETSP | The Marine Microbial Eukaryote Transcriptome Sequencing Project |
| NCBI | The National Center for Biotechnology Information |
| NIES | The National Institute for Environmental Studies |
| NM | nucelomorph |
| ORF | open reading frame |
| PBGD | porphobilinogen deaminase |
| pl-genome | plastid genome |
| POR | protochlorophyllide oxidoreductase |
| PPOX | protoporphyrinogen IX oxidase |
| SP | signal peptide |
| TP | transit peptide |
| UROD | uroporphyrinogen decarboxylase |
| UROS | uroporphyrinogen III synthase |
| VI-type | vertically inherited-type |

# Chapter 1: General Introduction

## 1. Evolution of photosynthetic eukaryotes

Oxygen-generating photosynthesis was initially established within prokaryotes, and then has been spread into phylogenetically diverse lineages of eukaryote. The photosynthetic reaction in eukaryotes occurs in organelles called plastids, which were most likely traced back to a single free-living cyanobacterium engulfed by a heterotrophic eukaryote (Archibald 2009; Howe et al. 2008; Palmer 2003; Reyes-Prieto and Bhattacharya 2007; Stiller et al. 2003). This endosymbiosis between a cyanobacterium and a eukaryote is estimated to arise approximately ˜2.1 billion years ago (Sánchez-Baracaldo et al, 2017), and referred to as primary endosymbiosis. The first photosynthetic eukaryote should have harbored a plastid surrounded by two membranes (primary plastid) that is the direct descendent of the endosymbiotic cyanobacterium. In the latter eukaryotic evolution, the ancestral eukaryotes with primary plastids were diverged into three extant groups of photosynthetic eukaryotes, namely glaucophytes, red algae and green plants comprising green algae and land plants. Phylogenetic diversity of photosynthetic eukaryotes was further expanded by "secondary endosymbioses" between primary plastid-bearing eukaryotes and heterotrophic eukaryotes, which gave rise to "complex" plastids surrounded by three or four membranes (Bhattacharya et al. 2004; Falkowski et al. 2004). Secondary endosymbioses occurred multiple times in eukaryotic evolution, as (i) complex plastids are sporadically distributed in the tree of eukaryotes and (ii) there are two types of complex plastids, one is of green algal origin and the other is of red algal origin. There are two separate secondary endosymbioses involved in green algae. The recent studies revealed that two phylogenetically distantly related groups, euglenids and chlorarachniophytes, established their current complex plastids from two distinct green algal groups, prasinophyte and ulvophyte green algae, respectively (Hibberd and Norris 1984; Gilson and McFadden 1996; Ishida et al. 1997; McFadden et al. 1994; Van de Peer et al. 1996; Hallick et al. 1993). The evolutionary process behind the diversity of red alga-derived complex plastids remains uncertain. Phylogenetic analyses based on plastid-encoded gene datasets repeatedly recovered the monophyly of red alga-derived complex plastids in stramenopiles, haptophytes, cryptophytes, chromerids and dinoflagellates

(Douglas and Penny 1999; Ishida et al. 1997; Moore et al. 2008; Sanchez-Puerta et al. 2007; Janouškovec et al, 2010). The plastid phylogenies indirectly but strongly suggest that the five eukaryotic lineages bearing red alga-derived complex plastids share an ancestry, and secondary endosymbiosis involved in a red alga occurred in their common ancestor. In contrary, no study has resolved the phylogenetic relationship among the host lineages bearing red alga-derived complex plastids. Without the robust evidence for the relationship among the host lineages bearing red alga-derived complex plastids, it is difficult to infer the precise number of red algal endosymbioses in the eukaryotic evolution.

Although the precise evolutionary process that transformed an endosymbiotic bacterium into a host-governed organelle has not been fully understood, the genome of the cyanobacterial endosymbiont most likely underwent severe reductive evolution during primary endosymbiosis (Martin et al. 2002). Principally, plastids retain their own genomes, which are descendents of the genome of the cyanobacterial endosymbiont engulfed by the common ancestor of glaucophytes, red algae and green plants. The plastid genomes (pl-genomes) sequenced to date range from ~70 to ~200 Kbp in size, being much smaller than the genomes of free-living cyanobacteria (approximately 3-11.6 Mbp) (Kaneko & Tabata 1997; Martin & Herrmann. 1998; Stucken et al, 2010; Ponce-Toledo et al, 2017). It is reasonable to assume that potentially a large number of genes became dispensable for an endosymbiotic life-style (e.g., those involved in cell wall synthesis or cell motility), and was discarded from the endosymbiont genome. At the same time, genes required to operate and maintain the obligate endosymbiont/plastid (henceforth I designate as plastid-related genes) have been translocated to the host nucleus. The gene transfer from the endosymbiont to host genomes, which is termed endosymbiotic gene transfer or EGT, has been considered as one of the key factors that enabled the host to enslave the endosymbiont/plastid. Likewise, severe reductive pressures most likely worked on the endosymbionts involved in secondary endosymbioses, disposing all of the intracellular structures of the endosymbiotic algae but plastids in the most of the lineages bearing complex plastids (Martin et al. 1998; 2002). It is noteworthy that cryptophytes and chlorarachniophytes still retain the residual nuclei (so-called nucelomorphs or NMs) in the compartments between the second and third membranes surrounding their plastids, which correspond to the cytosolic spaces of the endosymbiotic algae (Archibald, 2007; Eschbach et al, 1991; Gilson et al, 2002). Comparative genomic studies revealed that extremely reduced natures of both cryptophyte

and chlorarachniophyte NM genomes, nominating the two algal lineages as intermediates in the process transforming an endosymbiotic alga to the complex plastid. Secondary endosymbioses seemingly retouched the pl-genomes of green/red algal endosymbionts, as a certain degree of gene loss was observed by comparing the genomes of primary plastids to those of the corresponding complex plastids.

During primary and secondary endosymbioses, the host (nuclear) genomes acquired plastid-related genes via EGT. In the extant photosynthetic eukaryotes, diverse metabolisms are operated in plastids, including photosynthesis, depending on nucleus-encoded genes (Martin & Herrmann, 1998; van Dooren et al, 2001). The nucleus-encoded genes encoding proteins involved in plastid metabolisms are expressed by the host transcription and translation machineries. Such polypeptides need to be translocated from the cytosol to the plastid through multiprotein machinery on the organelle membranes. It can be easily presumed that the difference in the evolutions in both host and endosymbiont genomes is one of the major driving forces to diversify the extant photosynthetic lineages in eukaryotes. However, the detailed genome evolution associated with primary and secondary endosymbioses has yet to be understood, because the process transforming a photosynthetic endosymbiont into the plastid has been completed in most of extant photosynthetic eukaryotes. Thus, an eukaryote, which acquired the plastid through an endosymbiosis much more recently than the ancient endosymbioses, may preserve traits of the genome evolution associated with plastid acquisition.

## 2. "Non-canonical plastids" in dinoflagellates

In this study I focused on dinoflagellates, a group of aquatic unicellular photosynthetic eukaryotes that can provide important clues to unresolved issues in plastid evolution. Dinoflagellates belong to one of major taxonomic groups of eukaryotes, Alveolata, and about half species of dinoflagellates described to date are photosynthetic (Taylor et al, 2008). Typical photosynthetic dinoflagellates harbor golden-brown colored plastids, which were remnants of a red algal endosymbiont captured by the common ancestor of dinoflagellates (Hoek et al, 1995; Janouskovec et al, 2010). Comparing to red alga-derived plastids in other eukaryotic algae, the plastids in vast majority of dinoflagellates are unique in

containing a carotenoid called peridinin along with chlorophylls *a* and *c* as principal pigments (Jeffrey et al, 1975; Zapata et al, 2012). The pl-genomes of typical (i.e. peridinin-containing) dinoflagellates are known to comprise multiple, small circular DNA chromosomes (~2–6 Kbp) termed "minicircles" (Zang et al, 1999), each of which carries a single, or occasionally a few genes. So far, only ribosomal RNAs (rRNAs) and a limited number of proteins for photosynthesis were found in minicircles (Zhang et al. 2002). However, some species are known to possess "non-canonical plastids" lacking peridinin, which are distinctive from "peridinin-containing plastids" in typical photosynthetic dinoflagellates in evolutionary origin. The plastids in members of genera *Karenia* and *Karlodinium* (family Kareniaceae) contain 19'-hexanoyl-fucoxanthin along with chlorophylls *a* and *c*, which are remnants of an endosymbiotic haptophyte (Bjørnland et al, 2003; Tengs et al, 2000; Zapata et al, 2012). Members of the genus *Lepidodinium* (family Gymnodiniaceae) established the current plastids containing chlorophylls *a* and *b* through the endosymbiosis with a green alga. A multi-gene phylogenetic analysis based on the plastid-encoded genes robustly indicated that the green-colored plastids in *Lepidodinium* were derived from a small group of core Chlorophyta, namely pedinophytes. (Watanabe et al, 1987, 1990; Matsumoto et al, 2010; Kamikawa et al, 2015). The complete or nearly complete pl-genome data are currently available for two dinoflagellates bearing non-canonical plastids. Gabrielsen et al. (2011) reported non-minicircled pl-genome of a kareniacean dinoflagellate, *Karlodinium* (*Kl.*) *veneficum*. The *Kl. veneficum* pl-genome is originated from that of endosymbiotic haptophyte and its gene repertory is likely smaller than that of the pl-genome of a free-living haptophyte, *Emilinia huxleyi*, although the former genome has not been completed. The complete pl-genome of *L. chlorophorum* was reconstructed as a single circular chromosome (Kamikawa et al. 2014). Again, the *L. chlorophorum* pl-genome appeared to be reduced gene repertory-wise in comparison to the pl-genome of a closely related but free-living relative, *Pedinomonas minor*.

In the host (dinoflagellate) cells, both haptophyte endosymbionts in kareniacean species and green algal endosymbionts in *Lepidodinium* were extensively reduced, leaving only plastids in the endosymbiont-derived compartments. These ultrastructural features imply that both haptophyte and green algal endosymbionts have been already integrated as the host-controlled organelles. The aforementioned notion is further supported by "haptophyte/green algal genes" found in the nuclear genomes of kareniacean species/*L.*

*chlorophorum* (Takishita et al, 2004; Nosenko et al, 2006; Patron et al, 2006; Minge et al, 2010; Burki et al, 2014). Such EGT is regarded as the hallmark of the host-endosymbiont interlock at the genetic level. Thus, both haptophyte-derived and pedinophyte-derived plastids in dinoflagellates are considered as genuine organelles in the current host cells.

Besides the two lineages bearing peridinin-lacking plastids described above, another group of dinoflagellates (e.g., *Durinskia baltica* and *Kryptoperidinium foliaceum*) is known to retain obligate diatom endosymbionts, rather than peridinin-containing plastids (Inagaki et al, 2000; Horiguchi 2006; Imanian et al, 2010). The obligate diatom endosymbionts retain their own nuclei, mitochondria, and plastids in the dinoflagellate cells (Tomas and Cox 1973; Eschbach et al, 1990). A recent study detected no clear evidence for EGT in the host genome, suggesting that the interlock between the host and endosymbiont has yet to be established at the genetic level (Hunsperger et al, 2016).

## 3. Purposes of this work

In the following chapters, I will address two evolutionary issues associated with plastid replacements in dinoflagellates, (1) the impact of the endosymbiosis in a dinoflagellate cell on the pl-genome in an endosymbiont algae, and (2) the evolutionary processes that shaped the proteomes of the plastids in *Ke. brevis*/*Kl. veneficum* and *L. chlorophorum*. Chapter 2 accounts the former issue, "What impact did plastid replacement have on the pl-genome of endosymbiotic algae?" In this chapter, I will report features of pl-genome in novel dinoflagellate strain TRD-132, of which plastid is originated from green algal endosymbiont (green-colored dinoflagellate). Then, to reveal in detail what kind of modification the pl-genome of green-colored dinoflagellates including TRD-132 undergo, I compared the characters of pl-genome of TRD-132 with other two green-colored dinoflagellate species and also with extant green algae. The next chapter 3 describes the latter question, "What impact did the EGT associated with plastid replacement have on the host nucleus genome?" On this subject, I surveyed the transcripts encoding nucleus-encoded proteins involved in three metabolic pathways localized in the plastid (i) C5 pathway for the heme biosynthesis, (ii) the Chl *a* biosynthetic pathway and (iii) the non-mevalonate pathway for the isopentenyl diphosphate (IPP) biosynthesis in two kareniacean species (*Karenia* and

*Karlodinium*) and *Lepidodinium*, which bear haptophyte-derived and green alga-derived plastids, respectively. Individual proteins identified in this study were then subjected to phylogenetic analyses to evaluate how the three pathways were modified before and after the haptophyte/green algal endosymbiosis. In the last chapter 4 I briefly summarize the arguments I made through the chapter 2 to 3, and I propose a future perspective of investigation of genome evolution in dinoflagellates experienced plastid replacement.

# Chapter 2: Not open to the public.

# Chapter 3: The patterns in reorganization of host genome in *Karenia, Karlodinium* and *Lepidodinium*

## 1. Introduction and questions addressed in this chapter

**Origins of nucleus-encoded proteins in *Karenia*, *Karlodinium* and *Lepidodinium***

In this chapter I will report on evolution of nuclear genome in host dinoflagellates associated with plastid replacement, focusing on *Karenia* (*Ke*.) *brevis*, *Karlodinium* (*Kl*.) *veneficum* bearing haptophyte-derived plastids, and *L. chlorophorum* bearing pedinophyte-derived plastid. As mentioned in the chapter 1, pioneering works on *Ke. brevis*, *Kl. veneficum*, and *L. chlorophorum* identified substantial numbers of endosymbiotically transferred genes in the host genomes (Ishida and Green 2002; Nosenko et al, 2006; Patron et al, 2006; Minge et al, 2010). In addition, plastid-related genes, which bear phylogenetic affinities to the orthologous genes in peridinin-containing dinoflagellates, have been detected. For instance, the *Ke. brevis* nuclear gene for 1-deoxy-D-xylulose 5-phosphate synthase (DXS), and the *L. chlorophorum* nuclear gene for 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR or IspC), showed clear affinities to the orthologous identified in peridinin-containing dinoflagellates (Minge et al, 2010; Bentlage et al, 2016). As the peridinin-containing plastid was most likely established in the ancestral dinoflagellates, it is reasonable to assume that the genes described above have been inherited vertically beyond plastid replacements, rather than acquired from the endosymbiont. Furthermore, a certain fraction of plastid-related genes were unlikely to be vertically inherited from the ancestral dinoflagellates or endosymbiotically acquired from eukaryotic algae that gave rise to non-canonical plastids. For instance, *sppA* for serine protease IV of green algal origin was detected in *Ke. brevis* (Nosenko et al. 2006), and *csp41* for an mRNA-binding protein in *L. chlorophorum* was found to share the origin with the stramenopile orthologues (Minge et al. 2010). According to the previous gene surveys in the *Ke. brevis*/*Kl. veneficum* and *L. chlorophorum* transcriptomic data, it is clear that the proteomes in the non-canonical plastids in these dinoflagellates comprise the proteins with diverse evolutionary origins. However, to my knowledge, there is few study examined whether the trend in the evolutionary

"chimerism" in plastid proteome vary among non-canonical plastids established separately in dinoflagellate evolution. In the third chapter, I investigate the evolutionary origins of enzymes involved in plastid-localized metabolic pathways, namely those synthesizing heme, chlorophyll *a* (Chl *a*), and isopentenyl diphosphate (IPP), in *Ke. brevis*/*Kl. veneficum* and *L. chlorophorum*. I here note that this study does not consider C4 pathway for the heme biosynthesis and the mevalonate pathway for IPP biosynthesis, as the former and latter occurred in both mitochondria and cytosol, and in the cytosol, respectively (Vavilin and Vermaas, 2002; Kuzuyama, 2002). By comparing the evolutionary origins of individual proteins/enzymes involved in the three pathways in *Ke. brevis*/*Kl. veneficum* and *L. chlorophorum*, I explore the putative evolutionary driving forces that shaped the plastid metabolic pathways in the haptophyte-derived plastids in the the former species and the pedinophyte-derived plastids in the latter species.

## 2. Materials and Methods

**RNA-seq analysis of *L. chlorophorum***

　　*L. chlorophorum* NIES-1868 has been maintained in the laboratory as described in Kamikawa et al. (2015a). Total RNA was isolated from the harvested cells with TRIzol reagent (Thermo Fisher Scientific) according to the manufacture's instruction. I sent the RNA sample to Hokkaido System Science (Sapporo, Hokkaido, Japan) for library construction with Truseq RNA Sample Prep Kit (Illumina) followed by sequencing on a Illumina HiSeq 2000 platform. I yielded ~386 million paired-end reads (~39 Mbp in total). After exclusion of poorly called reads and removal of an adaptor sequence by FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/download.html), I finally assembled the curated reads into 77841 contigs by Trinity ver.2.1.0 (Haas et al, 2013). The RNA-seq data of *L. chlorophorum* was deposited in DDBJ Sequence Read Archive, accession number DRA006544.

　　I evaluated the coverage of the contig data yielded from the *L. chlorophorum* RNA-seq data against the "universally conserved genes" in eukaryotes with BUSCO ver. 3 (Waterhouse et al, 2017). The proportions of "complete," "fragmented" and "missing" BUSCOs were 84.2, 3.3 and 12.5%, respectively.

**Survey of the genes encoding proteins involved in the heme, Chl *a* and IPP biosyntheses**

In this study, I conducted phylogenetic analyses on the proteins involved in "Porphyrin and chlorophyll metabolism" (map00860) and "Terpenoid backbone biosynthesis" (map00900) in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (http://www.genome.jp/kegg/pathway.html). KEGG Orthology (KO) identifiers (KOIDs) of the proteins subjected to the investigation in this study were K02492, K01845, K01698, K01749, K01719, K01599, K00228, K02495, K00231, K01772, K03403, K03404, K03405, K03428, K04034, K04035, K04040, K00218, K04037, K04038, K04039, K01662, K00099, K00991, K00919, K01770, K03526 and K03527 (see also Fig .7). To generate amino acid sequence alignments covering phylogenetically diverse eukaryotic algae and prokaryotes, I surveyed the sequences of interest in the contigs generated from the RNA-seq data of *L. chlorophorum* (see above) and public databases including the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP; http://marinemicroeukaryotes.org), which provides the contigs from the combined assemblies of the *Ke. brevis* and *Kl. veneficum* RNA-seq data ("Karenia-brevis-Wilson," "Karenia-brevis-CCMP2229" and "Karlodinium-micrum-CCMP2283"; https://www.imicrobe.us), by TBLASTN or BLASTP. In the first BLAST search, multiple sequences registered in the KOIDs listed above were used as queries except the enzymes described below. Two evolutionarily distinct versions were known for protoporphyrinogen IX oxidase (PPOX) ("*hemY*-type" and "*hemJ*-type"), and I used KOID K00231 as the query to identify the former type of PPOX. I surveyed *hemJ*-type PPOX sequences, which is not registered in the current KEGG pathway, by using UPF0093 membrane protein encoded by *slr1790* in *Synechocystis* sp. strain PCC 6803 (UniProtKB P72793) as the query. Both N-DVR and F-DVR are included in a single KOID (K19073), but the sequences deposited in the GenBank database were subjected to the initial survey as queries (Meguro et al, 2011; Nagata et al, 2005). The accession numbers for the queries used in the initial BLAST searches are summarized in Table S1 (supplemental information, deposited at an online repository https://drive.google.com/open?id=1IzWlpN3vlqz76IkCls_TCQ5XWyUp7y8g). The candidate sequences matched with the queries with *E*-values smaller than $10^{-20}$ in the first BLAST searches were retained. These candidate (amino acid) sequences were then subjected to BLASTP searches against the NCBI nr database (threshold was set as *E*-value of $10^{-5}$). Based on the results of the second BLAST searches, I retrieved the sequences matched to the

proteins involved in the heme, Chl *a* and IPP biosyntheses.

