A peer-reviewed version of this preprint was published in PeerJ on 3 August 2018.

View the peer-reviewed version (peerj.com/articles/5345), which is the preferred citable publication unless you specifically need to cite this preprint.

Matsuo E, Inagaki Y. 2018. Patterns in evolutionary origins of heme, chlorophyll *a* and isopentenyl diphosphate biosynthetic pathways suggest non-photosynthetic periods prior to plastid replacements in dinoflagellates. PeerJ 6:e5345 https://doi.org/10.7717/peerj.5345

Trends in reconstruction of three nucleus-encoded, plastid-localized pathways for the heme, chlorophyll *a* and isopentenyl diphosphate biosynthesises in two separate dinoflagellate lineages bearing non-canonical plastids

Eriko Matsuo [1], **Yuji Inagaki** [Corresp. 1, 2]

[1] Graduate School of Biological and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

[2] Center for Computational Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

Corresponding Author: Yuji Inagaki
Email address: yuji@ccs.tsukuba.ac.jp

Background: The ancestral dinoflagellate most likely established a peridinin-containing plastid, which have been inherited to the extant photosynthetic descendants. However, kareniacean dinoflagellates and *Lepidodinium* species were known to bear “non-canonical” plastids lacking peridinin, which were established through haptophyte and a green algal endosymbioses, respectively. 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 how “exogenous” genes reshaped individual metabolic pathways localized in a non-canonical plastid.

Results: In this study, we surveyed transcriptomic data from two kareniacean species (*Karenia brevis* and *Karlodinium veneficum*) and *Lepidodinium 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 our 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. We observed a clear difference in the contribution of VI-type proteins across the three pathways. In both *Karenia*/*Karlodinium* and *Lepidodinium*, we 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.

Discussion: Pioneering works hypothesized that the ancestral kareniacean species had lost the photosynthetic activity prior to haptophyte endosymbiosis. The absence of VI-type proteins in the Chl *a* biosynthetic pathway in *Karenia* or *Karlodinium* is in good agreement with the putative non-photosynthetic nature proposed for their ancestor. The dominance of proteins with haptophyte origin in the *Karenia*/*Karlodinium* pathway suggests that their ancestor rebuilt the particular pathway by genes acquired from the endosymbiont. Likewise, we here propose that the ancestral *Lepidodinium* likely experienced a non-photosynthetic period and discarded the entire Chl *a* biosynthetic pathway prior to the green algal endosymbiosis. Nevertheless, *Lepidodinium* rebuilt the pathway by genes transferred from phylogenetically diverse organisms, rather than the green algal endosymbiont. We explore the reasons why green algal genes were barely utilized to reconstruct the *Lepidodinium* pathway.

Trends in reconstruction of three nucleus-encoded, plastid-localized pathways for the heme, chlorophyll *a* and isopentenyl diphosphate biosyntheses in two separate dinoflagellate lineages bearing non-canonical plastids.

Eriko Matsuo[1], Yuji Inagaki[1,2].

[1]Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan. [2]Center for Computational Sciences, University of Tsukuba, Tsukuba, Ibaraki.

To whom correspondence should be addressed: Yuji Inagaki, yuji@ccs.tsukuba.ac.jp

Abstract

Background: The ancestral dinoflagellate most likely established a peridinin-containing plastid, which has been inherited to the extant photosynthetic descendants. However, kareniacean dinoflagellates and *Lepidodinium* species were known to bear "non-canonical" plastids lacking peridinin, which were established through haptophyte and a green algal endosymbioses, respectively. 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 endosymbionts. 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 how "exogenous" genes reshaped individual metabolic pathways localized in a non-canonical plastid.

Results: In this study, we surveyed transcriptomic data from two kareniacean species (*Karenia brevis* and *Karlodinium veneficum*) and *Lepidodinium chlorophorum*, and identified proteins involved in three plastid metabolic pathways synthesizing chlorophyll *a* (Chl *a*), heme and isopentenyl diphosphate (IPP). The origins of the individual proteins of our 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. We observed a clear difference in the contribution of VI-type proteins across the three pathways. In both *Karenia*/*Karlodinium* and *Lepidodinium*, we observed a substantial contribution of VI-type proteins to the IPP and heme biosyntheses. In sharp contrast, VI-type protein was barely detected in the Chl *a* biosynthesis in the three dinoflagellates.

Discussion: Pioneering works hypothesized that the ancestral kareniacean species had lost the photosynthetic activity prior to haptophyte endosymbiosis. The absence of VI-type proteins in the Chl *a* biosynthetic pathway in *Karenia* or *Karlodinium* is in good agreement with the putative non-photosynthetic nature proposed for their ancestor. The dominance of proteins with haptophyte origin in the *Karenia*/*Karlodinium* pathway suggests that their ancestor rebuilt the particular pathway by genes acquired from the endosymbiont. Likewise, we here propose that the ancestral *Lepidodinium* likely experienced a non-photosynthetic period and discarded the entire Chl *a* biosynthetic pathway prior to the green algal endosymbiosis. Nevertheless, *Lepidodinium* rebuilt the pathway by genes transferred from phylogenetically diverse organisms, rather than the green algal endosymbiont. We explore the reasons why green algal genes were barely utilized to reconstruct the *Lepidodinium* pathway.

Introduction

Dinoflagellates are aquatic unicellular eukaryotes belonging to one of major taxonomic groups of eukaryotes, Alveolata. About half species of dinoflagellates described to date are photosynthetic (Taylor et al, 2008). Typical photosynthetic dinoflagellates harbor plastids containing chlorophylls *a* and *c* (Chl *a*+*c*), which are remnants of a red algal endosymbiont captured by the common ancestor of dinoflagellates (Hoek et al, 1995; Janouškovec et al, 2010). Comparing to red alga-derived, Chl *a*+*c*-containing plastids in other eukaryotic algae, namely cryptophytes, chromerids (e.g., *Chromera velia* and *Vitrella brassicaformis*), stramenopiles, and haptophytes, the plastids in vast majority of dinoflagellates are unique in containing a carotenoid called peridinin (Jeffrey et al, 1975; Zapata et al, 2012). However, some species are known to bear "non-canonical" plastids,

which are distinctive from "peridinin-containing plastids" in the majority of photosynthetic dinoflagellates in both pigment composition and evolutionary origin. The plastids in members of genera *Karenia* and *Karlodinium* (family Kareniaceae) contain 19′-hexanoyl-fucoxanthin along with Chl *a*+*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 of a pedinophyte green alga (Watanabe et al, 1987, 1990; Matsumoto et al, 2011; Kamikawa et al, 2015a). At the morphological level, the endosymbionts in the two lineages described above were extensively reduced in the host cells, leaving only plastids in the endosymbiont-derived compartments. At the genetic level, endosymbiont genes, particularly ones for plastid functions and maintenance (henceforth we designate as "plastid-related genes"), were transplanted into the host genomes of the two dinoflagellate lineages (Takishita et al, 2004; Nosenko et al, 2006; Patron et al, 2006; Minge et al, 2010; Burki et al, 2014). Such gene transfer from an endosymbiont to its host (endosymbiotic gene transfer or EGT) is regarded as one of the keys for the host-endosymbiont interlock at the genetic level. Thus, both haptophyte-derived and green alga-derived plastids in dinoflagellates are regarded as genuine organelles in the current host cells. In addition, a particular 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). Unlike the two non-canonical plastids described above, the obligate diatom endosymbionts retain their own nuclei, mitochondria, and plastids in the dinoflagellate cells (Tomas & 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 (Hehenberger et al, 2016). Finally, some non-photosynthetic dinoflagellates engulf and digest eukaryotic algae as preys, except their

plastids. These dinoflagellates have been known to maintain and utilize the plastids of the preys for a certain period (Geider & Gunter 1989; Hewes et al. 1998; Gast et al, 2007; Onuma & Horiguchi 2015). As these "stolen" plastids (kleptoplastids) are eventually digested, the host cells repeat engulfing the preys.

Pioneering works on *Karenia brevis*, *Karlodinium veneficum*, and *Lepidodinium chlorophorum* identified substantial numbers of endosymbiotically transferred genes in the host genomes (Ishida & 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 *Karenia* nuclear gene for 1-deoxy-D-xylulase 5-phosphate synthase (DXS), and the *Lepidodinium* nuclear gene for 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), 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 throughout the dinoflagellate evolution, rather than acquired from the endosymbiont. Furthermore, a certain fraction of plastid-related genes was 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 *Karenia* (Nosenko et al. 2006), and *csp41* for an mRNA-binding protein in *Lepidodinium* was found to share the origin with the stramenopile orthologues (Minge et al. 2010). In light of previous gene surveys in the *Karenia*/*Karlodinium* and *Lepidodinium* 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 our knowledge, it has been unclear whether the overall degree of evolutionary "chimerism" in plastid proteome vary among non-canonical plastids established

separately in dinoflagellate evolution. Likewise, we are uncertain whether the trend in evolutionary chimerism shared among multiple plastid-localized metabolic pathways within a non-canonical plastid. In this study, we investigate the evolutionary origins of enzymes involved in the biosyntheses of heme, chlorophyll *a* (Chl *a*), and isopentenyl diphosphate (IPP), occurred in the non-canonical plastids in *Karenia*/*Karlodinium* and *Lepidodinium*. We 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 & Vermaas, 2002; Kuzuyama, 2002).

The vast majority of the enzymes involved in C5 pathway for the heme biosynthesis, the non-mevalonate pathway for IPP biosynthesis, and the Chl *a* biosynthesis is nucleus-encoded (Oborník & Green, 2005; Bentlage et al, 2016). The three pathways in photosynthetic eukaryotes and those in cyanobacteria are principally homologous to each other, suggesting that the pathways in photosynthetic eukaryotes can be traced back to those endosymbiotically acquired from the cyanobacterial endosymbiont that gave rise to the first plastid (primary plastid). Consistent with the cyanobacterial ancestry, the three pathways described above occur in the plastids in photosynthetic eukaryotes. In the following paragraphs, we overview the three pathways in cyanobacteria and land plants, which have been studied experimentally as the models of photosynthetic bacteria and eukaryotes, respectively. We do aware of the heme biosynthetic pathway in apicomplexan parasites and their relatives being diversified (Kořený et al, 2011). However, we do not mention these exceptions below, as we can discuss the evolutions of the heme biosynthesis in *Karenia*/*Karlodinium* and *Lepidodinium* without acknowledging the unorthodox pathways in apicomplexan parasites and their relatives.

We briefly review C5 pathway for the heme biosynthesis here (Fig. 1). The first step in C5 pathway was catalyzed by glutamyl-tRNA reductase (GTR) and glutamate-1-semialdehyde

2,1-aminomutase (GSAT) to transform glutamyl-tRNA to aminolevulinic acid (ALA) (Panek & O'Brian, 2002). After ALA synthesis, two ALA molecules undergo a condensation reaction by delta-aminolevulinic acid dehydratase (ALAD) to form porphobilinogen with a pyrrole ring. Four porphobilinogen molecules are combined into a single hydroxymethylbilane molecule by porphobilinogen deaminase (PBGD). Uroporphyrinogen III synthase (UROS) then circularizes hydroxymethylbilane to yield uroporphyrinogen III. Four acetyl side chains in uroporphyrinogen III are removed by uroporphyrinogen decarboxylase (UROD) to generate coproporphyrinogen III. Then, protoporphyrinogen IX is formed from coproporphyrinogen III by coproporphyrinogen oxidase (CPOX). Two types of functionally homologous but evolutionarily distinct CPOX encoded by *hemN* and *hemF* are identified in cyanobacteria. Protoporphyrinogen IX oxidase (PPOX) converts protoporphyrinogen IX to protoporphyrin IX. As seen in COPX above, functionally homologous but evolutionarily distinct PPOX, which are encoded by *hemJ* and *hemY*, have been identified in cyanobacteria. Finally, the heme biosynthesis is completed by ferrochelatase (FeCH), which inserts an iron ion into protoporphyrin IX yielding protoheme. Land plants (and most of eukaryotic algae) appeared to use the nucleus-encoded, plastid-targeted enzymes for C5 pathway (Kořený et al, 2013). Overall, land plants used the pathway of cyanobacterial origin with modifications described below. (i) PBGD, UROS, and UROD, which are likely acquired from phylogenetically diverse organisms in non-endosymbiotic contexts, were identified. (ii) Neither *hemN*-version of CPOX nor *hemJ*-version of PPOX was detected.

The pathway synthesizing Chl *a* partially overlaps with C5 pathway for the heme biosynthesis (i.e. the first to eighth steps converting glutamyl-tRNA to protoporphyrin IX; see Fig. 1) (Reinbothe et al, 1996; Beale 1999). We regard the steps converting protoporphyrin IX to Chl *a* as the "Chl *a* biosynthetic pathway" in this study, and briefly overview below. A magnesium ion is inserted into protoporphyrin IX by Mg-chelatase (MgCH), which consists of three hetero

subunits encoded by *chlD*, *chlH* and *chlI*. Mg-protoporphyrin IX is converted into Mg-protoporphyrin IX monomethyl ester by *S*-adenosylmethionine:Mg-protoporphyrin *O*-methyltransferase (MgPMT). In the next step, the characteristic isocyclic ring of chlorophylls (also known as E-ring) is formed by Mg-protoporphyrin IX monomethyl ester (oxidative) cyclase (MgPME cyclase), generating divinyl protochlorophyllide. In the majority of cyanobacteria, a multi-subunit enzyme catalyzes the MgPME cyclase activity (Yamanashi et al, 2015; Chen et al, 2017), and a membrane-binding protein encoded by *chlA* was identified as the catalytic subunit (Minamizaki et al, 2008; Hollingshead et al, 2012). In addition, another membrane-related component of MgPME cyclase was identified recently in *Synechocystis* PCC 6803 (Hollingshead et al, 2012). Some cyanobacteria possess an evolutionarily distinct, single-subunit MgPME cyclase encoded by *chlE* (Yamanashi et al, 2015). The C8 vinyl group in divinyl protochlorophyllide is reduced by divinyl chlorophyllide *a* 8-vinyl-reductase (DVR) to yield protochlorophyllide. In cyanobacteria, two evolutionarily distinct types of DVR, namely N-DVR and F-DVR, are exclusively distributed, and the former and the latter use NADPH and reduced ferredoxin for electron donors, respectively (Ito et al, 2008; Ito & Tanaka, 2014). Land plants possess both N-DVR and F-DVR, but the latter likely participates chlorophyll *b* metabolism instead of Chl *a* biosynthesis (Ito & Tanaka, 2014). The C17=C18 double bond in protochlorophyllide is reduced to yield chlorophyllide *a* by protochlorophyllide reductase (POR). Again, two evolutionarily distinct types of POR (light-dependent and light-independent version) have been identified in cyanobacteria (Fujita et al, 2003). The light-independent version comprises three hetero subunits encoded by *chlB*, *chlL* and *chlN*, whereas the light-dependent version comprises a single polypeptide encoded by *por*. It is also important to note that the subunits composed of the light-independent POR are plastid-encoded in land plants. Both types of POR were identified in land plants, except flowering plants in which the light-independent version is absent (Suzuki & Bauer

1992; Schoefs, 2000). In cyanobacteria, the order of the two reactions carried out by DVR and POR are believed to be interchangeable, but Nagata et al. (2007) reported that N-DVR in *Arabidopsis* prefers divinyl chlorophyllide (the product of POR) to divinyl protochlorophyllide (the product of MgPME cyclase) as the substrate. Finally, Chl *a* is yielded by chlorophyll synthase encoded by *chlG*, which adds the phytol chain to C17 in chlorophyllide *a*.