I further subjected the *Ke. brevis* and *Kl. veneficum* sequences selected by the procedure described above to BLASTP against the entire MMETSP database (https://www.imicrobe.us/#/projects/104) to examine potential cross-contamination among multiple RNA-seq data deposited in the database (Marron et al, 2016; Dorrell & Gile et al, 2017). The candidate sequences did not match with any sequence derived from organisms rather than *Ke. brevis* or *Kl. veneficum* (data not shown). Thus, I conclude that the *Ke. brevis* and *Kl. veneficum* sequences retrieved from the MMETSP database are highly unlikely contaminants from non-kareniacean organisms.

The nucleotide sequences (and their detailed information) identified in this study are summarized in Tables S2. I also deposited their conceptual amino acid sequence at an online repository (see above)

**Phylogenetic analyses**

In this study, I investigated the origins of individual proteins involved in the three pathways for the heme, Chl *a* and IPP biosyntheses in *Ke. brevis*, *Kl. veneficum* and *L. chlorophorum* by phylogenetic analyses with the maximum-likelihood (ML) method. For each protein of interest, I aligned the amino acid sequences retrieved from the public sequence databases described in the previous chapter by MAFFT v7.149b (Katoh & Standley, 2013) with the L-INS-i option. The accession numbers of the sequences included in the alignments were listed in Table S3. The resultant alignments were manually refined, followed by exclusion of ambiguously aligned positions (see Table 2 for the details). The final alignments were individually subjected to RAxML 8.0.20 (Stamatakis, 2014). The ML tree was selected from 10 heuristic tree searches, each of which was initiated from a randomized stepwise addition parsimony tree. The most appropriate amino acid substitution model was selected for each alignment by ProtTest 3.4 (Darriba et al, 2011). The substitution models applied to the ML phylogenetic analyses are summarized in Table 2. MLBPs were calculated by summarizing 100 trees, each of which was inferred from a bootstrap data by a single heuristic tree search (see above). Bayesian analyses were conducted with PhyloBayes 4.1c under the CAT-GTR model with a discrete Γ distribution with four categories (Lartillot & Philippe, 2004). Four independent Markov Chain Monte Carlo (MCMC) chains were run in parallel. The detailed settings of the PhyloBayes analyses are summarized in Table 2. The

individual alignments and Newick format treefiles inferred from both ML and Bayesian methods are available in the online repository (see above).

**_In silico_ examination of plastid-localizing potential of N-terminal extensions**

Nucleus-encoded proteins localized to complex plastids usually bear "bipartite" plastid-localizing signals, which are composed of an endoplasmic reticulum-targeting signal peptide (SP) and transit peptide-like sequence at their N-termini (Bolte et al, 2009). Firstly, I identified extra amino acid residues at their N-termini (N-terminal extensions) in the putative *Ke. brevis/Kl. veneficum* and *L. chlorophorum* proteins involved in the heme, Chl *a* and IPP biosyntheses by comparing the corresponding bacterial homologues (most of them are from cyanobacteria). I subsequently searched the SP in each N-terminal extension by SignalP v.4.1 (Petersen et al. 2011). In case of the SP being predicted, the transit peptide (TP) potential in the non-SP portions of the N-terminal extension was examined by ChloroP v.1.1 (Emanuelsson et al, 1999). SignalP and ChloroP analyses were done with the default settings. The results from the analyses on the N-terminal extensions can be found in Table S4.

## 3. Results

Henceforth here, I designated (1) the proteins inherited from the ancestral dinoflagellate as "vertically inherited-type" or "VI-type," (2) those acquired from an endosymbiont (i.e. haptophyte and green alga, in *Ke. brevis/Kl. veneficum* and *L. chlorophorum*, respectively) as "endosymbiotically acquired-type" or "EA-type," and (3) those acquired from organisms distantly related to the host or endosymbiont lineages as "laterally acquired-type" or "LA-type." A certain fraction of the proteins investigated could not be categorized into any of the three types described above due to the lack of phylogenetic signal in alignments (see below).

**Heme biosynthetic pathway**

All or most of the enzymes required for the heme biosynthesis were successfully identified in the transcriptomic data of *Ke. brevis*, *Kl. veneficum* and *L. chlorophorum*. In *Ke. brevis* and *Kl. veneficum*, 8 out of the 9 enzymes were identified—only uroporphyrinogen III

synthase (UROS) and coproporphyrinogen oxidase (CPOX) were missed in the former and latter, respectively. It is most likely that the UROS/CPOX transcripts were simply missed from the *Ke. brevis*/*Kl. veneficum* cDNA libraries, as UROS is essential to convert hydroxymethylbilane to uroporphyrinogen III, and CPOX is indispensable to yield protoporphyrinogen IX from coproporphyrinogen III. Fortunately, the *Ke. brevis* UROS sequence has already deposited in the GenBank database under the accession number CO063310, and was phylogenetically analyzed in this study (see below). Fifty-four out of the 60 transcripts (those encoding putative cytosolic proteins were excluded; see below) investigated here were found to bear N-terminal extensions. The SP were predicted in 21 out of the 54 N-terminal extensions. Seven N-terminal extensions were predicted to bear both SP and TP. The details of the N-terminal extensions are summarized in Table S4.

**Proteins with evolutionarily diverse origins comprise the *Ke. brevis* and *Kl. veneficum* pathways**

To convert glutamyl-tRNA to glutamate-1-semialdehyde, *Ke. brevis* and *Kl. veneficum* possesses two versions of glutamyl-tRNA reductase (GTR). "*Karenia*-1" and "*Karlodinium*-1" sequences were considered to be as VI-type, as they were nested within a clade with those of peridinin-containing dinoflagellates, and the clade as a whole received full statistical support in the GTR phylogeny (Fig. 8A). On the other hand, the second version of GTR in the two kareniacean species (*Karenia*-2 and *Karlodinium*-2) were most likely acquired from the haptophyte endosymbiont (i.e. EA-type). The two GTR sequences were nested within the haptophyte clade, and the haptophyte clade (including the *Karenia*-2 and *Karlodinium*-2 sequences) was supported by a MLBP of 66% and a bayesian posterior probability (BPP) of 0.99 (Fig. 8A).

Aminolevulinic acid (ALA) is synthesized from glutamate-1-semialdehyde by glutamate-1-semialdehyde 2,1-aminomutase (GSAT). A single version of GSAT was identified in both *Ke. brevis* and *Kl. veneficum*. The possibility of the two GSAT sequences being acquired from the haptophyte endosymbiont can be excluded, as the haptophyte sequences (except that of *Pavlova* sp.) formed a robust clade and excluded the kareniacean sequences in the GSAT phylogeny (Fig. 8B). Nevertheless, the phylogenetic origin of either *Kl. veneficum* or *Ke. brevis* GSAT sequence could not be pinpointed any further. The GSAT sequences of *Ke. brevis* and peridinin-containing dinoflagellates grouped together in the both

ML and Bayesian phylogenies, but their monophyly was poorly supported (MLBP of 8% and BPP <0.50). The *Kl. veneficum* GSAT sequence showed no specific affinity to any sequence examined here. Thus, concluding the precise origins of the GSAT sequences of *Ke. brevis* and *Kl. veneficum* was withheld in this study.

Synthesis of porphobilinogen from ALA is likely catalyzed by a single aminolevulinic acid dehydratase (ALAD) homologue in both *Ke. brevis* and *Kl. veneficum*. Both *Ke. brevis* and *Kl. veneficum* ALAD sequences formed a robustly supported clade with those of peridinin-containing dinoflagellates (MLBP of 99% and BPP of 1.0; Fig. 8C), suggesting that the two sequences are of VI-type.

Two versions of porphobilinogen deaminase (PBGD), which deaminates porphobilinogen to synthesize hydroxymethylbilane, were identified in *Ke. brevis* (*Karenia*-1 and 2). The PBGD phylogeny (Fig. 8D) recovered (i) a clade of the sequences of peridinin-containing dinoflagellates and *Karenia*-1 sequence with a MLBP of 100% and a BPP of 0.99, and (ii) a clade of the haptophyte and *Karenia*-2 sequences with a MLBP of 73% and a BPP of 0.83. Thus, *Ke. brevis* seemingly uses both VI-type and EA-type enzymes for hydroxymethylbilane synthesis. Four PBGD sequences in *Kl. veneficum* (*Karlodinium*-1-4), which clearly share a single ancestral sequence were identified. The *Kl. veneficum* clade appeared to be distant from the clade comprising the sequences of peridinin-containing dinoflagellates and the *Karenia*-1 sequence, suggesting that the *Karlodinium*-1-4 sequences are not of VI-type. The *Kl. veneficum* clade was connected to the haptophyte clade (including the *Karenia*-2 sequence) with a MLBP of 47% and a BPP of 0.67 (highlighted by an arrowhead in Fig. 8D). The statistical support for the particular node is insufficient to conclude or exclude the haptophyte origin of the *Karlodinium* sequences with confidence. Thus, the origin of the *Karlodinium* sequences still remain uncertain.

The evolutionary origins of two distinct versions of UROS in *Kl. veneficum* (*Karlodinium*-1 and 2), and a single UROS sequence detected in an expressed-sequence tag data from *Ke. brevis* (Fig. 9A) were assessed. The *Karenia* sequence, *Karlodinium*-1 and peridinin-containing dinoflagellate sequences grouped together with a MLBP of 76% and BPP of 0.95. Thus, the *Karenia* and *Karlodinium*-1 sequences are most likely of VI-type. The *Karlodinium*-2 appeared to be remote from the dinoflagellate or haptophyte clade, suggesting this sequence was of LA-type (Fig. 9A). The UROS phylogeny connected the *Karlodinium*-2 sequence and that of a red alga *Rhodosorus marinus* with a MLBP of 95% and BPP of 0.93,

and the "*Karlodinium + Rhodosorus*" clade was nested within the bacterial sequences. The particular clade could have been generated by two separate gene transfers from a single bacterium (or two closely related bacteria) to *Kl. veneficum* and *Rhodosorus*. Alternatively, the combination of the first gene transfer from a bacterium to either *Kl. veneficum* or *Rhodosorus*, and the second one between the two eukaryotes may have produced the "*Karlodinium + Rhodosorus*" clade in the UROS phylogeny. In either of the two scenarios, the *Karlodinium*-2 sequence can be regarded as LA-type. It is necessary to increase the number of red algal UROS sequences in the future studies to retrace the precise origins of the *Karlodinium* and *Rhodosorus* sequences.

Four acetyl side chains in uroporphyrinogen III are removed by uroporphyrinogen decarboxylase (UROD) to generate coproporphyrinogen III. Pioneering studies revealed that photosynthetic eukaryotes with complex plastids possess evolutionarily distinct, multiple versions of UROD (Kořený et al, 2011; Cihlář & Füssy et al, 2016). The UROD sequences of peridinin-containing dinoflagellates were split into three clades in the UROD phylogeny (designated as D1, D2 and D3 clades in Fig. 9B), suggesting that the three distinct versions have already been established in the ancestral dinoflagellate. Likewise, haptophytes were found to possess three distinct versions of UROD (designated as H1, H2 and H3 clades in Fig. 9B). Five and four UROD sequences in *Ke. brevis* and *Kl. veneficum* were here identified, respectively. Among the five sequences identified in *Ke. brevis*, the "*Karenia*-1, 2 and 4" sequences were considered as EA-type, as they were placed within the haptophyte sequences in the UROD phylogeny (Fig. 9B). H1 clade including the *Karenia*-1 and 2 sequences received a MLBP of 100% and a BPP 1.0. The *Karenia*-4 sequence and the haptophyte sequences (except that of *Pavlova*) formed H3 clade, of which monophyly was supported by a MLBP of 79% and a BPP of 0.99. The "*Karenia*-3" sequence grouped with the sequence of a euglenid *Eutreptiella gymnastica* with a MLBP of 100% and a BPP of 0.99. As the *Karenia*-3 sequence showed no affinity to the dinoflagellate or haptophyte clade in the UROD phylogeny, it was categorized in LA-type. The intimate phylogenetic affinity between the *Karenia*-3 and *E. gymnastica* sequences hints either (i) gene transfer between the two organisms or (ii) two separate gene transfers from an as-yet-unknown organism to *Ke. brevis* and *E. gymnastica*. It is also important to improve the sequence sampling from euglenids in future phylogenetic studies to understand precisely the biological event generated the union of the *Karenia*-3 and *E. gymnastica* sequences. The "*Karenia*-5" sequence is most likely

descended from one of the UROD versions established in the ancestral dinoflagellate (i.e. VI-type), as this sequence participated in D3 clade, of which monophyly received a MLBP of 100% and a BPP of 1.0. The origins of four *Kl. veneficum* sequences (*Karlodinium*-1-4) were also assessed based on the UROD phylogeny (Fig. 9B). The *Karlodinium*-1 and 2 sequences were nested within H1 clade, and the *Karlodinium*-4 sequence were placed within H3 clade. Thus, the three UROD sequences were conclude to be of EA-type. The *Karlodinium*-3, one of the *L. chlorophorum* sequences (see below), and diatom sequences formed a robust clade (MLBP of 100% and BPP of 1.0; highlighted by an arrowhead in Fig. 9B), suggesting that the *Karlodinium*-3 sequence were laterally acquired from a diatom (i.e. LA-type).

Coproporphyrinogen III is oxidized by CPOX to yield protoporphyrinogen IX. A single version of CPOX in *Ke. brevis* was identified, but none in *Kl. veneficum*. The CPOX phylogeny (Fig. 9C) recovered a clade comprising the sequences of peridinin-containing dinoflagellates and *Ke. brevis* with a MLBP of 98% and a BPP of 0.99. Thus, *Ke. brevis* uses a VI-type CPOX.

Protoporphyrinogen IX is further oxidized by protoporphyrinogen IX oxidase (PPOX) to obtain protoporphyrin IX. The PPOX sequences of peridinin-containing dinoflagellates were separated into two distinct clades labeled as "D1" and "D2" (Fig. 10A), and both received MLBPs of 100% and BPPs ≥0.97. The PPOX sequences in D2 clade are likely cytosolic version, as these dinoflagellate sequences, as well as those of other photosynthetic eukaryotes bearing complex plastids (chlorarachniophytes, *Euglena gracilis* and *Vitrella brassicaformis*), formed a robust clade with the PPOX sequences of heterotrophic eukaryotes (MLBP of 93% and BPP of 0.99; highlighted by an arrowhead in Fig. 10A). Two versions of PPOX in *Ke. brevis* (*Karenia*-1 and 2) were identified, while a single version was identified in *Kl. veneficum*. The *Karenia*-1 sequence fell into D1 clade (Fig. 10A), suggesting that this PPOX sequence is of VI-type. On the other hand, the PPOX phylogeny united the *Karenia*-2 and a single PPOX of *Kl. veneficum* with the sequences of stramenopiles, haptophytes and *L. chlorophorum* with a MLBP of 100% and a BPP of 0.99 (highlighted by a double-arrowhead in Fig. 10A). As the bipartitions within the clade were principally unresolved (Fig. 10A), the possibility of the *Karenia*-2 and *Karlodinium*-1 sequences group with the haptophyte sequences in this clade cannot be excluded. The *Karenia*-2 and *Karlodinium*-1 sequences are definitely not of VI-type, but the phylogenetic resolution was not sufficient to classify the two sequences into of EA-type or LA-type. Thus,

20

I decide to leave the origins of the *Karenia*-2 and *Karlodinium*-1 sequences uncertain.

In the last step in the heme biosynthesis, ferrochelatase (FeCH) converts protoporphyrin IX to protoheme. None of the FeCH sequences identified in *Ke. brevis* and *Kl. veneficum* appeared to be of VI-type. *Ke. brevis* possesses three distinct versions of FeCH (*Karenia*-1-3). The FeCH phylogeny (Fig. 10B) united the *Karenia*-1 sequence with one of four versions of FeCH in *L. chlorophorum* (*Lepidodinium*-3) with a MLBP of 100% and a BPP of 1.0, and this union was then connected specifically to γ-proteobacterial sequences with a MLBP of 75% and a BPP of 0.99 (highlighted by an arrowhead in Fig. 10B). This subtree can be explained by two sequential gene transfer events, namely the first gene transfer from a γ-proteobacterium to either *L. chlorophorum* or *Ke. brevis*, and the second one between the two dinoflagellates. Consequently, the *Karenia*-1 and *Lepidodinium*-3 sequences can be of LA-type, and traced back to the bacterial sequence. The *Karenia*-2 sequence is unlikely to be of VI-type, as the sequences of peridinin-containing dinoflagellates (and one of the sequences of *L. chlorophorum*) were united with a MLBP of 100% and a BPP of 1.0 (Fig. 10B). The *Karenia*-2 sequence, which showed no clear affinity to the haptophyte sequences, is unlikely to be of EA-type (Fig. 10B). Thus, the *Karenia*-2 sequence was concluded to be of LA-type, although the precise donor remains uncertain. Finally, the FeCH phylogeny grouped the *Karenia*-3 sequence and a single FeCH sequence identified in *Kl. veneficum* with the cyanobacterial sequences, and their monophyly was supported by a MLBP of 76% and a BPP of 0.94 (highlighted by a double-arrowhead in Fig. 10B). That the two dinoflagellate sequences were concluded to be of cyanobacterial origin (i.e. LA-type).

Pioneering studies on kareniacean dinoflagellates revealed the evolutionary chimeric natures of the plastid proteomes of *Ke. brevis* and *Kl. veneficum*, which comprises VI-, EA- and LA-type proteins (Ishida & Green 2002; Nosenko et al., 2006; Patron et al., 2006; Burki et al., 2014). This study further enables me to evaluate the precise contribution of EA-type proteins on the evolution of the heme biosynthetic pathway in kareniacean dinoflagellates—EA-type proteins found only in three out of the 9 steps, namely (i) GTR in *Ke. brevis* and *Kl. veneficum*, (ii) PBGD in *Kl. veneficum*, and (iii) UROD in *Ke. brevis* and *Kl. veneficum*.