IPP is an essential precursor for various terpenoids including the phytol residue in Chl *a*. There are two distinct pathways for the IPP biosynthesis, namely mevalonate and non-mevalonate pathways (Lichtenthaler et al, 1997; Rohmer, 1999; Kuzuyama, 2002; Eisenreich et al, 2004). The former was likely established in early eukaryotic evolution, and the latter is found in prokaryotes including cyanobacteria as well as diverse photosynthetic eukaryotes. The mevalonate pathway is operated in the cytosol, while the non-mevalonate pathway in photosynthetic eukaryotes (including land plants) occurs in the plastid (Kuzuyama, 2002; Dubey et al, 2003). The genes encoding enzymes composed of the non-mevalonate pathway in the cyanobacterial endosymbiont were most likely transplanted into the host genome during primary endosymbiosis, as the enzymes involved in the non-mevalonate pathway are nucleus-encoded and share the evolutionary origins with the cyanobacterial counterparts (Grauvogel et al, 2007). In land plants, the non-mevalonate pathway appeared to supply IPP to the Chl *a* synthesis, suggesting that the two biosyntheses are tightly coupled together in the plastids (Lichtenthaler et al, 1997; Dubey et al, 2003). As land plants (and other photosynthetic eukaryotes) seemingly use the cyanobacterial non-mevalonate pathway with little modification, we simply describe the seven enzymatic steps with no discrimination between cyanobacteria and land plants below (Fig. 1). In the first step, 1-deoxy-D-xylulose-5-phosphate (DXP) is synthesized from pyruvate and glyceraldehyde 3-phosphate by DXP synthase (DXS). DXP is converted into 2-C-methyl-D-erythritol 4-phosphate (MEP) by DXP reductoisomerase (DXR). Then MEP cytidylyltransferase (IspD) converts MEP into 4-(cytidine

5'-diphospho)-2-C-methyl-D-erythritol (CDP-ME), which is then altered to 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol (CDP-MEP) by 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE). 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF) alters CDP-MEP to 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (MEcPP), which is further converted to 1-hydroxy-2-methyl-2-butenyl 4-diphosphate (HMB-PP) by HMB-PP synthase (IspG). Finally, HMB-PP reductase (IspH) yields IPP from HMB-PP.

We here surveyed the transcripts encoding enzymes involved in three plastid-localized metabolic pathways 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 during the haptophyte/green algal endosymbiosis. Our systematic assessment revealed that the impact of EGT was different among the three pathways in *Karenia*, *Karlodinium* and *Lepidodinium*. All the three dinoflagellates appeared to be common in "vertically inherited (VI)-type" proteins, which were descended from the ancestral dinoflagellate bearing a peridinin-containing plastid, being completely (or nearly completely) eliminated from the Chl *a* biosynthesis. The haptophyte endosymbiont is the major source of the proteins involved in the Chl *a* biosynthesis in *Karenia*/*Karlodinium*, while the *Lepidodinium* pathway was reconstituted by proteins with phylogenetically diverse origins. Unlike the Chl *a* biosynthetic pathway, VI-type proteins were found to contribute to both C5 and non-mevalonate pathways in *Karenia*/*Karlodinium* and *Lepidodinium*. We finally propose a biological reason why the impact of EGT varied among the three pathways in the dinoflagellates studied here.

Materials and Methods

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

In this study, we 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 pathway (KEGG pathway, http://www.genome.jp/kegg/pathway.html). 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, K04036, K00218, K04037, K04038, K04039, K01662, K00099, K00991, K00919, K01770, K03526 and K03527. To generate amino acid sequence alignments covering phylogenetically diverse eukaryotic algae and bacteria, we surveyed the sequences of interest in the contigs generated from our in-house RNA-seq data of *Lepidodinium* and public databases including the Marine Microbial Eukaryote Transcriptome Sequencing Project (http://marinemicroeukaryotes.org), in which the contigs from the *Karenia* and *Karlodinium* RNA-seq data are available, by TBLASTN. In the first BLAST search, both bacterial and eukaryotic sequences registered in the KEGG pathway were used as the queries. Two evolutionarily distinct versions were known for PPOX ("*hemY*-type" and "*hemJ*-type"), and we used KOID K00231 as the query to identify the former type of PPOX. We 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), and the two distinct types of DVR sequences were separately subjected to the TBLASTN survey as the queries. The candidate sequences matched with the queries with *E*-values smaller than 10^{-20} in the first BLAST searches were retained, and conceptually translated into amino acid sequences. These 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, we

selected the candidate sequences matched to the proteins known to involved in the heme, Chl *a* and IPP biosyntheses for phylogenetic analyses described in the next section. The detailed information on the sequences identified in this study is summarized in Tables S1-3.

Phylogenetic analyses

In this study, we investigated the origins of individual proteins involved in the three pathways for the heme, Chl *a* and IPP biosyntheses in *Karenia*, *Karlodinium* and *Lepidodinium* by phylogenetic analyses with the maximum-likelihood (ML) method. For each protein of interest, we aligned the amino acid sequences retrieved from the public sequence databases described in the previous section by MAFFT v7.149b (Katoh & Standley, 2013), and the resultant alignments were manually refined. After exclusion of ambiguously aligned positions, 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). ML bootstrap values (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-Poisson model with a discrete Γ distribution with four categories (Lartillot & Philippe, 2004). Otherwise, the PhyloBayes analyses were done with the default settings. Four independent MCMC chains were run in parallel with >10,000 cycles until maxdiff values became smaller than 0.3. The details of the alignments and substitution models applied to the ML and Bayesian phylogenetic analyses are summarized in Table S4.

Results

Henceforth here, we designated the proteins inherited from the ancestral dinoflagellate as "vertically inherited-type" or "VI-type," (ii) those acquired from an endosymbiont (i.e. haptophyte and green alga, in *Karenia*/*Karlodinium* and *Lepidodinium*, respectively) as "endosymbiotically acquired-type" or "EA-type," and (iii) 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

We successfully identified all or most of the enzymes required for the heme biosynthesis in the transcriptomic data of *Karenia*, *Karlodinium* and *Lepidodinium*. In *Kareina* and *Karlodinium*, 8 out of the 9 enzymes were identified—only UROS and CPOX were missed in the former and latter, respectively. It is most likely that the UROS/CPOX transcripts were simply missed from the *Karenia*/*Karlodinium* cDNA library, as UROS is essential to convert hydroxymethylbilane to uroporphyrinogen III, and CPOX is indispensable to yield protoporphyrinogen IX from coproporphyrinogen III. In the following sections, we discuss the origins of individual proteins involved in the heme biosynthesis in the three dinoflagellates. Fifty-three out of the 59 transcripts (those encoding putative cytosolic proteins were excluded; see below) investigated here were found to bear extra amino acid residues at their N-termini, which are absent in the cyanobacterial homologues (N-terminal extensions). We regard these N-terminal extensions as plastid-localizing signals for complex plastids, and 20 of them were predicted to have a bipartite structure, which is composed of a portion rich in hydrophobic amino acids (signal peptide or SP) and the region

predicted to work as a transit peptide (TP-like region) by SignalP v.4.1 (Petersen et al. 2011). The details of the N-terminal extensions are summarized in Table S1.