**Little impact of endosymbiotic gene transfer on the *L. chlorophorum* pathway**

Three versions of GTR in *L. chlorophorum* (*Lepidodinium*-1-3) were identified. The

*Lepidodinium*-1 sequence grouped with the sequence of peridinin-containing dinoflagellates (and the *Karenia*-1 and *Karlodinium*-1 sequences), and this "dinoflagellate" clade received a MLBP of 100% and a BPP of 1.0 (Fig. 8A). Thus, the *Lepidodinium*-1 sequence was concluded to be of VI-type. The *Lepidodinium*-2 and 3 sequences were tied to each other with a MLBP of 100% and a BPP of 1.0, and this clade was connected with the "dinoflagellate" clade described above (Fig. 8A). However, the phylogenetic affinity between the two clades received little statistical support (MLBP of 38% and BPP <0.50; highlighted by an arrowhead in Fig. 8A), suggesting that the *Lepidodinium*-2 and 3 sequences were unlikely of VI-type. The GTR sequences of land plants and green algae (plus a euglenid *Eutreptiella gymnastica*) formed a clade supported by a MLBP of 100% and a BPP of 0.99, and appeared to be distant from the two sequences of *L. chlorophorum* (Fig. 8A). Thus, the green algal (endosymbiont) origin of the *Lepidodinium*-2 and 3 sequences can be excluded. Combined, the two sequences are regarded as LA-type, although the organism donated a GTR gene to *L. chlorophorum* remains unclear.

Two out of the three versions of GSAT identified in *L. chlorophorum* could not be classified. In the GSAT phylogeny (Fig. 8B), the "*Lepidodinium*-1" and "*Lepidodinium*-2" sequences fell separately into the cluster of the sequences of peridinin-containing dinoflagellates and *Ke. brevis*, but this clade as a whole received no significant statistical support. Thus, I left the origins of the *Lepidodinium*-1 and 2 sequences uncertain in this study. On the other hand, the "*Lepidodinium*-3" sequence was excluded from the clade comprising the sequences of diverse photosynthetic eukaryotes and cyanobacteria with a MLBP of 100% and a BPP of 1.0, and connected to the sequences of *Streptomyces coelicolor*, *Mycobacterium tuberculosis* and *Corynebacterium diphtheriae* with a MLBP of 71% and a BPP of 0.91 (highlighted by an arrowhead in Fig. 8B). This tree topology suggests that the *Lepidodinium*-3 sequence was acquired from a bacterium (i.e. LA-type), albeit the bacterium donated a GSAT gene to *L. chlorophorum* cannot be pinpointed.

Two versions of ALAD were identified in *L. chlorophorum* (*Lepidodinium*-1 and 2). The sequences of peridinin-containing dinoflagellates, the *Lepidodinium*-1 sequence and the sequences of *Ke. brevis* and *Kl. veneficum* clustered with a MLBP of 99% and a BPP of 1.0 in the ALAD phylogeny (Fig. 8C). Thus, I conclude the *Lepidodinium*-1 sequence as VI-type. The *Lepidodinium*-2 sequence was distantly related to the sequences of peridinin-containing dinoflagellate or green algae/land plants, but showed no strong affinity to any clade/sequence

in the ALAD phylogeny (Fig. 8C). Thus, I propose the *Lepidodinium*-2 sequence as LA-type, albeit its precise origin was unresolved in the ALAD phylogeny.

Two versions of PBGD and a single version of UROS identified from *L. chlorophorum*. The PBGD phylogeny (Fig. 8D) united the two sequences of *L. chlorophorum* together with a MLBP of 100% and a BPP of 0.99, but this clade showed little phylogenetic affinity to the sequences of peridinin-containing dinoflagellates, green algae/land plants or any sequences considered in the phylogenetic analysis. Likewise, the UROS phylogeny (Fig. 9A) recovered no clear affinity of the sequence of *L. chlorophorum* to other sequences including those of peridinin-containing dinoflagellates or green algae/land plants. Thus, *L. chlorophorum* likely uses LA-type PBGD and UROS, but their precise origins remain uncertain.

I identified 7 versions of UROD in *L. chlorophorum*, and 6 of them showed clear affinities to the sequences of peridinin-containing dinoflagellates. In the UROD phylogeny (Fig. 9B), the sequences of peridinin-containing dinoflagellates formed three distinct clades (D1-3 clades), and each of these clades enclosed at least one sequence of *L. chlorophorum*, namely (i) the "*Lepidodinium*-2" sequence in D1 clade, (ii) "*Lepidodinium*-3 and 4" sequences in D2 clade, and (iii) "*Lepidodinium*-5, 6 and 7" sequences in D3 clade. Thus, the 6 sequences described above are concluded as VI-type. The "*Lepidodinium*-1" sequence and sequences of diatoms (and *Kl. veneficum*) formed a clade with a MLBP of 100% and a BPP of 1.0 (highlighted by an arrowhead in Fig. 9B), suggesting that this version was acquired from a diatom (i.e. LA-type).

Three versions of CPOX were identified in *L. chlorophorum*, but none of them was of VI-type or EA-type. The CPOX phylogeny (Fig. 9C) placed the "*Lepidodinium*-1" sequence in a remote position from the sequences of peridinin-containing dinoflagellates or green algae/land plants. Instead, the *Lepidodinium*-1 sequence grouped with the bacterial sequences, as well as the eukaryotic sequences for the cytosolic pathway, with a MLBP of 77% and a BPP of 0.87 (highlighted by an arrowhead in Fig. 9C). This protein most likely bears no N-terminal extension (Table S4). Altogether, I propose that the *Lepidodinium*-1 sequence encodes a cytosolic CPOX enzyme involved in C4 pathway, and omitted from the discussion below. The "*Lepidodinium*-2" and "*Lepidodinium*-3" sequences share high sequence similarity in the mature protein region, while their N-terminal regions are distinct from each other (Table S4). The two *Lepidodinium* sequences were united robustly with the

diatom sequences in the CPOX phylogeny (MLBP of 86% and BPP of 0.98; highlighted by a double-arrowhead in Fig. 9C). Thus, the ancestral CPOX of the two *L. chlorophorum* sequences was most likely acquired from a diatom (i.e. LA-type).

I conclude that *L. chlorophorum* possesses two distinct VI-type (*Lepidodinium*-1 and 3) and a single LA-type PPOX (*Lepidodinium*-2). The PPOX sequences of peridinin-containing dinoflagellates were split into two distinct clades (D1 and D2), and the two clades received strong statistical support from both ML bootstrap and Bayesian analyses (Fig. 10A). The *Lepidodinium*-1 and 3 sequences were included in D1 and D2 clades, respectively. As the PPOX sequences (including the *Lepidodinium*-3 sequence) in D2 clade can be considered as the cytosolic version, I did not discuss the *Lepidodinium*-3 sequence further. The PPOX phylogeny recovered a robust clade comprising the *Lepidodinium*-2 sequence and, the sequences of haptophytes, stramenopiles, *Kl. veneficum* and *Ke. brevis* (MLBP of 100% and BPP of 0.99; highlighted by a double-arrowhead in Fig. 10A). Thus, the *Lepidodinium*-2 sequence was acquired from an organism distantly related to dinoflagellates or green algae/land plants (i.e. LA-type).

In *L. chlorophorum*, I identified four versions of FeCH. The "*Lepidodinium*-1" sequence is of VI-type, as this sequence apparently shared the origin with the sequences of peridinin-containing dinoflagellates, and their monophyly was supported with a MLBP of 100% and a BPP of 1.0 (Fig. 10B). On the other hand, I consider the rest of the sequences of *L. chlorophorum* as LA-type, as the "*Lepidodinium*-2," "*Lepidodinium*-3" and "*Lepidodinium*-4" sequences appeared to be distantly related to the dinoflagellate clade described above or the green algae/land plant sequences in the FeCH phylogeny (Fig. 10B). The precise positions of the *Lepidodinium*-2 and *Lepidodinium*-4 sequences were unresolved, and it remains unclear how *L. chlorophorum* acquired the two versions of FeCH. On the other hand, the FeCH phylogeny united the "*Lepidodinium*-3" and *Karenia*-1 sequences together (MLBP of 100% and BPP of 1.0), and this dinoflagellate clade was then connected to two γ-proteobacterial sequences with a MLBP of 75% and a BPP of 0.99 (highlighted by an arrowhead in Fig. 10B). I have already proposed the two scenarios for the origin of the *Lepidodinium*-3 and *Karenia*-1 sequences, in which two lateral gene transfers were invoked (see the previous section for the details).

The most prominent feature in the *L. chlorophorum* pathway is that no gene from the green algal endosymbiont was detected in the heme biosynthesis. Instead, genes transferred

from organisms related to neither host (dinoflagellate) nor endosymbiont (green alga) largely contributed to the *L. chlorophorum* pathway. The impact of lateral gene transfer on the heme biosynthesis is the most prominent in the steps catalyzed by PBGD, UROS and CPOX, in which only LA-type proteins were identified.

**Chl *a* biosynthetic pathway**

I surveyed the transcripts encoding enzymes involved in the Chl *a* biosynthesis in *Ke. brevis*, *Kl. veneficum* and *L. chlorophorum*, and assessed their origins individually. Overall, all the enzymes required to synthesize Chl *a*, except Mg-protoporphyrin IX monomethyl ester (MgPME) cyclase, was retrieved from the transcriptomic data from the three dinoflagellates. No sign for nucleus-encoded MgPME cyclase, which converts MgPME to divinyl protochlorophyllide, has been detected in complex algae (Wilhelm et al. 2006; Nymark et al. 2009). Although not described in detail, I additionally surveyed the single-subunit MgPME cyclase encoded by *chlE*, which are phylogenetically distinct from the multi-subunit MgPME cyclase, but yielded no significant match. I suspect an as-yet-unidentified enzyme forming E-ring in the photosynthetic eukaryotes described above. Twenty-two out of the 27 transcripts investigated here were found to possess N-terminal extensions, which likely work as plastid-localizing signals (Note that 11 sequences were proceeded by the putative SPs, and two of their non-SP portion were predicted to have the TP potential; see Table S4 for the details).

**Large impact of EGT on the *Ke. brevis* and *Kl. veneficum* pathways**

Mg chelatase (MgCH), which comprises three subunits ChlD, ChlH and ChlI, inserts $Mg^{2+}$ to protoporphyrin IX. In *Kl. veneficum*, ChlI is plastid-encoded (Gabrielsen et al. 2011) and the rest of the subunits were nucleus-encoded (see below). Although no pl-genome data is available for *Ke. brevis*, *chlI* is predicted to reside in the pl-genome of *Karenia mikimotoi* based on a pioneering study on the plastid transcriptome of this species (Dorrell et al, 2016). In this study, I identified the partial *chlI* contig in the transcriptomic data of *Ke. brevis* (contig No. 0173787962). The 3′ terminus of this contig was not completed, I could not conclude whether the *chlI* transcript received the poly(U) tail, which is the unique RNA modification occurred in peridinin-containing plastids as well as the non-canonical plastids of *Ke. mikimoti*

and *Kl. veneficum* (Dorrell & Howe, 2012; Richardson et al, 2014). Although the precise genome harboring *chlI* in *Ke. brevis* remains uncertain, I assume that the *Ke. brevis chlI* sequence was transcribed from the pl-genome, based on the close relationship between *Ke. mikimotoi* and *Ke. brevis*. As the current study focuses on nucleus-encoded proteins involved in plastid metabolisms, I stopped examining the origin and evolution of ChlI in the two kareniacean species (and *L. chlorophorum*; see below) any further.

I here examine the evolutionary origins of two nucleus-encoded subunits of MgCH, ChlH and ChlD, in *Ke. brevis* and *Kl. veneficum*. A single version of ChlD was identified in *Ke. brevis* and *Kl. veneficum*. The ChlD phylogeny (Fig. 11A) recovered a clade of the sequences of haptophytes, *Ke. brevis* and *Kl. veneficum* with full statistical support, suggesting that the kareniacean ChlD sequences are of EA-type. Both *Ke. brevis* and *Kl. veneficum* possess two distinct versions of ChlH (*Karenia*-1 and 2, and *Karlodinium*-1 and 2). ChlH sequences can be split into two distinct clades, "ChlH-1" and "ChlH-2," as a previous study reported (Lohr et al, 2005, Fig. 11B). ChlH-1 sequences are ubiquitously distributed in photosynthetic organisms, while ChlH-2 sequences have been found in restricted lineages. The sequences of green algae/land plants formed two distinct clades (Gp1 and 2 clades). In the ChlH phylogeny (Fig. 11B), the haptophyte ChlH-1 sequences, the *Karenia*-1 and *Karlodinium*-1 sequences formed a clade supported with a MLBP of 98% and a BPP of 1.0. In contrast, the *Karenia*-2 and *Karlodinium*-2 sequences were nested within the Gp2 clade containing ChlH-2 sequences of green algae and a euglenid, and their monophyly received a MLBP of 81% and a BPP of 0.98 (highlighted by an arrowhead in Fig. 11B). Thus, I conclude that *Ke. brevis* and *Kl. veneficum* possess ChlH-1 sequences acquired from the haptophyte endosymbiont (i.e. EA-type), while their ChlH-2 sequences are of green algal origin (i.e. LA-type).

*S*-adenosylmethionine:Mg-protoporphyrin *O*-methyltransferase (MgPMT) converts Mg-protoporphyrin IX to Mg protoporphyrin IX monomethyl ester. The MgPMT sequences of *Ke. brevis* and *Kl. veneficum* grouped with the haptophyte sequences (except the one of *Pavlova*), and their monophyly was supported by a MLBP of 67% and a BPP of 0.99 (highlighted by an arrowhead in Fig. 11C). The two kareniacean species most likely use the MgPMT acquired from the haptophyte endosymbiont (i.e. EA-type).

No MgPME cyclase has been identified in any dinoflagellates regardless of plastid-type, and I could not examine the origin and evolution of this enzyme (see above).

However, divinyl protochlorophyllide reductase (DVR), which recognizes the product of MgPME cyclase (divinyl protochlorophyllide) as the substrate and generate protochlorophyllide, were identified in both dinoflagellates bearing peridinin and those bearing non-canonical plastids. I identified both N-DVR and F-DVR sequences in *Ke. brevis*, while only N-DVR sequence was found in *Kl. veneficum*. The F-DVR phylogeny (Fig. 11D) recovered the clade of the sequences of *Ke. brevis* and haptophytes with robust statistical support, suggesting that *Ke. brevis* acquired the F-DVR gene from the haptophyte endosymbiont (i.e. EA-type).

The N-DVR phylogeny (Fig. 11E) united the sequences of *Ke. brevis* and *Kl. veneficum* together with a MLBP of 98% and a BPP of 0.99, and the kareniacean clade showed any phylogenetic affinity to neither haptophyte sequences nor other dinoflagellate sequences. Instead, the kareniacean clade grouped with the sequences of stramenopiles, haptophytes, and two chromerids (*Chromera* and *Vitrella*) supported with a MLBP of 100% and a BPP of 1.0 (highlighted by an arrowhead in Fig. 11E). In this large clade, the affinity between the kareniacean clade and haptophyte sequences was not positively supported. I here propose that multiple stramenopile lineages donated N-DVR genes separately to haptophytes, kareniacean species, and chromerids (i.e. LA-type).

In land plants, conversion of protochlorophyllide to chlorophyllide *a* is catalyzed by the light-dependent and/or light-independent forms of protochlorophyllide oxidoreductase (POR). The light-dependent POR is nucleus-encoded, while the light-independent form comprises three plastid-encoded subunits (ChlB, ChlL and ChlN). No gene for light-independent POR was found in the sequenced region of the *Kl. veneficum* pl-genome (Gabrielsen et al. 2011) or plastid transcriptomic data of *Ke. mikimotoi* (Dorrell et al. 2016), implying that kareniacean species lack the light-independent version. I identified two and three distinct versions of the light-dependent POR in *Ke. brevis* and *Kl. veneficum*, respectively (*Karenia*-1 and 2, and *Karlodinium*-1-3), as demonstrated in Hunsperger et al, (2015). In the POR phylogeny (Fig. 12A), the *Karenia*-1 and *Karlodinium*-1 sequences grouped with the sequences of stramenopiles, cryptophytes and haptophytes, and their monophyly was supported by a MLBP of 95% and a BPP of 0.98 (highlighted by an arrowhead in Fig. 12A). The *Karenia*-1 and *Karlodinium*-1 sequences cannot be of VI-type, as the two sequences are distantly related to other dinoflagellate sequences included in the alignment. However, it is difficult to classify the *Karenia*-1 and *Karlodinium*-1 sequences

into EA-type or LA-type, as the relationship between the two dinoflagellate sequences and the haptophyte sequences was unresolved in the particular clade. Thus, I leave the origins of the two sequences uncertain in this study. The *Karenia*-2, *Karlodinium*-2 and *Karlodinium*-3 sequences robustly grouped together within the haptophyte sequences, and this "haptophyte" clade received a MLBP of 98% and a BPP of 0.99 (Fig. 12A). Thus, I conclude that these sequences were acquired from the haptophyte endosymbiont (i.e. EA-type).

The final step of the Chl *a* biosynthesis is catalyzed by Chl synthase (CS). A single version of CS was identified in *Ke. brevis* and *Kl. veneficum*. The two kareniacean sequences were placed within the haptophyte clade in the CS phylogeny, and the "haptophyte" clade as a whole was supported by a MLBP of 98% and a BPP of 1.0 (Fig. 12B). Thus, the sequences of *Ke. brevis* and *Kl. veneficum* are considered as EA-type.

The phylogenetic analyses described above revealed that EA-type proteins operate in all the steps converting protoporphyrin IX to Chl *a* in *Kl. veneficum* and/or *Ke. brevis* (except the step catalyzed by MgPME cyclase; see above). In addition, the common ancestor of *Ke. brevis* and *Kl. veneficum* should have possessed LA-type ChlH-2, POR and N-DVR, which were acquired from phylogenetically diverse eukaryotes distantly related to dinoflagellates or haptophytes.


**Genetic influx from phylogenetically diverse organisms shaped the *L. chlorophorum* pathway**

As discussed in the previous section, the Chl *a* biosynthetic pathway in *Ke. brevis* and *Kl. veneficum* appeared to be shaped by the genes transferred from the endosymbiont (i.e. a haptophyte in the above systems). Curiously, this is not the case for the same pathway in *L. chlorophorum*, of which plastid was derived from a pedinophyte green alga. Note that I present no result from the *L. chlorophorum* ChlI, which turned out to be plastid-encoded (Kamikawa et al., 2015a). I identified 8 proteins involved in the Chl *a* biosynthetic pathway in *L. chlorophorum*, and assess their phylogenetic origins individually (Figs. 11-12,). Among the 8 proteins examined here, I conclude MgPMT as a sole EA-type protein among those involved in the *L. chlorophorum* pathway. The MgPMT phylogeny (Fig. 11C) placed the *L. chlorophorum* sequence within a radiation of the sequences of green algae, land plants and chlorarachniophytes, and their monophyly was supported by a MLBP of 88% and a BPP of 0.98.

Our surveys and phylogenetic analyses revealed that VI-type proteins were almost entirely eliminated from the *L. chlorophorum* pathway. I identified only one of the two versions of POR (*Lepidodinium*-1) as VI-type. The POR phylogeny (Fig. 12A) recovered two distinct clades of the sequences of peridinin-containing dinoflagellates (D1 and D2 clades), and placed the *Lepidodinium*-1 sequence within D1 clade. D1 clade containing the *Lepidodinium*-1 sequence as a whole received a MLBP of 97% and a BPP of 0.99.

I could not clarify the origin of the N-DVR sequence of *L. chlorophorum*, of which position was resolved in neither ML nor Bayesian phylogenetic analyses (Fig. 11E). The *Lepidodinium* sequence was excluded from the robust clade of the sequences of peridinin-containing dinoflagellates, suggesting that this sequence cannot be of VI-type. However, it is difficult to pursue the origin of the *Lepidodinium* sequence any further, as the N-DVR phylogeny failed to exclude a potential affinity between the sequence of interest and the green algal/land plant sequences (Fig. 11E).

I classified ChlD, ChlH, one of the two versions of POR (*Lepidodinium*-2), F-DVR and CS into LA-type. In the ChlD phylogeny, the sequence of *L. chlorophorum* appeared to be excluded from the sequences of peridinin-containing dinoflagellates and those of the green algal/land plant sequences (Fig. 11A), suggesting that this sequence cannot be of VI-type or EA-type. Instead, the sequence of *L. chlorophorum*, as well as those of chromerids, cryptophytes and *Eutreptiella*, were placed within the radiation of the stramenopile sequences, and their monophyly was supported by a MLBP of 96% and a BPP of 0.99. This tree topology prompts us to propose that *L. chlorophorum* ChlD was acquired from a stramenopile (i.e. LA-type).

To my surprise, the ChlH, POR and CS phylogenies placed the sequences of *L. chlorophorum* within the radiation of the chlorarachniophyte sequences, and their monophylies received MLBPs >92% and by BPPs >0.62 (Figs. 11B, 12A and 12B). I here propose multiple gene transfers from chlorarachniophytes to the ancestral *Lepidodinium* cell to interpret the aforementioned tree topologies. The putative chlorarachniophyte origins of the three genes are not contradict to a recent molecular clock analysis, in which the green algal endosymbiosis leading to the *Lepidodinium* plastids was predicted to occur more recently than that leading to the chlorarachniophyte plastids (Jackson et al. 2018).

The F-DVR phylogeny reconstructed a robust affinity between the sequence of *L. chlorophorum* and the chlorarachniophyte clade within the land plant/green algal sequences

(Fig. 11D). This tree topology requires a combination of an EGT and LGT, as *Lepidodinium* species and chlorarachniophytes acquired their plastids commonly from green algae. Either ancestral *Lepidodinium* cell or the ancestral chlorarachniophyte acquired an F-DVR gene from the green algal endosymbiont (i.e. EGT), followed by the second gene transfer between the two organisms (i.e. LGT). Thus, I have to leave the evolutionary origin of the *L. chlorophorum* F-DVR uncertain.

The phylogenetic analyses described above revealed that EGT was much less significant in the *L. chlorophorum* pathway than the kareniacean pathway (see above). Instead, chlorarachniophytes seemingly donated the genes encoding the proteins involved in four out of the five steps in the *L. chlorophorum* pathway.

**Non-mevalonate pathway for the IPP biosynthesis**

The origin of all the enzymes involved in the non-mevalonate pathway of *Ke. brevis* and *Kl. veneficum* were investigated carefully in Bentlage et al. (2016). On the other hand, the entire picture of the *L. chlorophorum* pathway remain to be completed, leaving 5 out the 7 enzymes involved in this pathway unidentified (Minge et al. 2010). I successfully identified all the enzymes involved in the non-mevalonate pathway in *L. chlorophorum* (see below). In this section, I mainly examined the origins of individual enzymes involved in the non-mevalonate pathway of *L. chlorophorum*, coupled with a brief overview of the same pathway of the two kareniacean species. Thirty-one out of the 32 transcripts investigated here were found to possess N-terminal extensions, which likely work as plastid-localizing signals (Note that 12 sequences were proceeded by the putative SPs, and three of the non-SP portions were predicted to have the TP potential; see Table S4 for the details).

For the step synthesizing 1-deoxy-D-xylulose-5-phosphate (DXP) from pyruvate and glyceraldehyde 3-phosphate, *L. chlorophorum* was found to possess two versions of DXP synthase (DXS) (*Lepidodinium*-1 and 2). The DXS phylogeny robustly grouped the *Lepidodinium*-1 and 2 sequences with those of peridinin-containing dinoflagellates and *Ke. brevis* (Fig. 13A). The sequence of *Kl. veneficum* was found to be remote from the dinoflagellate clade, and showed an affinity to the haptophyte sequences (Fig. 13A). The clade comprising the sequences of *Kl. veneficum* and haptophytes received a MLBP of 63% and a BPP of 0.82 (if the *Pavlova* sequence was excluded, the "*Karlodinium* + haptophyte"

30

clade was supported by a MLBP of 98% and a BPP of 1.0). Thus, I conclude that the DXS sequences of *L. chlorophorum* and *Ke. brevis* were of VI-type, while that of *Kl. veneficum* was of EA-type.

The conversion of DXP to 2-C-methyl-D-erythritol 4-phosphate (MEP) is catalyzed by DXP reductase (DXR). Minge et al. (2010) detected a partial sequence of a VI-type DXR (GenBank accession number CCC15090). From my transcriptome data, two versions of DXR were identified in *L. chlorophorum* (*Lepidodinium*-1 and 2; the former corresponds to the previously reported DXR sequence). The DXR phylogeny (Fig. 13B) reconstructed a robust monophyly of the *Lepidodinium*-1 and 2 sequences, the sequences of peridinin-containing dinoflagellates, and those of *Ke. brevis* and *Kl. veneficum* (MLBP of 100% and BPP of 1.0), suggesting that *L. chlorophorum* and the two kareniacean species use VI-type proteins for this reaction.

*L. chlorophorum* was found to possess both LA-type and VI-type versions of MEP cytidylyltransferase (IspD) (*Lepdidodinium*-1 and 2) to convert MEP into 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol (CDP-ME). The *Lepidodinium*-1 sequence was concluded as LA-type based on its remote position from the sequences of peridinin-containing dinoflagellates or green plants in the IspD phylogeny (Fig. 13C). Instead, the *Lepidodinium*-1 sequence showed a specific affinity to the sequence of a stramenopile *Ochromonas* sp. with a MLBP of 100% and a BPP of 1.0. As the monophyly of the stramenopile sequences was not recovered probably due to insufficient phylogenetic signal in the IspD alignment, the two scenarios remain possible. The simplest scenario assumes that the *Lepidodinium*-1 sequence was originated to a stramenopile gene. Alternatively, a scenario, in which invokes separate gene transfers from an as-yet-unknown organism to *L. chlorophorum* and *Ochromonas* sp., is also possible. To examine the two scenarios in the future, I need to prepare a new IspD alignment including the sequences from stramenopiles closely related to *Ochromonas* sp. (i.e. chrysophycean algae) and reexamine a new alignment in the future. In contrast, the *Lepidodinium*-2 sequence, together with a single sequence of *Kl. veneficum*, were considered as VI-type, as they formed a clade with those of peridinin-containing dinoflagellates (MLBP of 81% and BPP of 0.68). The IspD sequence of *Ke. brevis* appeared to be nested within the clade of green algae/land plants and chlorarachniophytes, and being distantly related to the dinoflagellate clade (including the sequences of *L. chlorophorum* and *Kl. veneficum*) or the haptophyte clade (Fig. 13C). The

position of *Ke. brevis* IspD is consistent with the green algal origin of this enzyme proposed by Bentlage et al. (2016).

Minge et al. (2010) reported a VI-type 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE) (GenBank accession number CCC15094), which phosphorylates CDP-ME to 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol (CDP-MEP), in *L. chlorophorum*. In this study, I detected two distinct versions of IspE (*Lepidodinium*-1 and 2)—the *Lepidodinium*-1 sequence corresponds to the version reported in Minge et al. (2010) and the *Lepidodinium*-2 sequence is a novel version of IspE. The IspE phylogeny (Fig. 13D) recovered a monophyly of the two sequences of *L. chlorophorum*, the sequences of peridinin-containing dinoflagellates, the sequence of *Kl. veneficum*, and one of the two sequences of *Ke. brevis* (*Karenia*-2), which was supported by a MLBP of 88% and a BPP of 0.99. Thus, *L. chlorophorum*, *Ke. brevis* and *Kl. veneficum* possess VI-type versions of IspE. In addition, *Ke. brevis* possesses an EA-type IspE (*Karenia*-1), which was united with the haptophyte sequences with a MLBP of 100% and a BPP of 0.98.

The ancestral dinoflagellate likely possessed two versions of 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF), which converts CDP-MEP to 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (MEcPP). In the IspF phylogeny (Fig. 13E), the vast majority of the dinoflagellates sequences was split into two clades (D1 and D2), of which monophylies were supported by MLBPs of 85-94% and BPPs of 0.99-1.0, respectively. D1 clade appeared to contain one of the two sequences of *L. chlorophorum* (*Lepidodinium*-1), as well as the sequence of *Kl. veneficum* and one of the two sequences of *Ke. brevis* (*Karenia*-2). The other version of *L. chlorophorum* (*Lepidodinium*-2) was nested within D2 clade. Thus, I conclude that *L. chlorophorum*, *Ke. brevis* and *Kl. veneficum* possess VI-type versions of IspF. As reported in Bentlage et al. (2016), *Ke. brevis* possesses an additional IspF sequence (*Karenia*-1) with a phylogenetic affinity to the haptophyte sequences, suggesting that this version is of EA-type.

The origin and evolution of 1-hydroxy-2-methyl-2-butenyl 4-diphosphate (HMB-PP) synthase (IspG), which synthesize HMB-PP from MEcPP, seems straightforward in dinoflagellates. Two versions of IspG in *L. chlorophorum* (*Lepidodinium*-1 and 2) identified in this study were phylogenetically analyzed together with the sequences of *Ke. brevis* and *Kl. veneficum*. In the IspG phylogeny (Fig. 13F), the aforementioned dinoflagellate sequences tightly clustered with the sequences of peridinin-containing dinoflagellates (MLBP of 98%

and BPP of 1.0). Thus, I concluded that *Ke. brevis*, *Kl. veneficum* and *L. chlorophorum* uses VI-type enzymes to synthesize MEcPP.

The last step of the non-mevalonate pathway is catalyzed by HMB-PP reductase (IspH) to yield IPP from HMG-PP. Two versions of IspH in *L. chlorophorum* (*Lepidodinium*-1 and 2) were identified. The IspH phylogeny (Fig. 13G) recovered a clade comprising the sequences of peridinin-containing dinoflagellates, apicomplexan parasites and chromerids, as well as the two versions of *L. chlorophorum* and one of the three versions of *Ke. brevis* (*Karenia*-3), with a MLBP of 68% and a BPP of 0.92. The rapidly evolving natures of the apicomplexan and chromerid sequences were noticed, which highly likely biased the phylogenetic inference from the IspH alignment. Consequently, the same alignment was reanalyzed after the exclusion of the long-branch sequences. In both ML and Bayesian analyses of the second IspH alignment, the *Lepidodinium*-1 and 2, and *Karenia*-3 sequences grouped with those of peridinin-containing dinoflagellates with a MLBP of 100% and a BPP of 0.99 (Fig. 13G). The two versions of IspH in *L. chlorophorum* are concluded to be of VI-type. *Ke. brevis* was found to possess two additional versions of IspH (*Karenia*-1 and *Karenia*-2). The IspH phylogeny (Fig. 13G) united the *Karenia*-1 sequence and a single IspH sequence of *Kl. veneficum* with the sequences of haptophytes with a MLBP of 90% and a BPP of 0.98. The *Karenia*-2 sequence was excluded from the dinoflagellate or haptophyte clade, and instead connected to the sequence of a stramenopile *Ochromonas* sp. with a MLBP of 88% and a BPP of 0.99 (Fig. 13G). As the sequence of *Ochromonas* sp. showed no clear affinity to other stramenopile sequences, the origin of the *Karenia*-2 sequence remains unclear. Thus, as discussed in Bentlage et al. (2016), *Ke. brevis* uses VI-type (*Karenia*-3), EA-type (*Karenia*-1) and LA-type (*Karenia*-2) enzymes to yield IPP, while *Kl. veneficum* possesses a single, EA-type version.

As Bentlage et al. (2016) demonstrated, the non-mevalonate pathways in *Ke. brevis* and *Kl. veneficum* are evolutionary hybrids of VI-type and EA-type enzymes. In sharp contrast, the same pathway in *L. chlorophorum* appeared to be dominated by VI-type proteins, except a single LA-type protein (one of the two IspD versions).

# 4. Discussion

**Perspectives toward the evolution of kareniacean dinoflagellates and their plastids**

As anticipated from the haptophyte origin of the kareniacean plastids, the *Ke. brevis* and *Kl. veneficum* pathways investigated here appeared to be composed of three evolutionary types of proteins, namely EA-, VI- and LA-types (Patron et al. 2006; Nosenko et al. 2006; Hunsperger et al, 2015; Bentlage et al. 2016). Nevertheless, the impact of the genetic influx from the haptophyte endosymbiont was different among the three pathways in *Ke. brevis*/*Kl. veneficum* (Fig. 14). In the two kareniacean species, EA-type proteins, together with a few LA-type proteins, found to dominate the Chl *a* biosynthesis, albeit no VI-type protein was detected. In sharp contrast, VI-type proteins persist in 5-6 out of the 7 steps required for the non-mevalonate pathway for the IPP biosynthesis, albeit the contributions of EA-type and LA-type proteins may not be negligible. The evolutionary chimerism is most advanced in heme biosynthesis, in which all the three protein types are identified (Fig. 14).

Based on the difference in degree of evolutionary chimerism among the three pathways discussed in the previous section, I here explore the early evolution of kareniacean species. Saldarriaga et al. (2001) hinted non-photosynthetic intermediate stages for the dinoflagellates bearing non-canonical plastids based on the dinoflagellate phylogeny inferred from small subunit ribosomal DNA sequences, albeit they made no mention of the presence/absence of a residual plastid. Patron et al. (2006) and Nosenko et al. (2006) hypothesized that the ancestral kareniacean species possessed a non-photosynthetic plastid prior to the haptophyte endosymbiosis, although both studies assessed a restricted number of plastid-localized proteins. The above proposal is plausible, as secondarily non-photosynthetic eukaryotes often possess plastids with no photosynthetic activity but diverse metabolic capacities including those to synthesize heme and/or IPP (Lim & McFadden, 2010; Lohr et al, 2012; Kamikawa et al, 2015b, 2015c, 2017; Janouškovec et al, 2017). Notably, Kareniaceae includes a kleptoplastic species found in the Ross Sea, Antarctica, in addition to the species bearing haptophyte-derived plastids (Gast et al, 2006, 2007). The coexistence of the species bearing the haptophyte-derived non-canonical plastids and that leading a kleptoplastic lifestyle in the same family lends an additional support for the non-photosynthetic nature of their common ancestor. The hypothesis for the

non-photosynthetic nature in the ancestral kareniacean species can explain well the elimination of VI-type proteins from the Chl *a* biosynthesis in both *Ke. brevis* and *Kl. veneficum* (Fig. 14). During the putative non-photosynthetic period in the early kareniacean evolution, the proteins involved in the Chl *a* biosynthesis may have been dispensable, leading to discard of the corresponding genes from the dinoflagellate genome. Interestingly, Patron et al. (2006) identified *atpC* transcript encoding a subunit of VI-type ATP synthase in *Kl. veneficum*, implying that ATP synthase on plastid thylakoid membrane has not been discarded during tertiary endosymbiosis. Nevertheless, the presence of VI-type ATP synthase in the current *Kl. veneficum* plastid is not necessary to negate the putative non-photosynthetic period for the early kareniacean evolution, as the plastid ATP synthase can persist in secondarily non-photosynthetic organisms (Donaher et al, 2009; Wicke et al, 2013; Kamikawa et al, 2015c; 2017; the equivalent observation is also reported in a nonphotosynthetic cyanobacterium by Nakayama et al, 2014). In the later kareniacean evolution, the entire pathway for the Chl *a* biosynthesis was most likely reconstructed in the haptophyte-derived plastid by incorporating exogenous genes (acquired mainly from the endosymbiont). In contrast, both *Ke. brevis* and *Kl. veneficum* seemingly use VI-type proteins to synthesize both heme and IPP, suggesting that the proteins originally worked in the peridinin-containing plastid persisted in the ancestral kareniacean species beyond the haptophyte endosymbiosis. The two pathways have been modified after the haptophyte endosymbiosis by incorporating exogenous genes acquired from phylogenetically diverse organisms (including the endosymbiont), as I observed both EA- and LA-type proteins in the current pathways in both *Ke. brevis* and *Kl. veneficum* (Fig. 14).

I propose that the putative non-photosynthetic plastid in the ancestral kareniacean species possessed the residual genome as argued below. Poly-uridylation of the 3′ termini of plastid transcripts was found primarily in peridinin-containing dinoflagellates, but also *Ke. brevis* and *Kl. veneficum* (Dorrell & Howe, 2012; Richardson et al, 2014; Dorrell et al, 2016). Thus, the machinery for the RNA modification have been inherited from the photosynthetic ancestor bearing a peridinin-containing plastid to the extent kareniacean species beyond the putative non-photosynthetic period. In line with the above speculation, the putative non-photosynthetic plastid in the ancestral kareniacean species most likely retained a transcriptionally active genome.

I can retrieve an additional insight into the early kareniacean evolution by comparing

the phylogenetic inventories of VI-, EA- and LA-type proteins in the heme, Chl *a* and IPP biosynthetic pathways between *Ke. brevis* and *Kl. veneficum*. For instance, the contribution of VI-type proteins to the heme biosynthesis seemingly differs between *Ke. brevis* and *Kl. veneficum*, which retain 6 and 3 VI-type proteins for the 9 steps in the heme biosynthesis, respectively (Fig. 14). Coincidently, the significance of LA-type proteins in the particular pathway seems to be expanded in the *Kl. veneficum* pathway comparing to the *Ke. brevis* pathway. These observations suggest that the reconstruction of metabolic pathways in the haptophyte-derived plastids (i.e. gene acquisitions/losses) was not completed before the separation of the genera *Karenia* and *Karlodinium*. However, I need to assess carefully whether the differences between the *Ke. brevis* and *Kl. veneficum* pathways observed in my comparisons stemmed from the incomplete coverages of gene repertories in the two kareniacean species in future studies. Alternatively, the differences between the *Ke. brevis* and *Kl. veneficum* pathways observed in my comparisons might reflect the separate haptophyte endosymbioses in *Karenia* and *Karlodinium*, as a plastid small subunit ribosomal DNA phylogeny placed the two dinoflagellate species in two remote positions within the haptophyte clade (Gast et al. 2007). The precise origin (or origins) of haptophyte-derived plastids needs to examine by pl-genome-based multigene alignments including broad members of Kareniaceae (e.g., *Takayama* spp.) in the future.