Proteins with evolutionarily diverse origins comprise the *Karlodinium* and *Karenia* pathways

To convert glutamyl-tRNA to glutamate-1-semialdehyde, *Karlodinium* and *Karenia* possesses two versions of GTR. We consider "*Karenia*-1" and "*Karlodinium*-1" sequences as VI-type, as they were nested within a clade with those of peridinin-containing dinoflagellates, and the clade as a whole received a MLBP of 100% and a BPP of 0.99 in the GTR phylogeny (Fig. 2A). 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 BPP of 0.98 (Fig. 2A).

Aminolevulinic acid (ALA) is synthesized from glutamate-1-semialdehyde by GSAT. We identified a single version of GSAT in both *Karlodinium* and *Karenia*. 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. 2B). Nevertheless, the phylogenetic origin of either *Karlodinium* or *Karenia* GSAT sequence could not be pinpointed any further due to the lack of phylogenetic resolution in the GSAT phylogeny. The GSAT sequences of *Karenia* 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 *Karlodinium* GSAT sequence showed no specific affinity to any sequence examined here. Thus, we withhold

to conclude the precise origins of the *Karenia* and *Karlodinium* sequences in this study.

Synthesis of porphobilinogen from ALA likely catalyzed by a single ALAD homologue in both *Karenia* and *Karlodinium*. Both *Karenia* and *Karlodinium* ALAD sequences formed a robustly supported clade with those of peridinin-containing dinoflagellates (MLBP of 99% and BPP of 1.0; Fig. 2C), suggesting that the two sequences are of VI-type, which were vertically descended from the ancestral dinoflagellate.

We identified two versions of PBGD, which deaminates porphobilinogen to synthesize hydroxymethylbilane, in *Karenia* (*Karenia*-1 and 2). The PBGD phylogeny (Fig. 2D) 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, *Karenia* seemingly uses both VI-type and EA-type enzymes for hydroxymethylbilane synthesis. We identified four PBGD sequences in *Karlodinium* (*Karlodinium*-1-4), which clearly share a single ancestral sequence. The *Karlodinium* clade appeared to be distant from the clade comprising the sequences of peridinin-containing dinoflagellates and the *Karenia*-1 sequence, suggesting that the *Karlodinium* sequences are not of VI-type. The *Karlodinium* 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. 2D). The statistical support for the particular node is insufficient to conclude or exclude the haptophyte origin of the *Karlodinium* sequences with confidence. Thus, we determined to leave the origin of the *Karlodinium* sequences uncertain.

We identified two distinct versions of UROS in *Karlodinium*, but none in *Karenia*. One of the two versions of the *Karlodinium* UROS (*Karlodinium*-1) branched at the base of the clade of the sequences of peridinin-containing dinoflagellates, albeit the affinity between the *Karlodinium*-1 and other dinoflagellate sequences was not strongly supported (MLBP of 51% and

BPP <0.50; Fig. 2E). Likewise, it is difficult to ascertain whether the haptophyte origin of the *Karlodinium*-1 sequence due to the low phylogenetic resolution. Thus, we decided to leave the origin of the *Karlodinium*-1 sequence uncertain. The second *Karlodinium* sequence (*Karlodinium*-2) tied with the sequence of a red alga *Rhodosorus marinus* with a MLBP of 94% and a BPP of 0.76 (Fig. 2E), suggesting that the *Karlodinium*-2 sequence was acquired from a red alga (i.e. LA-type).

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. 2F), 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. 2F). We here identified five and four UROD sequences in *Karenia* and *Karlodinium*, respectively. Among the five sequences identified in *Karenia*, 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. 2F). H1 clade including the *Karenia*-1 and 2 sequences received a MLBP of 100% and a BPP 0.99. 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 0.98. The "*Karenia*-3" sequence grouped with the sequence of a euglenozoan *Eutreptiella gymnastica* with a MLBP of 100% and a BPP of 0.99, suggesting that *Karenia* acquired a UROD gene from a euglenozoan (i.e. LA-type). 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 0.99. We also assessed the origins of four *Karlodinium* sequences (*Karlodinium*-1-4) based on the UROD

phylogeny (Fig. 2F). The *Karlodinium*-1 and 2 sequences were nested within H1 clade, and the *Karlodinium*-4 sequence were placed within H3 clade. Thus, we conclude that the three UROD sequences are of EA-type. The *Karlodinium*-3, one of the *Lepidodinium* sequences (see below), and diatom sequences formed a robust clade (MLBP of 100% and BPP of 0.99; highlighted by an arrowhead in Fig. 2F), 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. We identified a single version of CPOX in *Karenia*, but none in *Karlodinium*. The CPOX phylogeny (Fig. 2G) recovered a clade comprising the *Karenia* sequence and the sequences of peridinin-containing dinoflagellates with a MLBP of 98% and a BPP of 0.99. Thus, *Karenia* uses a VI-type CPOX descended from the ancestral dinoflagellate.

Protoporphyrinogen IX is further oxidized by PPOX to obtain protoporphyrin IX. The PPOX sequences of peridinin-containing dinoflagellates were separated into two distinct clades labeled as "D1" and "D2" (Fig. 2H), 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. 2H). We identified two versions of PPOX in *Karenia* (*Karenia*-1 and 2), while a single version in *Karlodinium*. The *Karenia*-1 sequence fell into D1 clade (Fig. 2H), suggesting that this PPOX sequence is of VI-type, which was descended from the ancestral dinoflagellate. On the other hand, the PPOX phylogeny united the *Karenia*-2 and a single PPOX of *Karlodinium* with the sequences of stramenopiles, haptophytes and *Lepidodinium* with a MLBP of 100% and a BPP of 0.99 (highlighted by a double-arrowhead in Fig. 2H). As the bipartitions within the clade were

principally unresolved (Fig. 2H), we cannot exclude the possibility of the *Karenia*-2 and *Karlodinium*-1 sequences group with the haptophyte sequences in this clade. 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, we decide to leave the origins of the *Karenia*-2 and *Karlodinium*-1 sequences uncertain.

In the last step in the heme biosynthesis, FeCH converts protoporphyrin IX to protoheme. None of the FeCH sequences identified in *Karenia* and *Karlodinium* appeared to be of VI-type. *Karenia* possesses three distinct versions of FeCH (*Karenia*-1-3). The FeCH phylogeny (Fig. 2I) united the *Karenia*-1 sequence with one of four versions of FeCH in *Lepidodinium* (*Lepidodinium*-3) with a MLBP of 100% and a BPP of 0.99, and this union was then connected specifically to γ-proteobacterial sequences with a MLBP of 75% and a BPP of 0.88 (highlighted by an arrowhead in Fig. 2I). This subtree can be explained by two sequential gene transfer events, namely the first gene transfer from a γ-proteobacterium to either *Lepidodinium* or *Karenia*, and the second one between the two dinoflagellates. Consequently, the *Karenia*-1 and *Lepidodinium*-3 sequences can be traced back to the bacterial sequence (i.e. LA-type). The *Karenia*-2 sequence is unlikely to be of VI-type, as the sequences of peridinin-containing dinoflagellates (and one of the *Lepidodinium* sequences) were united with a MLBP of 100% and a BPP of 0.98 (Fig. 2I). The *Karenia*-2 sequence, which showed no clear affinity to the haptophyte sequences, is unlikely to be of EA-type (Fig. 2I). Thus, we conclude that the *Karenia*-2 sequence is 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 *Karlodinium* with the cyanobacterial sequences, and their monophyly was supported by a MLBP of 76% and a BPP of 0.51 (highlighted by a double-arrowhead in Fig. 2I). We conclude that the two dinoflagellate sequences are of cyanobacterial origin (i.e. LA-type).

In *Karenia* and *Karlodinium*, the heme biosynthetic pathway appeared to be composed of (i) VI-type, (ii) EA-type and (iii) LA-type proteins. Pioneering studies have documented such evolutionarily chimeric plastid proteomes in dinoflagellates bearing non-canonical plastids, but primarily emphasized the presence of EA-type proteins to demonstrate the integration of a haptophyte endosymbiont into the dinoflagellate cell as the plastid. Nevertheless, our systematic evaluation on individual proteins involved in the heme biosynthesis identified EA-type proteins only in three out of the 9 steps of this particular pathway—the steps catalyzed by (i) GTR in *Karenia* and *Karlodinium*, (ii) PBGD in *Karlodinium*, and (iii) UROD in *Karenia* and *Karlodinium*. In other word, the results presented above suggest that VI-type proteins still play major roles in the heme biosynthesis in the two kareniacean species.

Little impact of endosymbiotic gene transfer on the *Lepidodinium* pathway

We identified three versions of GTR in *Lepidodinium* (*Lepidodinium*-1-3). 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 0.99 (Fig. 2A). Thus, we conclude that the *Lepidodinium*-1 sequence is of VI-type. The *Lepidodinium* 2 and 3 sequences were tied to each other with a MLBP of 100% and a BPP of 0.99, and this clade was connected with the "dinoflagellate" clade described above (Fig. 2A). 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. 2A), suggesting that the *Lepidodinium* 2 and 3 sequences were unlikely of VI-type. The GTR sequences of land plants and green algae (plus a euglenozoan *Eutreptiella gymnastica*) formed a clade supported by a MLBP of 100% and a BPP of 0.94, and appeared to be distantly related to any *Lepidodinium* sequences (Fig.

2A). Thus, we can exclude the green algal (endosymbiont) origin of the *Lepidodinium* 2 and 3 sequences, and classified the two sequences as LA-type, although the organism donated a GTR gene to *Lepidodinium* remains unclear.

We failed to classify two out of the three versions of GSAT identified in *Lepidodinium*. In the GSAT phylogeny (Fig. 2B), the "*Lepidodinium*-1" and "*Lepidodinium*-2" sequences fell separately into the cluster of the sequences of peridinin-containing dinoflagellates and *Karenia*, but this clade as a whole received no significant statistical support. Thus, we left the origin of the *Lepidodinium*-1 sequence uncertain in this study. On the other hand, the "*Lepidodinium*-3" sequence was excluded from the sequences of diverse photosynthetic eukaryotes and cyanobacteria with a MLBP of 100% and a BPP of 0.99, and connected to the sequences of *Streptomyces*, *Mycobacterium* and *Corynebacterium* with a MLBP of 71% and a BPP of 0.51 (highlighted by an arrowhead in Fig. 2B). This tree topology suggests that the *Lepidodinium*-3 sequence was acquired from a bacterium (i.e. LA-type), albeit we cannot pinpoint the bacterium donated a GSAT gene to *Lepidodinium*.

Two versions of ALAD were identified in *Lepidodinium*. The sequences of peridinin-containing dinoflagellates, one of the two *Lepidodinium* sequences (*Lepidodinium*-1) and the sequences of *Karenia* and *Karlodinium* clustered with a MLBP of 99% in the ALAD phylogeny (Fig. 2C). Thus, we 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. 2C). Thus, we 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 *Lepidodinium*. The PBGD phylogeny (Fig. 2D) united the two *Lepidodinium* sequences together with a MLBP of

100% and a BPP of 0.99, but the *Lepidodinium* 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. 2E) recovered no clear affinity of the *Lepidodinium* sequence to other sequences including those of peridinin-containing dinoflagellates or green algae/land plants. Thus, *Lepidodinium* likely uses LA-type PBGD and UROS, but their precise origins remain uncertain.

We identified 7 versions of UROD in *Lepidodinium*, and 6 of them showed clear affinities to the sequences of peridinin-containing dinoflagellates. In the UROD phylogeny (Fig. 2F), the sequences of peridinin-containing dinoflagellates formed three distinct clades (D1-3 clades), each of these clades enclosed at least one of the *Lepidodinium* sequences, 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 *Karlodinium*) formed a clade with a MLBP of 100% and a BPP of 0.99 (highlighted by an arrowhead in Fig. 2F), suggesting that this version was acquired from a diatom (i.e. LA-type).

Three versions of CPOX were identified in *Lepidodinium*, but none of them was of VI-type or EA-type. The CPOX phylogeny 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. 2G). This enzyme most likely bears no N-terminal extension (Table S1). Altogether, we 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,

493 while their N-terminal regions are distinct from each other (data not shown). The two
494 *Lepidodinium* sequences were united robustly with the diatom sequences in the CPOX phylogeny
495 (MLBP of 86% and BPP of 0.98; highlighted by a double-arrowhead in Fig. 2G). Thus, the
496 ancestral CPOX of the two *Lepidodinium* sequences was most likely acquired from a diatom (i.e.
497 LA-type).

498 We conclude that *Lepidodinium* possesses two distinct VI-type and a single LA-type
499 PPOX. The PPOX sequences of peridinin-containing dinoflagellates were split into two distinct
500 clades (D1 and D2), and the two clades received strong statistical support from both ML bootstrap
501 and Bayesian analyses (Fig. 2H). The "*Lepidodinium*-1" and "*Lepidodinium*-3" sequences were
502 included in D1 and D2 clades, respectively. As the PPOX sequences (including the *Lepidodinium*-
503 3 sequence) in D2 clade can be considered as the cytosolic version, we did not discuss the
504 *Lepidodinium*-3 sequence further. The PPOX phylogeny recovered a robust clade comprising the
505 "*Lepidodinium*-2" sequence and, the sequences of haptophytes, stramenopiles, *Karlodinium* and
506 *Karenia* (MLBP of 100% and BPP of 0.99; highlighted by a double-arrowhead in Fig. 2H). Thus,
507 the *Lepidodinium*-2 sequence was acquired from an organism distantly related to dinoflagellates
508 or green algae/land plants (i.e. LA-type).

509 In *Lepidodinium*, we identified four versions of FeCH. The "*Lepidodinium*-1" sequence
510 is of VI-type, as this sequence apparently shared the origin with the sequences of peridinin-
511 containing dinoflagellates, and their monophyly was supported with a MLBP of 100% and a BPP
512 of 0.98 (Fig. 2I). On the other hand, we consider the rest of the *Lepidodinium* sequences as LA-
513 type, as the "*Lepidodinium*-2," "*Lepidodinium*-3" and "*Lepidodinium*-4" sequences appeared to
514 be distantly related to the dinoflagellate clade described above or the green algae/land plant
515 sequences in the FeCH phylogeny (Fig. 2I). The precise positions of the *Lepidodinium*-2 and
516 *Lepidodinium*-4 sequences were unresolved, and it remains unclear how *Lepidodinium* acquired

517 the two versions of FeCH. On the other hand, the FeCH phylogeny united the "*Lepidodinium*-3"
518 and *Karenia*-1 sequences together (MLBP of 100% and BPP of 0.99), and this dinoflagellate clade
519 was then connected to two γ-proteobacterial sequences with a MLBP of 75% and a BPP of 0.88
520 (highlighted by an arrowhead in Fig. 2I). We have already proposed the two scenarios for the
521 origin of the *Lepidodinium*-3 and *Karenia*-1 sequences, in which two lateral gene transfers were
522 invoked (see the previous section for the details).

523 Minge et al. (2010) reported nucleus-encoded genes encoding green algal proteins
524 involved in the plastid functions, indicating that the ancestral *Lepidodinium* has genetically
525 integrated the green algal endosymbiont as a plastid. Nevertheless, it was surprising that no gene
526 from the green algal endosymbiont was detected in the heme biosynthetic pathway in
527 *Lepidodinium*. Instead, genes transferred from organisms related to neither host (dinoflagellate)
528 nor endosymbiont (green alga) largely contributed to the *Lepidodinium* pathway. The impact of
529 lateral gene transfer on the heme biosynthesis is the most prominent in the steps catalyzed by
530 PBGD, UROS and CPOX, in which only LA-type proteins were identified.