**Perspectives toward the evolution of *Lepidodinium* and its plastids**

The phylogenetic inventories of VI-, EA- and LA-type proteins appeared to be different among the heme, Chl *a* and IPP biosynthetic pathways in *L. chlorophorum* (Fig. 14). The IPP synthesis in this species retains VI-type proteins in all of the 7 steps, and no EA-type protein was found. In sharp contrast, the contribution of LA-type proteins to the Chl *a* biosynthesis is likely much greater than that of VI- or EA-type proteins. The heme biosynthesis appeared to be distinct from the two pathways described above, as I detected both VI- and LA-type proteins but no EA-type protein. Interestingly, the trend, of which VI-type proteins contribute to the heme and IPP biosyntheses at much greater magnitudes than the Chl *a* biosynthesis, is common between *L. chlorophorum* and *Ke. brevis*/*Kl. veneficum* (Fig. 14). Thus, as discussed the putative ancestral state of kareniacean species (see above), I speculate that the ancestral *Lepidodinium*, which engulfed a green algal endosymbiont, experienced a non-photosynthetic period and discarded most of the genes

encoding proteins involved in the Chl *a* biosynthesis, but retained a non-photosynthetic plastid with the capacities for synthesizing both heme and IPP. Ferredoxin-NADP$^+$ reductase (FNR) involved in photosynthetic electron transport chain has been identified in photosynthetic, as well as secondarily non-photosynthetic organisms (Balconi et al, 2009; Nakayama et al, 2014). Thus, the transcript encoding plastid-type FNR identified in the *L. chlorophorum* transcriptome (Minge et al. 2010) does not contradict the putative non-photosynthetic period in the ancestral *Lepidodinium* species.

I unexpectedly revealed a potentially large contribution of chlorarachniophyte genes to the Chl *a* biosynthesis in *L. chlorophorum* (Fig. 14). In the organismal tree of eukaryotes, chlorarachniophytes and dinoflagellates belong to two distantly related taxonomic assemblages, Rhizaria and Alveolata, respectively. Likewise, the current plastids in chlorarachniophytes and *Lepidodinium* were derived from distinct green algal groups, ulvophytes and pedinophytes, respectively (Suzuki et al, 2016; Kamikawa et al. 2015a). Thus, the relationship between their host lineages or that between their endosymbiont lineages (plastids) can provide no ground for the presence of chlorarachniophyte genes in the *L. chlorophorum* genome. If the ancestral *Lepidodinium* cell fed on chlorarachniophytes in the natural environment, such predator-prey relationship led to the genetic influx from the prey (chlorarachniophyte) genome to the predator (*Lepidodinium*) genome. Nevertheless, under the circumstance postulated above, genes transferred from chlorarachniophytes could not have been restricted to a single metabolic pathway. To understand the biological reasons for the genetic contribution from chlorarachniophytes to the Chl *a* biosynthesis in *L. chlorophorum*, I need to explore (i) potential interaction between dinoflagellates and chlorarachniophytes in the natural environment and (ii) the biochemical and/or physiological commonality in the proteins involved in the Chl *a* biosynthesis between chlorarachniophytes and *L. chlorophorum*. Nevertheless, my proposal for the presence of the "chlorarachniophyte genes" in the extant *L. chlorophorum* genome, which incorporates lateral gene transfers, needs to be re-assessed in the future. In particular, much broader sampling of ChlH, POR, CS and F-DVR sequences would be critical to infer the precise origins of the four "LA-type" proteins involved in the *L. chlorophorum* Chl *a* biosynthesis. Depending on the future ChlH, POR, CS and F-DVR phylogenies with improved sequence sampling, I may need to revise their phylogenetic origins proposed in the current study.

I also noticed a clear difference in contribution of EA-type proteins to the three

pathways between *L. chlorophorum* and *Ke. brevis*/*Kl. veneficum* (Fig. 14). EA-type proteins are most likely indispensable for the heme, Chl *a* and IPP biosyntheses in *Ke. brevis*/*Kl. veneficum*. On the other hand, only MgPMT in the Chl *a* biosynthetic pathway appeared to be of green algal origin in *L. chlorophorum*. One potential factor, which could introduce the marked difference between the *L. chlorophorum* and *Ke. brevis*/*Kl. veneficum* pathways, is the difference in plastid-targeting signal of nucleus-encoded plastid-related proteins between their endosymbionts. In both eukaryotes bearing primary plastids (e.g., green algae) and those bearing complex plastids (e.g., haptophytes and dinoflagellates), the vast majority of plastid-related proteins are nucleus-encoded, which are synthesized in the cytosol and localized to the plastid. In green algae, nucleus-encoded plastid-related proteins are synthesized with N-terminal extensions (i.e. transit peptides or TP), which act as the "tags" to pass through the two membranes surrounding their plastids (Patron & Waller, 2007). On the other hand, the "tag" sequences, which enable nucleus-encoded proteins to localize in complex plastids surrounded by three or four membranes, are more complicated than green algal plastids (Bolte et al, 2009). Patron et al. (2005) revealed that dinoflagellates with peridinin-containing plastids and haptophytes appeared to share a bipartite structure of plastid-targeting signal, which is composed of signal peptide and the TP-like region. Consequently, without any substantial modification on plastid-targeting signals, the ancestral kareniacean species could have targeted the proteins encoded by endosymbiotically transferred genes back to the haptophyte-derived plastid. In contrast, nucleus-encoded plastid-related proteins in the green algal endosymbiont engulfed by the ancestral *Lepidodinium* unlikely possessed bipartite plastid-targeting signals. Thus, in the ancestral *Lepidodinium*, the proteins encoded by endosymbiotically transferred genes needed to acquire bipartite plastid-targeting signals to be localized in the green alga-derived plastid surrounded by four membranes. Altogether, I here propose the initial presence/absence of bipartite plastid-targeting signals was one of the major factors affecting the EGT in dinoflagellates bearing non-canonical plastids. As anticipated from the above scenario, many LA-type proteins acquired from diverse eukaryotes bearing complex plastids (e.g., stramenopiles and chlorarachniophytes) were identified in the heme, Chl *a* and IPP biosynthetic pathways in *L. chlorophorum*. Nevertheless, the factor discussed above may not be dominant enough to exclude EA-type proteins from a green alga (e.g., MgPMT; Fig. 11C) and LA-type proteins from bacteria (e.g., GSAT; Fig. 8B) from the plastid proteome in *L.*

*chlorophorum*. In the future, I need to expand the detailed phylogenetic analyses into the entire *L. chlorophorum* plastid proteome to examine the hypothesis based on the biosynthetic pathways for heme, Chl *a* and IPP.

## 5. Conclusion

I here assessed the evolutionary origins of the proteins involved in the three plastid-localized pathways for the heme, Chl *a* and IPP biosyntheses in two separate dinoflagellate lineages bearing non-canonical plastids, namely kareniacean dinoflagellates bearing haptophyte-derived plastids (i.e. *Ke. brevis* and *Kl. veneficum*), and *L. chlorophorum* established a green alga-derived plastid. In each of the two dinoflagellate lineages, the three pathways have been modified differently during the process reducing an algal endosymbiont to a non-canonical plastid. I interpreted that the observed difference stemmed from the nature of the ancestral dinoflagellate engulfed a haptophyte/green algal endosymbiont. When individual pathways were compared between *Ke. brevis*/*Kl. veneficum* and *L. chlorophorum*, EGT appeared to contribute to the pathways in the former lineage much more substantially than those in the latter lineage. I proposed that this observation emerged partially from the structural difference in plastid-localizing signal (i.e. presence or absence of the SP) between the proteins acquired from the haptophyte endosymbiont and those from a green algal endosymbiont. The discussion based on the *Ke. brevis* and *Kl. veneficum* sequence data need to be reexamined in future studies incorporating the data from additional kareniacean species (e.g., members belonging to the genus *Takayama*), as well as their relative operating kleptoplastidy (Gast et al, 2006, 2007).

## Chapter 4: Summary and feature study

I briefly summarize my investigations and propose perspectives in future study. The initial interest of this work is establishments of complex plastids through serial endosymbiosis, a great driving force of diversification of photosynthetic eukaryotes. My aim is to resolve the questions how endosymbiotic algae were genetically integrated into the host cell, and contrariwise what kind of genetic reorganization to maintain complex plastids the host cell underwent. In this thesis I focused on dinoflagellates that experienced plastid replacements as they are likely to remain clues to the questions, and I reported genome evolution in dinoflagellates harboring non-canonical plastids from the both aspects, pl-genome and host nucleus genome. I described parallel reduction of pl-genomes of endosymbiotic pedinophytes in TRD-132, MRD-151 and *L. chlorophorum* in chapter 2. The loss of IRs, the loss of seven common genes and the expansion of gene overlapping/fusion suggest that a similar reductive pressure acted on pl-genome of endosymbiotic pedinophytes in the different dinoflagellates. These features also may give me a hint for pl-genome reduction occurred in when peridinin-plastids were established through endosymbiosis with red alga. Although IRs seem to be originally absent in red algae, it is imaginable that repertoire of genes that were lost from red algal pl-genome in the early stage of the process may be similar to those in green-colored dinoflagellates.

In the chapter 3, systematic investigation was conducted to reveal the pattern in chimerism in the proteins involved in biosynthetic pathways of heme, Chl *a* and IPP in *Karenia*, *Karlodinium* and *Lepidodinium*, by using phylotranscriptomic approach. Intriguingly the patterns of chimerism observed in the three pathways were common among the three dinoflagellates: VI-type protein was almost absent from Chl *a* biosynthetic pathway while exogenous protein (i.e. EA-type or LA-type) occupied this pathway. In contrast, VI-type proteins were conserved in biosynthetic pathways for heme and IPP and impact of EGT was less than Chl *a* synthetic pathway regardless of origin of plastids. The contrastive pattern of distribution in the VI-type proteins involved in the pathways is supporting evidence for presence of non-photosynthetic plastid that has its origin in peridinin plastid on the branch leading to *Ke. brevis*, *Kl. veneficum* or *L. chlorophorum*.

In the future study, it is necessary to expand taxon sampling in dinoflagellates harboring non-canonical plastid and also in other key organisms e.g. pedinophytes, haptophytes and chlorarachniophyte that are considered to be deeply involved in the evolution of the non-canonical plastids. For instance, pl-genome information is limited in dinoflagellates harboring haptophyte-derived plastids except for *Kl. veneficum* of which pl-genome is sequenced and *Ke. mikimotoi* of which pl-genome composition was predicted by trascriptomic data. Other species with haptophyte-derived plastids such as *Takayama* and *Ke. brevis* as well as kleptoplastidic dinoflagellate isolated from Ross sea should be surveyed. For further understanding of nuclear genome reorganization, it is important to investigate other metabolic pathways localized in plastids namely biosynthetic pathway for fatty acids, iron-sulfur cluster and shikimate pathway. Investigation of origin of proteins comprise these pathways will allow me to shed light on intermediate state of plastid replacement.

## Supplementary data

The nucleotide sequences and their conceptual amino acid sequence used in the analyses in chapter 3 and their detailed information are available online as supplementary data, Tables S1-S4 (see https://drive.google.com/open?id=1IzWlpN3vlqz76IkCls_TCQ5XWyUp7y8g).

# Acknowledgements

# References

Archibald, J. M., 2009. The puzzle of plastid evolution. *Current Biology* 19:R81-R88. DOI: 10.1016/j.cub.2008.11.067.

Balconi, E., Pennati, A., Crobu, D., Pandini, V., Cerutti, R., Zanetti, G. and Aliverti, A. 2009. The ferredoxin-NADP+ reductase/ferredoxin electron transfer system of *Plasmodium falciparum*. *Federation of European Biochemical Societies Journal* 276:3825-3836. DOI: 10.1111/j.1742-4658.2009.07100.x.

Bhattacharya, D., Yoon, H. S. and Hackett, J. D. 2004. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *Bioessays* 26:50-60. DOI: 10.1002/bies.10376.

Beale, S.I. 1999. Enzymes of chlorophyll biosynthesis. *Photosynthesis Research* 60:43-73. DOI: https://doi.org/10.1023/A:1006297731456.

Bentlage, B., Rogers, T. S., Bachvaroff, T. R. and Delwiche, C. F. 2016. Complex ancestries of isoprenoid synthesis in dinoflagellates. *Journal of Eukaryotic Microbiology* 63:123-137. DOI:10.1111/jeu.12261.

Bjørnland, T., Haxo, F. T. and Liaaen-Jensen, S. 2003. Carotenoids of the Florida red tide dinoflagellate *Karenia brevis*. *Biochemical Systematics and Ecology* 31:1147-1162. DOI: 10.1016/S0305-1978(03)00044-9.

Bolte, K., Bullmann, L., Hempel, F., Bozarth, A., Zauner, S. and Maier, U. G. 2009. Protein targeting into secondary plastids. *Journal of Eukaryotic Microbiology* 56:9-15. DOI: 10.1111/j.1550-7408.2008.00370.x.

Burki, F., Imanian, B., Hehenberger, E., Hirakawa, Y., Maruyama, S. and Keeling, P. J. 2014. Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. *Eukaryotic Cell* 13:246-255. DOI: 10.1128/EC.00299-13.

Chen, G. E., Canniffe, D. P. and Hunter, C. N. 2017. Three classes of oxygen-dependent cyclase involved in chlorophyll and bacteriochlorophyll biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* 114:6280-6285. DOI: 10.1073/pnas.1701687114.

Cihlář, J., Füssy, Z., Horák, A. and Oborník, M. 2016. Evolution of the tetrapyrrole biosynthetic pathway in secondary algae: conservation, redundancy and replacement. *PLoS ONE* 11 e0166338. DOI: 10.1371/journal.pone.0166338.

Darriba, D., Taboada, G.L., Doallo, R. and Posada, D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27:1164-1165. DOI: 10.1093/bioinformatics/btr088.

Donaher, N., Tanifuji, G., Onodera, N. T., Malfatti, S. A., Chain, P. S. G., Hara, Y. and Archibald, J. M. 2009. The complete plastid genome sequence of the secondarily nonphotosynthetic alga *Cryptomonas paramecium*: reduction, compaction, and accelerated evolutionary rate. *Genome Biology and Evolution* 1:439–448. DOI: 10.1093/gbe/evp047.

Dorrell, R. G., Gile, G., McCallum, G., Méheust, R., Bapteste, E. P., Klinger, C. M., Brillet-Guéguen, L., Freeman, K. D., Richter, D. J. and Bowler, C. 2017. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *eLife* 6:e23717. DOI: 10.7554/eLife.23717.

Dorrell, R. G., Hinksman, G. A. and Howe, C. J. 2016. Diversity of transcripts and transcript processing forms in plastids of the dinoflagellate alga *Karenia mikimotoi. Plant Molecular Biology* 90:233-247. DOI: 10.1007/s11103-015-0408-9.

Dorrell, R. G. and Howe, C. J. 2012. Functional remodeling of RNA processing in replacement chloroplasts by pathways retained from their predecessors. *Proceedings of the National Academy of Sciences of the United States of America* 109:18879-18884. DOI: 10.1073/pnas.1212270109.

Douglas, S. E and Penny, S. L. 1999. The plastid genome of the cryptophyte alga, Guillardia theta: complete sequence and conserved synteny groups confirm its common ancestry with red algae. *Journal of Molecular Evolution* 48:236-244. DOI: 10.1007/PL00006462.

Dubey, V. S., Bhalla, R. and Luthra, R. 2003. An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants. *Journal of Biosciences* 28:637-646.

Emanuelsson, O., Nielsen, H., von Heijne, G. 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science: a Publication of the Protein Society* 8:978-984. DOI: 10.1110/ps.8.5.978.

Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A.,, Raven, J. A., Schofield, O. and Taylor, F. J. 2004. The evolution of modern eukaryotic phytoplankton. *Science* 305:354-360 DOI: 10.1126/science.1095964.

Fujita, Y. and Bauer, C. E. 2003. The light-independent protochlorophyllide reductase: a nitrogenase-like enzyme catalyzing a key reaction for greening in the dark. In: Kadish K. M., Smith K. M., Guilard, R. eds. *The Porphyrin Handbook: Chlorophylls and bilins : biosynthesis, synthesis, and degradation.* USA: Elsevier Science, 109-156. DOI: 10.1016/B978-0-08-092387-1.50010-2.

Gabrielsen, T. M., Minge, M. A., Espelund, M., Tooming-Klunderud, A., Patil, V., Nederbragt, A. J., Otis, C., Turmel, M., Shalchian-Tabrizi, K., Lemieux, C. and Jakobsen, K. S. 2011. Genome evolution of a tertiary dinoflagellate plastid. *PLoS ONE* 6:e19132. DOI: 10.1371/journal.pone.0019132.

Gast, R. J., Moran, D. M., Beaudoin, D. J., Dennett, M. R. and Caron, D. A. 2006. Abundance of a novel dinoflalgelate phylotype in the Ross Sea, Antarctica. *Journal of Phycology*, 42:233-242. DOI: 10.1111/j.1529.8817.2006.00183.x.

Gast, R. J., Moran, D. M., Dennett, M. R. and Caron, D. A. 2007. Kleptoplasty in an Antarctic dinoflagellate: caught in evolutionary transition? *Environmental Microbiology*, 9:39-45. DOI: 10.1111/j.1462-2920.2006.01109.x.

Gilson, P. R. and McFadden, G. I. 1996. The miniaturized nuclear genome of a eukaryotic endosymbiont contains genes that overlap, genes that are cotranscribed, and the smallest known spliceosomal introns. *Proceedings of the National Academy of Sciences of the United States of America* 93:7737-7742 DOI:10.1073/pnas.93.15.7737.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., Leduc, R. D., Friedman, N. and Regev, A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8:1494-1512. DOI: 10.1038/nprot.2013.084.

Hallick, R. B., Hong, L., Drager, R. G., Favreau, M. R., Monfort, A., Orsat, B., Spielmann, A. and Stutz, E. 1993. Complete sequence of Euglena gracilis chloroplast DNA. *Nucleic Acids Research* 21:3537-3544.

Hibberd, D. J. and Norris, R. E. 1984. Cytology and ultrastructure of Chlorarachnion reptans (Chlorarachniophyta divisio nova, Chlorarachniophyceae classis nova). *Journal of Phycology* 20:310-330. DOI: 10.1111/j.0022-3646.1984.00310.x.

Hoek, C., Mann, D. and Jahns, H. M. 1995. Algae: an introduction to phycology. Cambridge: Cambridge University Press.

Howe, C. J., Barbrook, A. C., Nisbet, R. E., Lockhart, P. J. and Larkum, A. W. 2008. The origin of plastids. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences.* 363:2675-2685. DOI: 10.1098/rstb.2008.0050.