531

532 **Chl *a* biosynthetic pathway**

533

534 We surveyed the transcripts encoding enzymes involved in the Chl *a* biosynthesis in *Karenia*,
535 *Karlodinium* and *Lepidodinium*, and assessed their origins individually. Overall, all the enzymes
536 required to synthesize Chl *a*, except MgPME cyclase, was retrieved from the transcriptomic data
537 from the three dinoflagellates. To our knowledge, no sign for MgPME cyclase, which converts
538 Mg-protoporphyrin IX monomethyl ester to divinyl protochlorophyllide, was detected in
539 peridinin-containing dinoflagellate, diatoms, cryptophytes or haptophytes. Although not described
540 in detail, we 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. We 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 predicted to have a bipartite structure; see Table S2 for the details).

Large impact of EGT on the *Karenia* and *Karlodinium* pathways

Mg chelatase (MgCH), which comprises three subunits ChlD, ChlH and ChlI, inserts Mg^{2+} to protoporphyrin IX. In *Karlodinium*, ChlI is plastid-encoded (Gabrielsen et al. 2011) and the rest of the subunits were nucleus-encoded (see below). Although no plastid genome data is available for *Karenia*, we identified the ChlI sequence in the transcriptomic data of this species (contig No., 0173787962). According to the intimate organismal relationship between *Karlodinium* and *Karenia*, we considered that the *Karenia* ChlI sequence was transcribed from the plastid genome. As the current study focuses on nucleus-encoded proteins involved in plastid metabolisms, we stopped examining the origin and evolution of ChlI in the two kareniacean species (and *Lepidodinium*; see below) any further.

We here examine the evolutionary origins of two nucleus-encoded subunits of MgCH, ChlH and ChlD, in *Karenia* and *Karlodinium*. A single version of ChlD was identified in *Karlodinium* and *Karenia*. The ChlD phylogeny (Fig. 3A) recovered a clade of the sequences of haptophytes, *Karlodinium* and *Karenia* with full statistical support, suggesting that the kareniacean ChlD sequences are of EA-type. Both *Karenia* and *Karlodinium* possess two distinct versions of ChlH (*Karenia*-1 and 2, and *Karlodinium*-1 and 2). ChlH sequences can be split in to two distinct clades, “ChlH-1” and “ChlH-2,” as a previous study reported (Lohr et al, 2005, Fig. 3B). ChlH-1

565 sequences are ubiquitously distributed in photosynthetic organisms, while ChlH-2 sequences have
566 been found in restricted lineages. The sequences of green algae/land plants formed two distinct
567 clades (Gp1 and 2 clades). In the ChlH phylogeny (Fig. 3B), the haptophyte ChlH-1 sequences,
568 the *Karenia*-1 sequence and *Karlodinium*-1 sequences formed a clade supported with a MLBP of
569 98% and a BPP of 0.94. In contrast, the *Karenia*-2 and *Karlodinium*-2 sequences were nested
570 within the Gp2 clade containing ChlH-2 sequences of green algae and a euglenid, and their
571 monophyly received a MLBP of 81% and a BPP of 0.74 (highlighted by an arrowhead in Fig. 3B).
572 Thus, we conclude that *Karenia* and *Karlodinium* possess ChlH-1 sequences acquired from the
573 haptophyte endosymbiont (i.e. EA-type), while their ChlH-2 sequences are of green algal origin
574 (i.e. LA-type).

575 MgPMT converts Mg-protoporphyrin IX to Mg protoporphyrin IX monomethyl ester.
576 The MgPMT sequences of *Karlodinium* and *Karenia* grouped with the haptophyte sequences
577 (except the one of *Pavlova*), and their monophyly was supported by a MLBP of 67% and a BPP
578 of 0.98 (highlighted by an arrowhead in Fig. 3C). The two kareniacean species most likely use the
579 MgPMT acquired from the haptophyte endosymbiont (i.e. EA-type).

580 No MgPME cyclase has been identified in any dinoflagellates regardless of plastid-type,
581 and we could not examine the origin and evolution of this enzyme (see above). However, DVR,
582 which recognizes the product of MgPME cyclase (divinyl protochlorophyllide) as the substrate
583 and generate protochlorophyllide, were identified in both dinoflagellates bearing peridinin and
584 those bearing non-canonical plastids. We identified both N-DVR and F-DVR sequences in
585 *Karenia*, while only N-DVR sequence was found in *Karlodinium*. The F-DVR phylogeny (Fig.
586 3D) recovered the clade of the *Karenia* and haptophyte sequences with full statistical support,
587 suggesting that this sequence was acquired from the haptophyte endosymbiont (i.e. EA-type).

588 The N-DVR phylogeny (Fig. 3E) united the *Karlodinium* and *Karenia* sequences together

589 with a MLBP of 98% and a BPP of 0.99, and the kareniacean clade showed any phylogenetic
590 affinity to neither haptophyte sequences nor other dinoflagellate sequences. Instead, the
591 kareniacean clade grouped with the sequences of stramenopiles, haptophytes, and two chromerids
592 (*Chromera* and *Vitrella*) supported with a MLBP of 100% and a BPP of 1.0 (highlighted by an
593 arrowhead in Fig. 3E). In this large clade, the affinity between the kareniacean clade and
594 haptophyte sequences was not positively supported. We here propose that multiple stramenopile
595 lineages donated N-DVR genes separately to haptophytes, kareniacean species, and chromerids
596 (i.e. LA-type).

597 In land plants, conversion of protochlorophyllide to chlorophyllide *a* is catalyzed by the
598 light-dependent and/or light-independent forms of POR. The light-dependent POR is nucleus-
599 encoded, while the light-independent form comprises three plastid-encoded subunits (ChlB, ChlL
600 and ChlN). No gene for light-independent POR was found in the plastid genome of *Karlodinium*
601 (Gabrielsen et al. 2011), implying that kareniacean species lack the light-independent version. We
602 identified two and three distinct versions of the light-dependent POR in *Karenia* and *Karlodinium*,
603 respectively, as demonstrated in Hunsperger et al, (2015). In the POR phylogeny (Fig. 3F), the
604 *Karenia*-1 and *Karlodinium*-1 sequences grouped with the sequences of stramenopiles,
605 cryptophytes and haptophytes, and their monophyly was supported by a MLBP of 95% and a BPP
606 of 0.98 (highlighted by an arrowhead in Fig. 3F). The *Karenia*-1 and *Karlodinium*-1 sequences
607 cannot be of VI-type, as the two sequences are distantly related to other dinoflagellate sequences
608 included in the alignment. However, it is difficult to classify the *Karenia*-1 and *Karlodinium*-1
609 sequences into EA-type or LA-type, as the relationship between the two dinoflagellate sequences
610 and the haptophyte sequences was unresolved in the particular clade. Thus, we leave the origins
611 of the two sequences uncertain in this study. The *Karenia*-2, *Karlodinium*-2 and *Karlodinium*-3
612 sequences robustly grouped together within the haptophyte sequences, and this "haptophyte" clade

613 received a MLBP of 98% and a BPP of 0.99 (Fig. 3F). Thus, we conclude that these sequences
614 were acquired from the haptophyte endosymbiont (i.e. EA-type).

615 The final step of the Chl *a* biosynthesis is catalyzed by Chl synthase (CS). A single version
616 of CS was identified in *Karlodinium* and *Karenia*. The two kareniacean sequences were placed
617 within the haptophyte clade in the CS phylogeny, and the "haptophyte" clade as a whole was
618 supported by a MLBP of 98% and a BPP of 0.99 (Fig. 3G). Thus, both *Karlodinium* and *Karenia*
619 sequences are considered as EA-type, acquired from the haptophyte endosymbiont.

620 The phylogenetic analyses described above revealed that EA-type proteins, which were
621 acquired from the haptophyte endosymbiont resided in the ancestral kareniacean species, operate
622 in all the steps converting protoporphyrin IX to Chl *a* in *Karlodinium* and/or *Karenia* (except the
623 step catalyzed by MgPME cyclase; see above). In addition, the common ancestor of *Karlodinium*
624 and *Karenia* should have possessed LA-type ChlH-2, POR and N-DVR, which were acquired from
625 phylogenetically diverse eukaryotes distantly related to dinoflagellates or haptophytes. Overall,
626 the phylogenetic analyses described above strongly suggest that the Chl *a* biosynthetic pathway in
627 the ancestral kareniacean species has almost entirely reconstructed by genetic materials acquired
628 from the haptophyte endosymbiont.

629

630 Genetic influx from phylogenetically diverse organisms shaped the *Lepidodinium* pathway

631

632 As discussed in the previous section, the Chl *a* biosynthetic pathway in *Karenia* and *Karlodinium*
633 appeared to be shaped by the genes transferred from the endosymbiont (i.e. a haptophyte in the
634 above systems). Curiously, this is not the case for the same pathway in *Lepidodinium*, of which
635 plastid was derived from a pedinophyte green alga. Note that we present no result from the
636 *Lepidodinium* ChlI, which turned out to be plastid-encoded (Kamikawa et al. 2015a). We identified

637 8 proteins involved in the Chl *a* biosynthetic pathway in *Lepidodinium*, and assess their
638 phylogenetic origins individually (Figs. 3A-G). Among the 8 proteins examined here, we conclude
639 MgPMT as a sole EA-type protein among those involved in the *Lepidodinium* pathway. The
640 MgPMT phylogeny (Fig. 3C) placed the *Lepidodinium* sequence within a radiation of the
641 sequences of green algae, land plants and chlorarachniophytes, and their monophyly was supported
642 by a MLBP of 88% and a BPP of 0.96.

643 Our surveys and phylogenetic analyses revealed that VI-type proteins were almost
644 entirely eliminated from the *Lepidodinium* pathway. We identified only one of the two versions of
645 POR (*Lepidodinium*-1) as VI-type. The POR phylogeny (Fig. 3F) recovered two distinct clades of
646 the sequences of peridinin-containing dinoflagellates (D1 and D2 clades), and placed the
647 *Lepidodinium*-1 sequence within D1 clade. D1 clade containing the *Lepidodinium*-1 sequence as
648 a whole received a MLBP of 97% and a BPP of 0.99.

649 We could not clarify the origin of the *Lepidodinium* N-DVR sequence. In the N-DVR
650 phylogeny (Fig. 3E), the *Lepidodinium* sequence grouped with an euglenid *Eutreptiella* with a
651 MLBP of 35 % and a BPP of 0.57, and was excluded from the clade of the sequences of peridinin-
652 containing dinoflagellates, of which monophyly received full statistical support. Thus, the
653 *Lepidodinium* sequence cannot be of VI-type. However, it is difficult to pursue the origin of the
654 *Lepidodinium* sequence any further, as the N-DVR phylogeny failed to exclude a potential affinity
655 between the sequence of interest and the green algal/land plant sequences (Fig. 3E).

656 We classified ChlD, ChlH, one of the two versions of POR (*Lepidodinium*-2), F-DVR
657 and CS into LA-type. In the ChlD phylogeny, the *Lepidodinium* sequence appeared to be excluded
658 from the sequences of peridinin-containing dinoflagellates and those of the green algal/land plant
659 sequences (Fig. 3A), suggesting that the *Lepidodinium* sequence cannot be of VI-type or EA-type.
660 Instead, the *Lepidodinium* sequences, as well as those of chromerids, cryptophytes and

661 *Eutreptiella*, were placed within the radiation of the stramenopile sequences, and their monophyly
662 was supported by a MLBP of 96% and a BPP of 0.99. This tree topology prompts us to propose
663 that *Lepidodinium* ChlD was acquired from a stramenopile (i.e. LA-type). To our surprise, ChlH,
664 F-DVR, one of the two versions of POR (*Lepidodinium*-2) and CS appeared to be acquired
665 commonly from chlorarachniophytes (Figs. 3B, 3D, 3F and 3G). In each of the ChlH, POR, F-
666 DVR and CS phylogenies, the *Lepidodinium* sequence showed an intimate affinity to the
667 chlorarachniophyte sequences, and their monophyly was supported by MLBPs greater than 92%
668 and by BPPs >0.85.

669 The phylogenetic analyses described above revealed that only one out of the five steps in
670 this pathway appeared to be catalyzed by an EA-type protein, suggesting that EGT was much less
671 significant in the *Lepidodinium* pathway than the kareniacean pathway (see above). Instead,
672 chlorarachniophytes seemingly donated the genes encoding the proteins involved in four out of the
673 five steps in the *Lepidodinium* pathway. The phylogenetically chimeric nature of the Chl *a*
674 biosynthesis in this species is represented well by MgCH comprising three subunits bearing
675 distinct evolutionary backgrounds. (i) ChlI is principally of green alga (plastid-encoded), (ii) ChlH
676 was acquired from a chlorarachniophyte, and (iii) ChlD acquired from a stramenopile.

677

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

679

680 The origin of all the enzymes involved in the non-mevalonate pathway of *Karlodinium* and
681 *Karenia* were investigated carefully in Bentlage et al. (2016). On the other hand, the entire picture
682 of the *Lepidodinium* pathway remain to be completed, leaving 5 out the 7 enzymes involved in this
683 pathway unidentified (Minge et al. 2010). We successfully identified all the enzymes involved in
684 the non-mevalonate pathway in *Lepidodinium* (see below). In this section, we mainly examined

685 the origins of individual enzymes involved in the non-mevalonate pathway of *Lepidodinium*,
686 coupled with a brief overview of the same pathway of the two kareniacean species. Thirty-one out
687 of the 32 transcripts investigated here were found to possess N-terminal extensions, which likely
688 work as plastid-localizing signals (Note that 12 sequences were predicted to have a bipartite
689 structure; see Table S3 for the details).

690 For the step synthesizing DXP from pyruvate and glyceraldehyde 3-phosphate,
691 *Lepidodinium* was found to possess two versions of DXS (*Lepidodinium*-1 and 2). The DXS
692 phylogeny robustly grouped the *Lepidodinium*-1 and 2 sequences with those of peridinin-
693 containing dinoflagellates and *Karenia* (Fig. 4A). The sequence of *Karlodinium* was found to be
694 remote from the dinoflagellate clade, and showed an affinity to the haptophyte sequences (Fig.
695 4A). The clade of the *Karlodinium* sequence and haptophytes received a MLBP of 63% and a BPP
696 of 0.56 (if the *Pavlova* sequence was excluded, the "*Karlodinium* + haptophyte" clade was
697 supported by a MLBP of 98% and a BPP of 1.0). Thus, we conclude that the DXS sequences of
698 *Lepidodinium* and *Karenia* were vertically inherited from the ancestral dinoflagellate (i.e. VI-
699 type), while that of *Karlodinium* was acquired from the haptophyte endosymbiont (i.e. EA-type).