Hunsperger, H. M., Randhawa, T. and Cattolico, R. A. 2015. Extensive horizontal gene transfer, duplication, and loss of chlorophyll synthesis genes in the algae. *BMC Evolutionary Biology* 15:16. DOI: 10.1186/s12862-015-0286-4.

Hunter, W. N. 2007. The non-mevalonate pathway of isoprenoid precursor biosynthesis. *The Journal of Biological Chemistry* 282:21573-21577. DOI: 10.1074/jbc.R700005200.

Ishida, K., Cao, Y., Hasegawa, M., Okada, N. and Hara, Y. 1997. The origin of chlorarachniophyte plastids, as inferred from phylogenetic comparisons of amino acid sequences of EF-Tu. *Journal of Molecular Evolution* 45:682-687. DOI: 10.1007/PL00006272.

Ishida, K. and Green, B. R. 2002. 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. *Proceedings of the National Academy of Sciences of the United States of America* 99:9294-9299. DOI: 10.1073/pnas.142091799.

Ito, H., Yokono, M., Tanaka, R. and Tanaka, A. 2008. Identification of a novel vinyl reductase gene essential for the biosynthesis of monovinyl chlorophyll in *Synechocystis* sp. PCC6803. *The Journal of Biological Chemistry* 283:9002-9011. DOI: 10.1074/jbc.M708369200.

Ito, H. and Tanaka A. 2014. Evolution of a new chlorophyll metabolic pathway driven by the dynamic changes in enzyme promiscuous activity. *Plant and Cell Physiology* 55:593-603. DOI: 10.1093/pcp/pct203.

Jackson, C., Knoll, A. H., Chan, C. X., Verbruggen, H. 2018. Plastid phylogenomics with broad taxon sampling further elucidates the distinct evolutionary origins and timing of secondary green plastids. *Scientific Reports* 8:1523. DOI: 10.1038/s41598-017-18805-w.

Janouškovec, J., Horák, A., Oborník, M., Lukeš, J. and Keeling, P. J. 2010. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proceedings of the National Academy of Sciences of the United States of America* 107:10949-10954. DOI: 10.1073/pnas.1003335107.

Janouškovec, J., Gavelis, G. S., Burki, F., Dinh, D., Bachvaroff, T. R., Gornik, S. G., Bright, K. J., Imanian, B., Strom, S. L., Delwiche, C. F., Waller, R. F., Fensome, R. A., Leander, B. S., Rohwer, F. L. and Saldarriaga, J. F. 2017. Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proceedings of the National Academy of Sciences of the United States of America* 114:E171-E180. DOI: 10.1073/pnas.1614842114.

Jeffrey, S. W., Sielicki, M. and Haxo, F. T. 1975. Chloroplast pigment patterns in dinoflagellates. *Journal of Phycology* 11: 374-384. DOI: 10.1111/j.1529-8817.1975.tb02799.x.

Kamikawa, R., Tanifuji, G., Kawachi, M., Miyashita, H., Hashimoto, T. and Inagaki, Y. 2015a. Plastid genome-based phylogeny pinpointed the origin of the green-colored plastid in the dinoflagellate *Lepidodinium chlorophorum*. *Genome Biology and Evolution* 7:1133-1140. DOI: 10.1093/gbe/evv060.

Kamikawa, R., Yubuki, N., Yoshida, M., Taira, M., Nakamura, N., Ishida, K., Leander, B. S., Miyashita, H., Hashimoto, T., Mayama, S. and Inagaki, Y. 2015b. Multiple losses of photosynthesis in *Nitzschia* (Bacillariophyceae). *Phycological Research* 63:19-28. DOI: 10.1111/pre.12072.

Kamikawa R., Tanifuji G., Ishikawa S. A., Onodera N. T., Ishida K., Hashimoto T., Miyashita H., Mayama S. and Inagaki Y. 2015c. Proposal of a twin-arginine translocator system-mediated constraint against loss of ATP synthase genes from nonphotosynthetic plastid genomes. *Molecular Biology and Evolution* 32:2598-2604. DOI: 10.1093/molbev/msv134.

Kamikawa R., Moog D., Zauner S., Tanifuji G., Ishida K., Miyashita H., Mayama S., Hashimoto T., Maier U. G., Archibald J. A. and Inagaki Y. 2017. A non-photosynthetic diatom reveals early steps of reductive evolution in plastids. *Molecular Biology and Evolution* 34:2355-2366. DOI: 10.1093/molbev/msx172.

Katoh, K. and Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772-780. DOI: 10.1093/molbev/mst010.

Kobayashi, K., Masuda, T., Tajima, N., Wada, H. and Sato, N. 2014. Molecular phylogeny and intricate evolutionary history of the three isofunctional enzymes involved in the oxidation of protoporphyrinogen IX. *Genome Biology and Evolution* 6:2141-2155. DOI: 10.1093/gbe/evu170.

Kořený, L., Sobotka, R., Janouškovec, J., Keeling, P. J. and Oborník, M. 2011. Tetrapyrrole synthesis of photosynthetic chromerids is likely homologous to the unusual pathway of apicomplexan parasites. *The Plant Cell* 23:3454-3462. DOI: 10.1105/tpc.111.089102.

Lartillot, N. and Philippe, H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution* 21:1095-1109. DOI: 10.1093/molbev/msh112.

Lichtenthaler, H. K., Schwender, J., Disch, A. and Rohmer, M. 1997. Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway. *Federation of European Biochemical Societies Letters* 400:271-274. DOI: 10.1016/S0014-5793(96)01404-4.

Lim, L. and McFadden, G. I. 2010. The evolution, metabolism and functions of the apicoplast. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 365:749–763. DOI: 10.1098/rstb.2009.0273.

Lohr, M., Im, C. S. and Grossman, A. R. 2005. Genome-based examination of chlorophyll and carotenoid biosynthesis in *Chlamydomonas reinhardtii*. *Plant Physiology* 138:490-515. DOI: 10.1104/pp.104.056069.

Lohr, M., Schwender, J. and Polle, J. E. 2012. Isoprenoid biosynthesis in eukaryotic phototrophs: a spotlight on algae. *Plant Science* 185-186:9-22. DOI: 10.1016/j.plantsci.2011.07.018.

Lowe, T. M. and Eddy, S. R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25:955-964.

McFadden, G. I., Gilson, P. R., Hofmann, C. J., Adcock, G. J. and Maier, U. G. 1994. Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga. *Proceedings of the National Academy of Sciences of the United States of America* 91:3690-3694. DOI: 10.1073/pnas.91.9.3690.

Marron, A. O., Ratcliffe, S., Wheeler, G. L., Goldstein, R. E., King. N., Not, F., de Vargas, C. and Richter, D. J. 2016. The Evolution of Silicon Transport in Eukaryotes. *Molecular Biology and Evolution* 33:3226-3248. DOI: 10.1093/molbev/msw209.

Martin, W. 1., Stoebe, B., Goremykin, V., Hapsmann, S., Hasegawa, M. and Kowallik, K. V. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393:162-165. DOI: 10.1038/30234.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M. and Penny, D. 2002. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proceedings of the National Academy of Sciences of the United States of America* 99:12246-12251. DOI: 10.1073/pnas.182432999.

Martin, W. and Herrmann, R. G. 1998. Gene transfer from organelles to the nucleus : How

much, what happens, and why ? *Plant Physiology* 118:9-17. DOI: 10.1104/pp.118.1.9.

Matsumoto, T., Shinozaki, F., Chikuni, T., Yabuki, A., Takishita, K., Kawachi, M., Nakayama, T., Inouye, I., Hashimoto, T. and Inagaki Y. 2011. Green-colored plastids in the dinoflagellate genus *Lepidodinium* are of core chlorophyte origin. *Protist* 162:268-276. DOI: 10.1016/j.protis.2010.07.001.

Meguro, M., Ito, H., Takabayashi, A., Tanaka, R. and Tanaka, A. 2011. Identification of the 7-hydroxymethyl chlorophyll a reductase of the chlorophyll cycle in *Arabidopsis*. *The Plant Cell* 23:3442-3453. DOI: 10.1105/tpc.111.089714.

Minge, M. A., Shalchian-Tabrizi, K., Tørresen, O. K., Takishita, K., Probert, I., Inagaki, Y., Klaveness, D. and Jakobsen K. S. 2010. A phylogenetic mosaic plastid proteome and unusual plastid-targeting signals in the green-colored dinoflagellate *Lepidodinium chlorophorum*. *BMC Evolutionary Biology* 10:191. DOI: 10.1186/1471-2148-10-191.

Moore, R. B., Oborník, M., Janouskovec, J., Chrudimský, T., Vancová, M., Green, D. H., Wright, S. W., Davies, N. W., Bolch, C. J., Heimann, K., Slapeta, J., Hoegh-Guldberg, O., Logsdon, J. M. and Carter, D. A. 2008. *Nature* 451:959-963. DOI: 10.1038/nature06635.

Nagata, N., Tanaka, R., Satoh, S. and Tanaka, A. 2005. Identification of a vinyl reductase gene for chlorophyll synthesis in *Arabidopsis thaliana* and implications for the evolution of Prochlorococcus species. *The Plant Cell* 17:233-240. DOI: 10.1105/tpc.104.027276.

Nakayama, T., Kamikawa, R., Tanifuji G., Kashiyama, Y., Ohkouchi, N., Archibald, J. M. and Inagaki, Y. 2014. Complete genome of a nonphotosynthetic cyanobacterium in a diatom reveals recent adaptations to an intracellular lifestyle. *Proceeding of the National Academy of Science of the United States of America* 111:11407-11412. DOI: 10.1073/pnas.1405222111.

Nosenko, T., Lidie, K. L., Van Dolah, F. M., Lindquist, E., Cheng, J. F. and Bhattacharya, D. 2006. Chimeric plastid proteome in the Florida "red tide" dinoflagellate *Karenia brevis*. *Molecular Biology and Evolution*, 23: 2026-2038. DOI: 10.1093/molbev/msl074.

Nymark, M., Valle, K. C., Brembu, T., Hancke, K., Winge, P., Andresen, K., Johnsen, G. and Bones, A. M. 2009. An integrated analysis of molecular acclimation to high light in the marine diatom *Phaeodactylum tricornutum*. *PLoS ONE* 4:e7743. DOI: 10.1371/journal.pone.0007743.

Oborník, M. and Green, B. R. 2005. Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Molecular Biology and Evolution* 22:2343-2353. DOI: 10.1093/molbev/msi230.

Palmer, J., D. 2003. The symbiotic birth and spread of plastids: how many times and whodunit? *Journal of Phycology* 39:1-9 DOI:10.1046/j.1529-8817.2003.02185.x.

Panek, H. and O'Brian, M. R. 2002. A whole genome view of prokaryotic haem biosynthesis. *Microbiology* 148:2273-2282. DOI: 10.1099/00221287-148-8-2273.

Patron, N. J, Waller, R. F., Archibald, J. M. and Keeling, P. J. 2005. Complex protein targeting to dinoflagellate plastids. *Journal of Molecular Biology* 348:1015-1024. DOI: 10.1016/j.jmb.2005.03.030.

Patron, N. J., Waller, R. F. and Keeling, P. J. 2006. A tertiary plastid uses genes from two endosymbionts. *Journal of Molecular Biology* 357:1373-1382. DOI: 10.1016/j.jmb.2006.01.084.

Patron, N. J and Waller, R. F. 2007. Transit peptide diversity and divergence: A global analysis of plastid targeting signals. *BioEssays* 29:1048-58. DOI: 10.1002/bies.20638.

Petersen, T. N., Brunak, S., von Heijne, G. and Nielsen, H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods* 8:785-786. DOI: 10.1038/nmeth.1701.

Reinbothe, S. and Reinbothe, C. 1996. The regulation of enzymes involved in chlorophyll biosynthesis. *European Journal of Biochemistry* 237:323-343. DOI: 10.1111/j.1432-1033.1996.00323.x.

Reyes-Prieto, A. and Bhattacharya, D. 2007. Phylogeny of nuclear-encoded plastid-targeted proteins supports an early divergence of glaucophytes within Plantae. *Molecular Biology and Evolution* 24:2358-2361. DOI:10.1093/molbev/msm186.

Richardson, E., Dorrell, R. G. and Howe, C. J. 2014. Genome-wide transcript profiling reveals the coevolution of plastid gene sequences and transcript processing pathways in the fucoxanthin dinoflagellate *Karlodinium veneficum*. *Molecular Biology and Evolution* 31:2376-2386. DOI: 10.1093/molbev/msu189.

Rogers, M. B., Gilson, P. R., Su, V., McFadden, G. I. and Keeling, P. J. 2007. The complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Molecular Biology and Evolution* 24:54-62. DOI:10.1093/molbev/msl129.

Saldarriaga, J. F., Taylor, F. J., Keeling, P. J. and Cavalier-Smith, T. 2001. Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses and replacements. *Journal of Molecular Evolution* 53:204-213. DOI: 10.1007/s002390010210.

Sánchez-Baracaldo, P., Raven, J. A., Pisani, D. and Knoll, A. H 2017. Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proceedings of the National Academy of Sciences of the United States of America* 114:E7737-E7745. DOI: 10.1073/pnas.1620089114.

Sánchez Puerta, M. V., Bachvaroff, T. R. and Delwiche, C. F. 2005. The complete plastid genome sequence of the haptophyte *Emiliania huxleyi*: a comparison to other plastid genomes. *DNA research : an international journal for rapid publication of reports on genes and genomes.* 12:151-156. DOI: 10.1093/dnares/12.2.151

Sanchez-Puerta, M. V., Lippmeier, J. C., Apt, K. E. and Delwiche, C. F. 2007. Plastid genes in a non-photosynthetic dinoflagellate. *Protist* 158:105-117. DOI: 10.1016/j.protis.2006.09.004.

Sousa, F. L., Shavit-Grievink, L., Allen, J. F. and Martin, W. F. 2013. Chlorophyll biosynthesis gene evolution indicates photosystem gene duplication, not photosystem merger, at the origin of oxygenic photosynthesis. *Genome Biology and Evolution* 5:200-216. DOI: 10.1093/gbe/evs127.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313. DOI: 10.1093/bioinformatics/btu033.

Stiller, J. W., Reel, D. C. and Johnson, J. C. 2003. A single origin of plastids revisited: convergent evolution in organellar genome content. *Journal of Phycology*. 39:95-105 DOI:10.1046/j.1529-8817.2003.02070.x.

Suzuki, S., Hirakawa, Y., Kofuji, R., Sugita, M. and Ishida, K. 2016. Plastid genome sequences of *Gymnochlora stellata*, *Lotharella vacuolata*, and *Partenskyella glossopodia* reveal remarkable structural conservation among chlorarachniophyte species. *Journal of Plant Research* 129:581-590. DOI: 10.1007/s10265-016-0804-5.

Takishita, K., Ishida, K. and Maruyama, T. 2004. Phylogeny of nuclear-encoded plastid-targeted *gapdh* gene supports separate origins for the peridinin- and the fucoxanthin derivative-containing plastids of dinoflagellates. *Protist* 155:447-458. DOI: 10.1078/1434461042650325.

Takishita, K., Kawachi, M., Noël, M. H., Matsumoto, T., Kakizoe, N., Watanabe, M. M., Inouye, I., Ishida, K., Hashimoto, T. and Inagaki, Y. 2008. Origins of plastids and glyceraldehyde-3-phosphate dehydrogenase genes in the green-colored dinoflagellate *Lepidodinium chlorophorum*. *Gene* 410:26-36. DOI: 10.1016/j.gene.2007.11.008.

Taylor, F. J. R., Hoppenrath, M. and Saldarriaga, J. F. 2008. Dinoflagellate diversity and distribution. *Biodiversity and Conservation* 17:407-418. DOI: 10.1007/s10531-007-9258-3.

Tengs, T., Dahlberg, O. J., Shalchian-Tabrizi, K., Klaveness, D., Rudi, K., Delwiche, C. F. and Jakobsen, K. S. 2000. Phylogenetic analyses indicate that the 19' hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Molecular Biology and Evolution* 17:718-729. DOI: 10.1093/oxfordjournals.molbev.a026350.

Turmel, M., Otis, C. and Lemieux, C. 2009. The chloroplast genomes of the green algae *Pedinomonas minor*, *Parachlorella kessleri*, and *Oocystis solitaria* reveal a shared ancestry between the Pedinomonadales and Chlorellales. *Molecular Biology and Evolution* 26:2317-2331. DOI: 10.1093/molbev/msp138.

Van de Peer, Y., Rensing, S. A., Maier, U. G. and De Wachter, R. 1996. Substitution rate calibration of small subunit ribosomal RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. *Proceedings of the National Academy of Sciences of the United States of America* 93:7732-7736. DOI: 10.1073/pnas.93.15.7732.

van Dooren, G. G., Schwartzbach, S. D., Osafune, T. and McFadden, G. I. 2001. Translocation of proteins across the multiple membranes of complex plastids. *Biochimica et Biophysica Acta*. 1541:34-53. DOI: 10.1016/S0167-4889(01)00154-9.

Watanabe, M. M., Takeda, Y., Sasa, T., Inouye, I., Suda, S., Sawaguchi, T. and Chihara, M. 1987. A green dinoflagellate with chlorophylls *a* and *b*: morphology, fine structure of the chloroplast and chlorophyll composition. *Journal of Phycology* 23(s2):382-389. DOI: 10.1111/j.1529-8817.1987.tb04148.x.

Watanabe, M. M., Suda, S., Inouye, I., Sawaguchi, T. and Chihara, M. 1990. *Lepidodinium viride* gen. et sp. nov. (Gymnodinaiales, Dinophyta), a green dinoflagellate with a chlorophyll *a*- and *b*-containing endosymbiont. *Journal of Phycology* 26:741-751. DOI: 10.1111/j.0022-3646.1990.00741.x.

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V. and Zdobnov, E. M. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* 35:543–548. DOI: 10.1093/molbev/msx319.

Wicke, S., M€uller, K. F., de Pamphilis, C. W., Quandt, D., Wickett, N. J., Zhang, Y., Renner, S. S. and Schneeweiss, G. M. 2013. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the Broomrape family. *Plant Cell* 25:3711-3725. DOI: 10.1105/tpc.113.113373.

Wilhelm, C., Büchel, C., Fisahn, J., Goss, R., Jakob, T., Laroche, J., Lavaud, J., Lohr, M., Riebesell, U., Stehfest, K., Valentin, K. and Kroth, P. G. 2006. The regulation of carbon and nutrient assimilation in diatoms is significantly different from green algae. *Protist* 157:91-124. DOI: 10.1016/j.protis.2006.02.003.

Yamanashi, K., Minamizaki, K. and Fujita, Y. 2015. Identification of the *chlE* gene encoding oxygen-independent Mg-protoporphyrin IX monomethyl ester cyclase in cyanobacteria. *Biochemical and Biophysical Research Communications* 463:1328-1333. DOI: 10.1016/j.bbrc.2015.06.124.

Zapata, M., Fraga, S., Rodríguez, F. and Garrido, J. L. 2012. Pigment-based chloroplast types in dinoflagellates. *Marine Ecology Progress Series* 465:33-52. DOI: 10.3354/meps09879.

Zhang, Z. 1., Cavalier-Smith, T. and Green, B. R. 2002. Evolution of dinoflagellate unigenic minicircles and the partially concerted divergence of their putative replicon origins. *Molecular Biology and Evolution* 19:489-500. DOI: 10.1093/oxfordjournals.molbev.a004104.

Zhang, Z. 1., Green, B. R. and Cavalier-Smith, T. 1999. Single gene circles in dinoflagellate chloroplast genomes. *Nature* 400:155-159. DOI: 10.1038/22099.

**Tables**

Table 1 is not open to the public.

Table 2 Substitution models used in the phylogenetic analyses.

| | | | | ML analysis by RAxML | Bayesian analysis by PhyloBayes | | |
|---|---|---|---|---|---|---|---|
| | Orthologue name | Amino acid positions | Number of sequences | Substitution model | Number of generations (min. of 4 chains) | burn-in (for 20% of total generations) | bpcomp max diff. (* indicates max diff.>0.3 ) |
| heme synthesis | GTR | 361 | 83 | WAG+Γ | 19688 | 3937 | 0.239804 |
| | GSAT | 415 | 72 | WAG+Γ | 13067 | 2613 | 0.221443 |
| | ALAD | 300 | 114 | LG+Γ | 12178 | 2435 | 0.273875 |
| | PBGD | 267 | 109 | WAG+Γ | 22872 | 4574 | 0.660393* |
| | UROS | 162 | 77 | WAG+Γ | 11696 | 2339 | 0.268252 |
| | UROD | 302 | 206 | WAG+Γ | 11173 | 2234 | 0.445838* |
| | CPOX | 251 | 134 | WAG+Γ | 22185 | 4437 | 0.492636* |
| | PPOX | 300 | 87 | LG+Γ | 16547 | 3309 | 0.267774 |
| | FeCH | 214 | 129 | WAG+Γ | 19181 | 3836 | 0.196077 |
| Chl a synthesis | MgCH (ChlD) | 608 | 68 | LG+Γ | 25725 | 5145 | 0.537385* |
| | MgCH (ChlH) | 1095 | 96 | LG+Γ | 9172 | 1834 | 0.291622 |
| | MgPMT | 209 | 65 | WAG+Γ | 8605 | 1721 | 0.290851 |
| | POR | 285 | 132 | WAG+Γ | 18858 | 3771 | 0.266246 |
| | N-DVR | 290 | 63 | WAG+Γ | 26900 | 5380 | 0.443192* |
| | F-DVR | 358 | 37 | WAG+Γ | 15996 | 3199 | 0.248411 |
| | CS | 294 | 67 | Blosum62+Γ | 8863 | 1772 | 0.279679 |
| IPP synthesis | DXS | 604 | 57 | LG+Γ | 29914 | 5982 | 0.0660835 |
| | DXR | 358 | 55 | WAG+Γ | 29145 | 5829 | 0.0956384 |
| | IspD | 201 | 58 | WAG+Γ | 11490 | 2298 | 0.166566 |
| | IspE | 227 | 55 | WAG+Γ | 25279 | 5055 | 0.163873 |
| | IspF | 133 | 63 | LG+Γ | 8306 | 1661 | 0.184344 |
| | IspG | 595 | 55 | WAG+Γ | 29622 | 5924 | 0.146481 |
| | IspH | 328 | 68 | WAG + Γ | 8061 | 1612 | 0.274343 |

**Figures**

Fig. 1-6 are not open to the public.
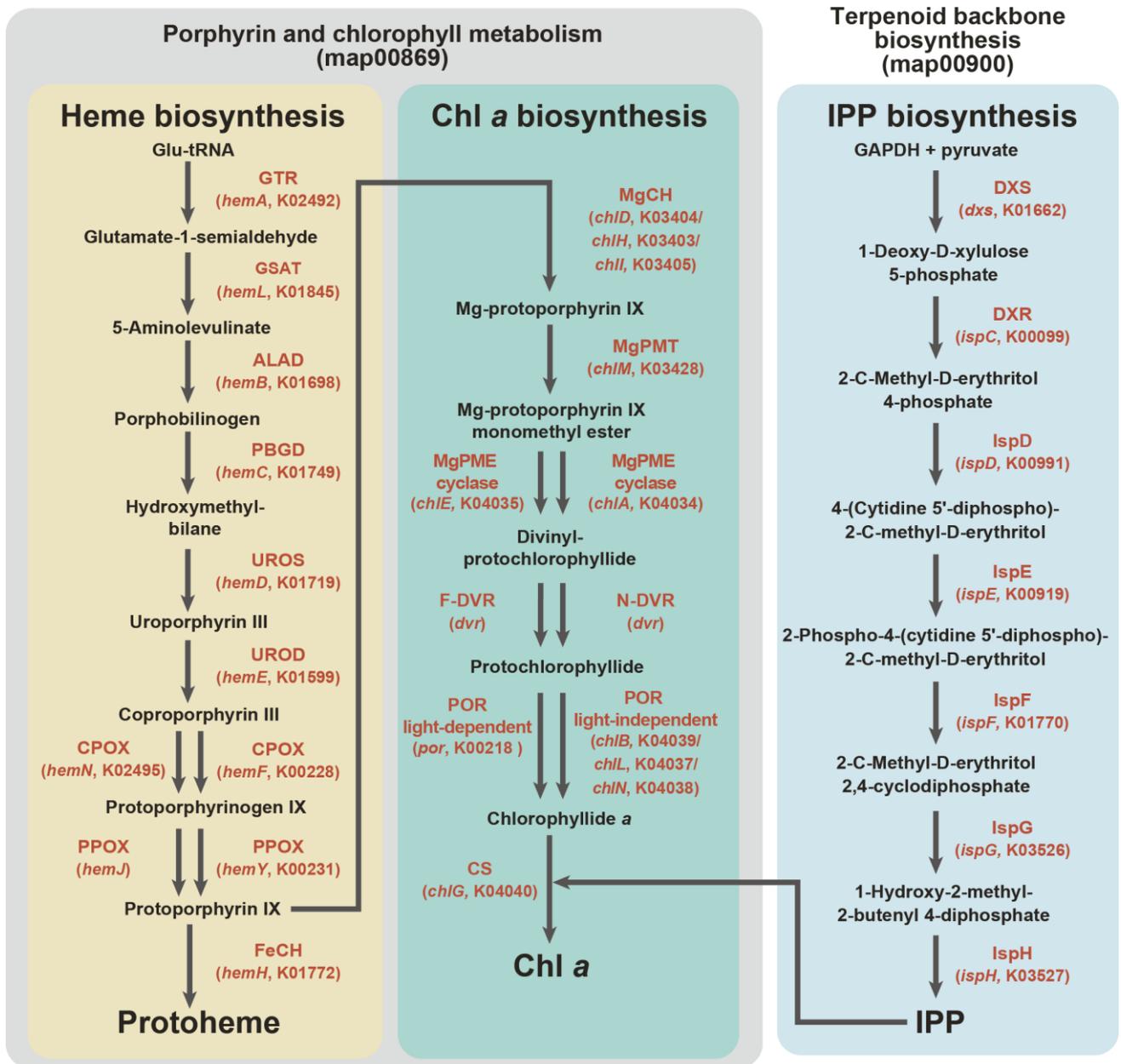
Fig. 7 Enzymes examined in the section 3, and their substrates and products.

C5 pathway for the heme biosynthesis is shaded in yellow. In this study, we regard the steps converting protoporphyrin IX to Chl a as the "Chl a biosynthetic pathway," and shaded in green. The non-mevalonate pathway is shaded in blue. The first two pathways belong to "Porphyrin and chlorophyll metabolism" (map00860), while the third pathway belongs to "Terpenoid backbone biosynthesis" (map00900) in Kyoto Encyclopedia of Genes and Genomes pathway (KEGG pathway, http://www.genome.jp/kegg/pathway.html). Enzymes involved in the three pathways, as well as their gene names and corresponding KOIDs, are shown in red.

**A** GTR

**B** GSAT

**C** ALAD

**D** PBGD
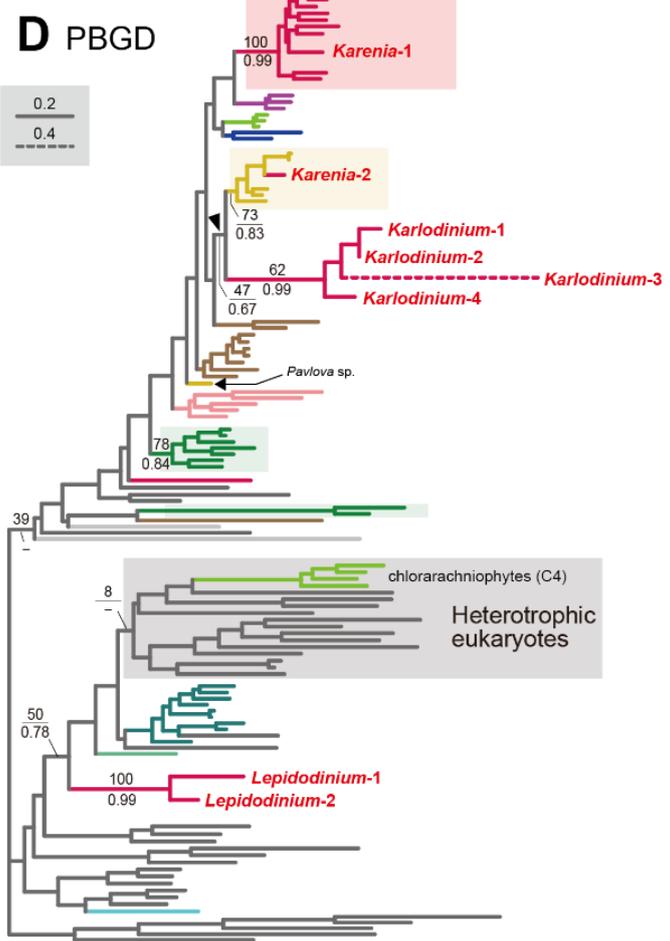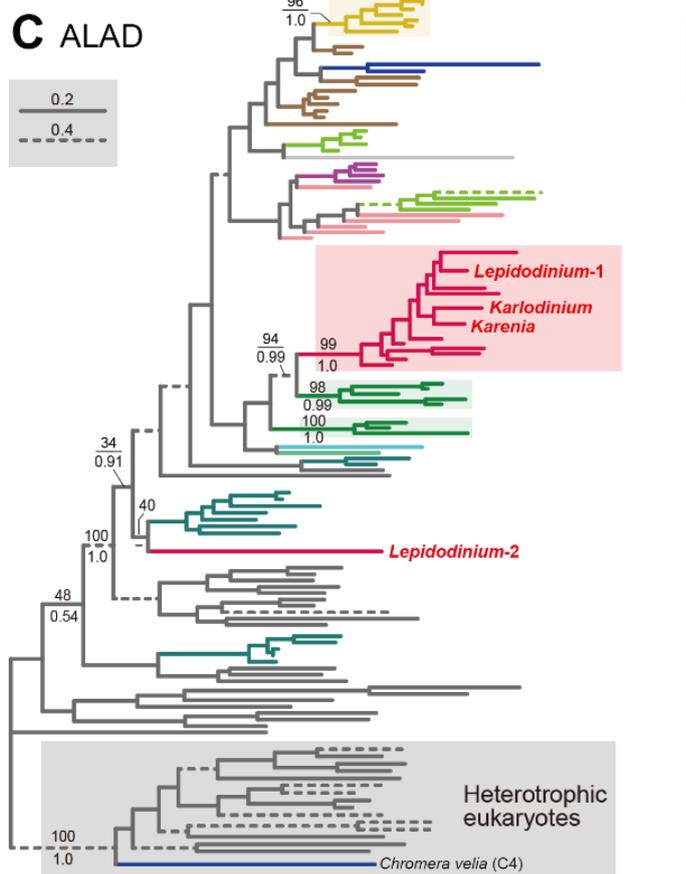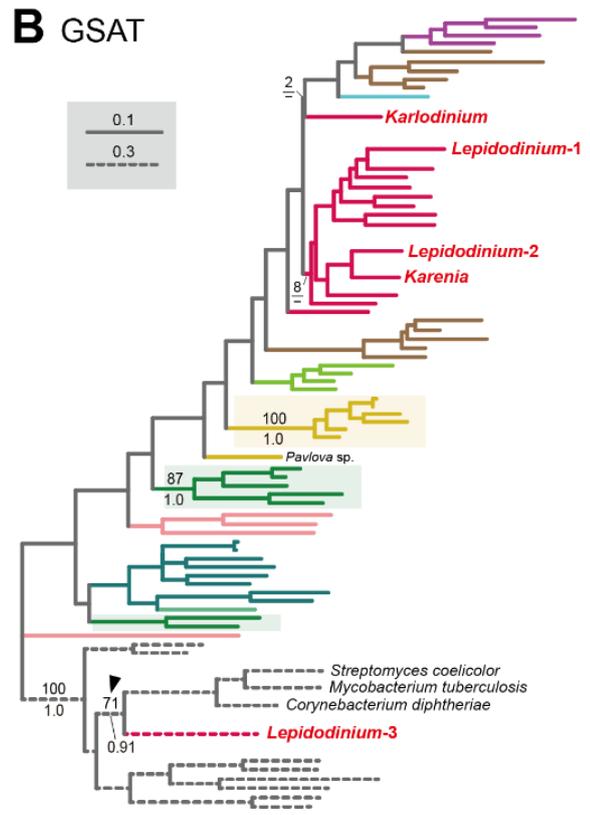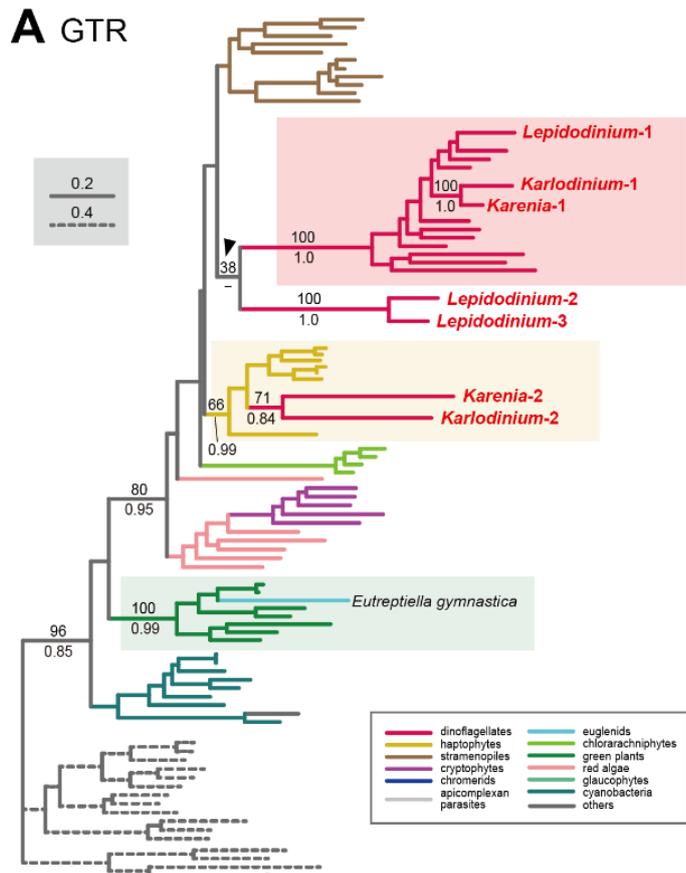
Fig. 8 Maximum-likelihood phylogenies of proteins involved in from the first step to the fourth step of C5 pathway for the heme biosynthesis.

We provide the maximum-likelihood bootstrap values (MLBPs), as well as Bayesian posterior probabilities (BPPs), only for the selected nodes, which are important to infer the origins of the proteins of *Karenia brevis*, *Karlodinium veneficum* and *Lepidodinium chlorophorum*. Dash marks represent the corresponding BPPs <0.50. MLBPs and BPPs are shown above and beneath the corresponding nodes, respectively. Subtrees/branches are color-coded (see the inset for the details). Statistically supported clades of the dinoflagellate, haptophyte and green plant sequences are highlighted by red, yellow and green backgrounds, respectively. Clades comprising the sequences of heterotrophic eukaryotes, which were predicted to be involved in C4 pathway, were shaded in gray. (A) Glutamyl-tRNA reductase (GTR). (B) Glutamate-1-semialdehyde 2,1-aminomutase (GSAT). (C) Delta-aminolevulinic acid dehydratase (ALAD). (D) Porphobilinogen deaminase (PBGD). The remaining proteins involved in C5 pathway are presented in following Fig.9 and Fig.10. Note that the substitution rates of the dashed branches are different from that of the solid lines as indicated in each figure.

**A UROS**

0.2
0.4

46
100 / 0.99
18 / –
37 / –
*Pavlova* sp.
*Cyanidioschyzon merolae*
***Lepidodinium***
43 / –

83 / 0.92
76 / 0.95
***Karenia***
***Karlodinium*-1**

93 / 0.95
*Euglena gracilis*(C4)
Heterotrophic eukaryotes

95 / 0.94
***Karlodinium*-2**
*Rhodosorus marinus*

**B UROD**

0.2
0.4

***Karlodinium*-1**
***Karlodinium*-2**
***Karenia*-1**
***Karenia*-2**
**H1**
100 / 1.0
100 / 1.0

93 / 0.91
*Pavlova* sp.

78 / –
99 / 0.99
100 / 0.99

***Karlodinium*-3**
*Chaetoceros* cf. *neogracile*
***Lepidodinium*-1**
*Odontella aurita*
diatom ensymbionts in *Durinskia* & *Kryptoperidinium*
100 / 1.0

100 / 1.0
*Eutreptiella gymnastica*
***Karenia*-3**
**H2**

100 / 1.0
***Lepidodinium*-2**
**D1**
*Pavlova* sp.

39 / –
79 / 0.99
**H3**
***Karlodinium*-4**
***Karenia*-4**
*Pavlova* sp.

89 / 0.99
***Lepidodinium*-3**
***Lepidodinium*-4**
**D2**

100 / 1.0
100 / 1.0

***Lepidodinium*-5**
***Lepidodinium*-6**
***Karenia*-5**
***Lepidodinium*-7**
**D3**
100 / 1.0

98 / 0.97
100 / 1.0
20 / –
Heterotrophic eukaryotes

72 / –
32
87 / 0.99

100 / 0.99

**C CPOX**

0.2
0.4

***Karenia***
98 / 0.99

95 / 0.99
**H1**

Heterotrophic eukaryotes
99 / 0.98
100 / 1.0
19 / –

75 / 0.87
*Gloeochaete wittrockiana* (C4)
100 / 0.99

77 / 0.87
***Lepidodinium*-1** (C4)

**H2**
86 / 0.98
100 / 1.0
***Lepidodinium*-2**
***Lepidodinium*-3**

97 / 0.99

100 / 0.98

| dinoflagellates | euglenids |
|---|---|
| haptophytes | chlorarachniophytes |
| stramenopiles | green plants |
| cryptophytes | red algae |
| chromerids | glaucophytes |
| apicomplexan parasites | cyanobacteria |
| | others |

81

Fig. 9 Maximum-likelihood phylogenies of seven proteins involved in from the fifth step to the seventh step of C5 pathway for the heme biosynthesis. The details of this figure are same as those of Fig. 8. (A) Uroporphyrinogen III synthase (UROS). (B) Uroporphyrinogen decarboxylase (UROD). (C) Coproporphyrinogen oxidase (CPOX). The remaining proteins involved in C5 pathway are presented in following Fig.10. Note that the substitution rates of the dashed branches are different from that of the solid lines as indicated in each figure.

**A** PPOX

0.2
0.4

100
0.97
D1
*Lepidodinium*-1
*Karenia*-1

92
0.99

38
0.96
*Chrysochromulina polylepis*
*Prymnesium parvum*
100
0.99
*Lepidodinium*-2
*Karlodinium*
*Karenia*-2
25
—
*Gephyrocapsa oceanica*
*Emiliania huxleyi*
*Isochrysis galbana*
*Coccolithus pelagicus*

*Pavlova* sp.

*Euglena gracilis* (C4)
chlorarachniophytes (C4)

93
0.99
100
0.98
100
0.98
D2
*Lepidodinium*-3
*Vitrella brassicaformis* (C4)
C4

Heterotrophic eukaryotes

dinoflagellates
haptophytes
stramenopiles
cryptophytes
chromerids
apicomplexan parasites
euglenids
chlorarachniophytes
green plants
red algae
glaucophytes
cyanobacteria
others

**B** FeCH

0.2
0.4

100
1.0
*Lepidodinium*-1

32
0.61
*Lepidodinium*-2

100
1.0

75
0.99
*Vibrio cholerae*
*Escherichia coli*
100
1.0
*Lepidodinium*-3
*Karenia*-1

*Gloeochaete wittrockiana* (C4)
cryptophytes (C4)
chlorarachniophytes (C4)
Heterotrophic eukaryotes
89
0.96

89
0.95
57
0.66
H1
19
—
*Karenia*-2
73
0.99
H2
*Pavlova* sp.
70
0.96

6
—
*Lepidodinium*-4

*Karenia*-3
*Synechococcus* sp.
76
0.94
*Paulinella chromatophora*
*Prochlorococcus marinus*
*Karlodinium*

Fig. 10 Maximum-likelihood phylogenies of proteins involved in from the eighth step to the final step of C5 pathway for the heme biosynthesis.

The details of this figure are same as those of Fig. 8. (A) Protoporphyrinogen IX oxidase (PPOX). (B) Ferrochelatase (FeCH). Note that the substitution rates of the dashed branches are different from that of the solid lines as indicated in each figure.

**A** MgCH(ChlD)

cryptophytes
chromerids
*Eutreptiella gymnastica*

*Lepidodinium*

*Karlodinium*
*Karenia*

chlorarachniophytes

**B** MgCH(ChlH)

D1

*Lepidodinium*
chlorarachniophytes

*Karenia-1*
*Karlodinium-1*

Gp1

ChlH-1

dinoflagellates
haptophytes
stramenopiles
cryptophytes
chromerids
apicomplexan
parasites
euglenids
chlorarachniphytes
green plants
red algae
glaucophytes
cyanobacteria
others

D2

*Karlodinium-2*
*Karenia-2*

Gp2

*Eutreptiella gymnastica*

*Isochrysis galbana*

ChlH-2

**C** MgPMT

chlorarachniophytes
*Lepidodinium*

*Karlodinium*
*Karenia*
*Pavlova* sp.

**E** N-DVR

*Pavlova* sp.
*Pavlova* sp.

*Karenia*
*Karlodinium*

*Coccolithus pelagicus*

*Lepidodinium*
*Eutreptiella gymnastica*

**D** F-DVR

*Karenia*

chlorarachniophytes
*Lepidodinium*
*Eutreptiella gymnastica*

chlorarachniophytes

85

Fig. 11 Maximum-likelihood phylogenies of proteins involved in the first, second and fourth step of Chl *a* biosynthetic pathway. The details of this figure are same as those of Fig. 8. (A) ChlD, one of the two nucleus-encoded subunits of Mg-chelatase (MgCH). (B) ChlH, the other nucleus-encoded subunit of MgCH. (C) S-adenosylmethionine:Mg-protoporphyrin O-methyltransferase (MgPMT). (D) divinyl chlorophyllide *a* 8-vinyl-reductase using ferredoxin for electron donor (F-DVR). (E) divinyl chlorophyllide *a* 8-vinyl-reductase using NADPH for electron donor (N-DVR). Note that I present no phylogeny of ChlI (involved in the first step of this pathway)or MgPME cyclase (involved in the third step of this pathway), as the former is plastid-encoded, and the latter was not identified in dinoflagellates (regardless of plastid-type), diatoms, cryptophytes or haptophytes. The remaining proteins involved in Chl *a* biosynthetic pathway are presented in following Fig.12.Note that the substitution rates of the dashed branches are different from that of the solid lines as indicated in each figure.

**A** POR

65
0.71

*Karenia*-1
*Karlodinium*-1

95
0.98

D1

*Lepidodinium*-1

97
0.99

chlorarachniophytes

99
0.62

92
0.85

*Lepidodinium*-2

0.2
0.4

*Karlodinium*-2
*Karlodinium*-3
*Karenia*-2

99
99
0.99
0.99

100
0.99

98
0.99

100
0.98

86
—

93
—

100
0.99

D2

84
0.95

**B** CS

0.1

18
0.88

chlorarachniophytes

13
—

96
0.98

80
0.98

*Lepidodinium*-1
*Lepidodinium*-2

8
—

100
1.0

100
1.0

*Karlodinium*
*Karenia*

98
1.0

17
—

96
0.97

*Eutreptiella gymnastica*

4
=

31
—

73
1.0

87

Fig. 12 Maximum-likelihood phylogenies of proteins involved in from the fifth step to the final step of Chl *a* biosynthetic pathway.

The details of this figure are same as those of Fig. 8. (A) light-dependent protochlorophyllide reductase (POR). (B) chlorophyll synthase (CS). Note that the substitution rates of the dashed branches are different from that of the solid lines as indicated in each figure.
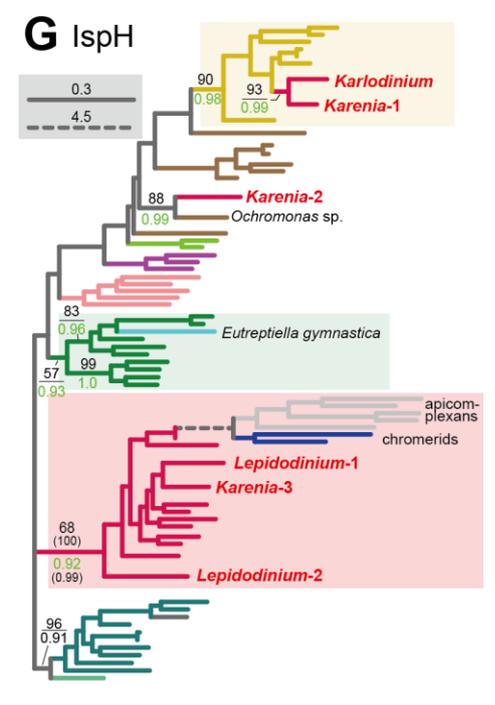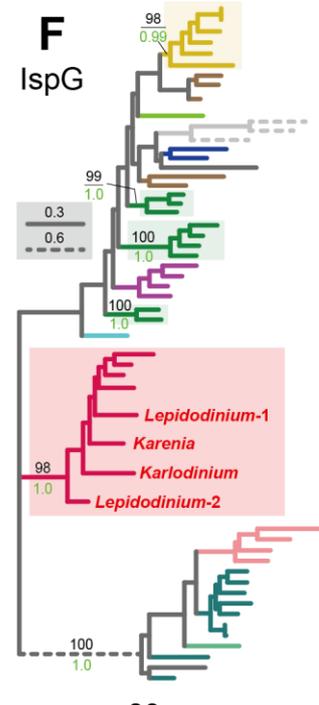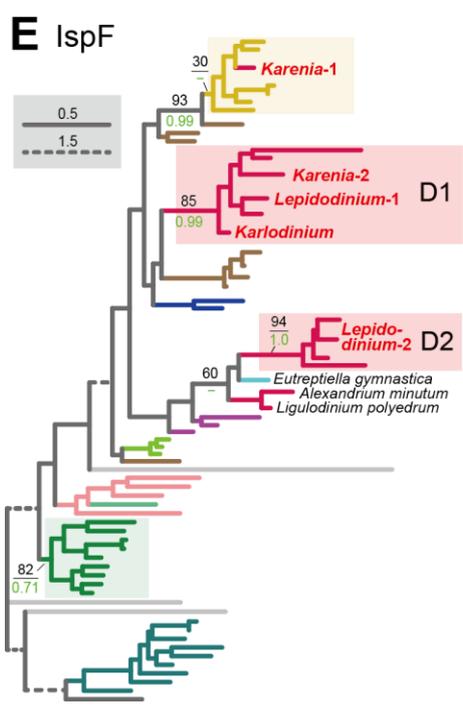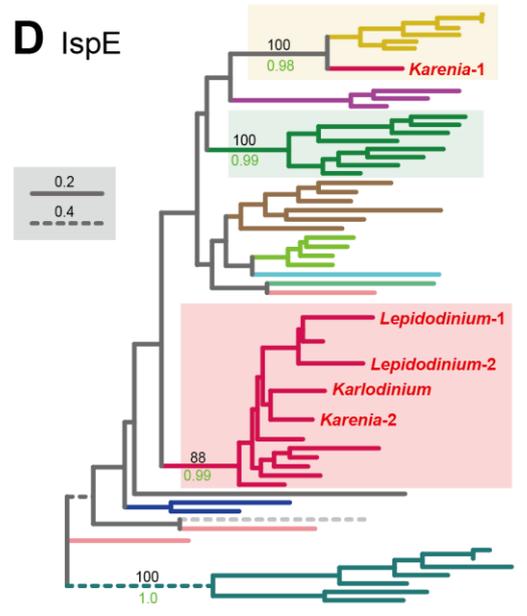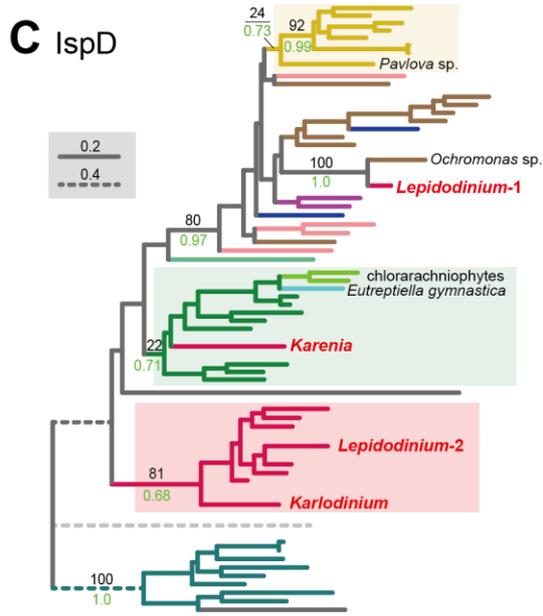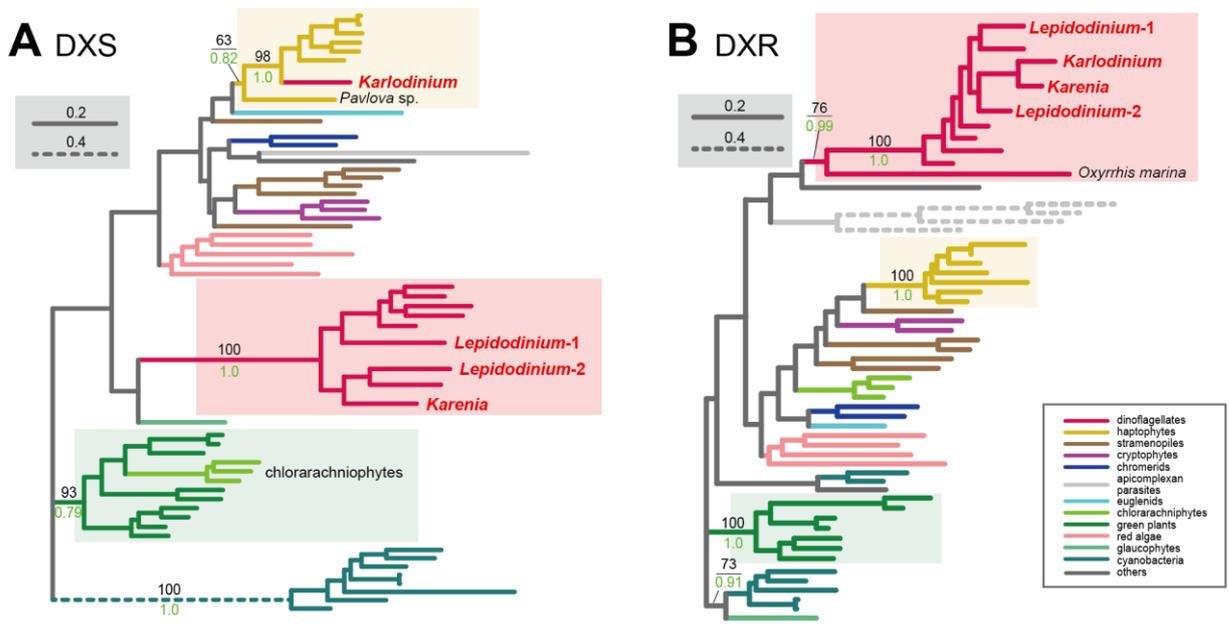
Fig. 13 Maximum-likelihood phylogenies of seven proteins involved in the non-mevalonate pathway for IPP biosynthesis.

The details of this figure are same as those of Fig. 8. (A) 1-deoxy-D-xylulose-5-phosphate (DXP) synthase (DXS). (B) DXP reductoisomerase (DXR). (C) 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (IspD). (D) 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE). (E) 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF). (F) 1-hydroxy-2-methyl-2-butenyl 4-diphosphate (HMB-PP) synthase (IspG). (G) HMB-PP reductase (IspH). Note that the substitution rates of the dashed branches are different from that of the solid lines as indicated in each figure.
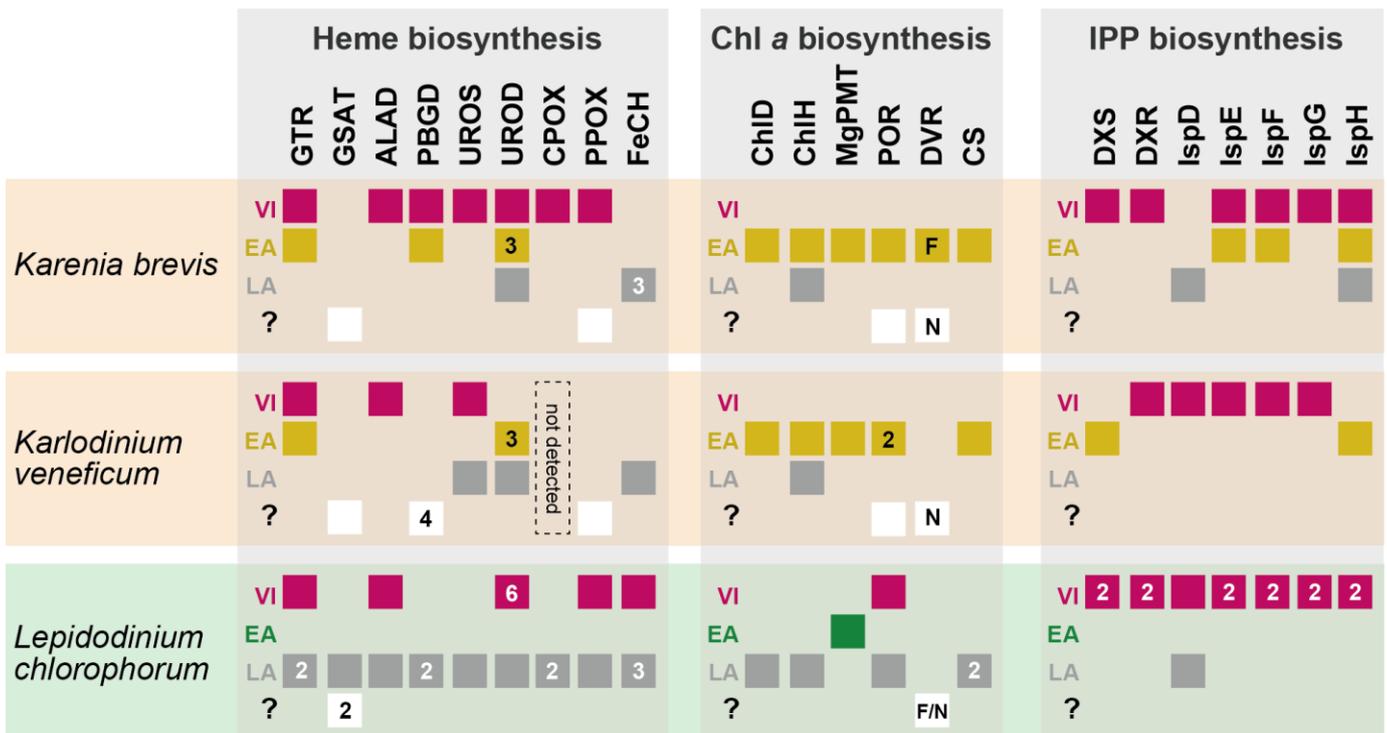
Fig. 14 Overview of the origins of proteins involved in three plastid-localized biosynthetic pathways in *Karenia brevis*, *Karlodinium veneficum* and *Lepidodinium chlorophorum*.

The origins of proteins of interest were classified into three types, (i) "VI-type" which were vertically inherited from the ancestral dinoflagellate beyond haptophyte/green algal endosymbiosis, (ii) "EA-type" which were acquired from the endosymbiont, and (iii) "LA-type" which were acquired from organisms distantly related to the host (dinoflagellates) or endosymbiont (haptophytes or green algae). Squares indicate the numbers and types of proteins of interest in the three dinoflagellates. In case of multiple versions being identified in one species, the numbers of the versions are shown in the corresponding squares. The squares are color-coded as follows: red, VI-type protein; yellow, EA-type in *Karenia brevis* and *Karlodinium veneficum*; green, EA-type in *Lepidodinium*; grey, LA-type. For DVR involved in the Chl *a* biosynthetic pathway, we distinguish N-DVR and F-DVR by labeling "N" and "F," respectively. The open squares in the fourth lows labeled with question marks represent the sequences of which origins remain uncertain.