700 The conversion of DXP to MEP is catalyzed by DXR. Minge et al. (2010) detected a
701 partial sequence of a VI-type DXR (GenBank accession number CCC15090). From our
702 transcriptome data, two versions of DXR were identified in *Lepidodinium* (*Lepidodinium*-1 and 2;
703 the former corresponds to the previously reported DXR sequence). The DXR phylogeny (Fig. 4B)
704 reconstructed a robust monophyly of the two versions of DXR in *Lepidodinium*, the sequences of
705 peridinin-containing dinoflagellates, and those of *Karenia* and *Karlodinium* (MLBP of 100% and
706 BPP of 0.99), suggesting that *Lepidodinium* and the two kareniacean species use VI-type proteins
707 for this reaction.

708 *Lepidodinium* was found to possess both LA-type and VI-type versions of IspD

(*Lepdidodinium*-1 and 2) to convert MEP into CDP-ME. The *Lepidodiniu*-1 sequence showed a specific affinity to the sequence of a stramenopile *Ochromonas* sp. with a MLBP of 100% and a BPP of 0.98 (Fig. 4C), indicating that *Lepidodinium* acquired this version from a stramenopile (i.e. LA-type). In contrast, the *Lepidodinium*-2 sequence, together with the *Karlodinium* sequence, were considered as VI-type, as they formed a clade with those of peridinin-containing dinoflagellates (MLBP of 81% and BPP <0.50). The IspD sequence of *Karenia* appeared to be nested within the clade of green algae/land plants and chlorarachniophytes, and being distantly related to the dinoflagellate clade (including the *Lepidodinium*-2 and *Karlodinium* sequences) or the haptophyte clade (Fig. 4C). The position of *Karenia* 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 IspE (GenBank accession number CCC15094), which phosphorylates CDP-ME to CDP-MEP, in *Lepidodinium*. In this study, we 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. 4D) recovered a monophyly of the two *Lepidodinium* sequences, the sequences of peridinin-containing dinoflagellates, the *Karlodinium* sequence, and one of the two *Karenia* sequences (*Karenia*-2), which was supported by a MLBP of 88% and a BPP of 0.98. Thus, *Lepidodinium*, *Karlodinium* and *Karenia* possess VI-type versions of IspE. In addition, *Karenia* possesses an EA-type IspE (*Karenia*-1), which was united with the haptophyte sequences with a MLBP of 100% and a BPP of 0.99.

The ancestral dinoflagellate likely possessed two versions of IspF, which converts CDP-MEP to MEcPP. In the IspF phylogeny (Fig. 4E), 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 1.0, respectively. D1 clade appeared to contain one of the two *Lepidodinium*

733 sequences (*Lepidodinium*-1), as well as the sequence of *Karlodinium* and one of the two *Karenia*
734 sequences (*Karenia*-2). The other version of *Lepidodinium* (*Lepidodinium*-2) was nested within
735 D2 clade. Thus, we conclude that *Lepidodinium*, *Karenia* and *Karlodinium* possess VI-type
736 versions of IspF. As reported in Bentlage et al. (2016), *Karenia* possess an additional IspF
737 sequence (*Karenia*-1) with a phylogenetic affinity to the haptophyte sequences, suggesting that
738 this version is of EA-type.

739 The origin and evolution of IspG, which synthesize HMB-PP from MEcPP, seems
740 straightforward in dinoflagellates. We phylogenetically analyzed two versions of IspG in
741 *Lepidodinium* (*Lepidodinium*-1 and 2) identified in this study, together with the *Karenia* and
742 *Karlodinium* sequences. In the IspG phylogeny (Fig. 4F), the aforementioned dinoflagellate
743 sequences tightly clustered with the sequences of peridinin-containing dinoflagellates (MLBP of
744 98% and BPP of 0.92). Thus, we concluded that *Karenia*, *Karlodinium* and *Lepidodinium* uses VI-
745 type enzymes to synthesize MEcPP.

746 The last step of the non-mevalonate pathway is catalyzed by IspH to yield IPP from HMG-
747 PP. We identified two versions of IspH in *Lepidodinium* (*Lepidodinium*-1 and 2). The IspH
748 phylogeny (Fig. 4G) recovered a clade with the sequences of peridinin-containing dinoflagellates,
749 apicomplexan parasites, chromerids, the two versions of *Lepidodinium* and one of the three
750 versions of *Karenia* (*Karenia*-3), which received a MLBP of 68% and a BPP <0.50. This tree
751 topology suggests that the *Lepidodinium* and *Karenia* IspH sequences can be trace back to that in
752 the common ancestor of dinoflagellates, apicomplexan parasites and chromerids (the ancestral
753 myzozoan). When we reanalyzed the same alignment after the exclusion of the rapidly evolving
754 apicomplexan and chromerid sequences, the *Lepidodinium*-1 and 2, and *Karenia*-3 sequences and
755 those of peridinin-containing dinoflagellates grouped together with a MLBP of 100% and a BPP
756 of 0.99 (Fig. 4G). We conclude that the two versions of IspH in *Lepidodinium* are of VI-type,

757 descended from the ancestral dinoflagellate (or even from the ancestral myzozoan). *Karenia* were
758 found to possess two additional versions of IspH (*Karenia*-1 and *Karenia*-2). The IspH phylogeny
759 (Fig. 4G) united the *Karenia*-1 sequence and a single IspH sequence of *Karlodinium* with the
760 sequences of haptophytes with a MLBP of 90% and a BPP of 0.71. The *Karenia*-2 sequence was
761 connected to the sequence of a stramenopile *Ochromonas* sp. with a MLBP of 88% and a BPP of
762 0.98 (Fig. 4G). Thus, as discussed in Bentlage et al. (2016), *Karenia* uses VI-type (*Karenia*-3),
763 EA-type (*Karenia*-1) and LA-type (*Karenia*-2) enzymes to yield IPP, while *Karlodinium* possesses
764 a single, EA-type version.

765 As Bentlage et al. (2016) demonastrated, the non-mevalonate pathways in *Karenia* and
766 *Karlodinium* are evolutionary hybrids of VI-type enzymes inherited vertically from the ancestral
767 dinoflagellate and EA-type enzymes acquired from the haptophyte endosymbiont, except two LA-
768 type enzymes identified in *Karenia*—IspD and one of the three IspH versions. In sharp contrast,
769 the same pathway in *Lepidodinium* appeared to be dominated by VI-type proteins, except a single
770 LA-type protein (one of the two IspD versions). Thus, these observations clearly suggest that the
771 gene transfer from a green algal endosymbiont had little impact on the non-mevalonate pathway
772 in *Lepidodinium*.

773

774 **Discussion**

775

776 In this study, we surveyed the nucleus-encoded, plastid-localized proteins involved in the heme,
777 Chl *a*, and IPP biosyntheses in the transcriptomic data from *Karenia*, *Karlodinium* and
778 *Lepidodinium*. By assessing the phylogenetic origins of the individual proteins of interest
779 rigorously, we successfully revealed how the haptophyte and green algal endosymbioses altered
780 the aforementioned pathways in the three dinoflagellates.

781

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

783 The common ancestor of *Karenia* and *Karlodinium* discarded the canonical (peridinin-
784 containing) plastid and established the current plastid derived from a haptophyte-endosymbiont.
785 As anticipated from the plastid evolution in kareniacean species, the *Karenia* and *Karlodinium*
786 pathways investigated here appeared to be composed of three types of proteins, namely (i) proteins
787 acquired from the haptophyte endosymbiont (EA-type), (ii) proteins descended from the ancestral
788 dinoflagellate (VI-type), and (iii) proteins acquired from organisms distantly related to the host or
789 endosymbiont (LA-type) (Patron et al. 2006; Nosenko et al. 2006; Hunsperger et al, 2015; Bentlage
790 et al. 2016). Nevertheless, the impact of the genetic influx from the haptophyte endosymbiont was
791 different among the three pathways in *Karenia* and *Karlodinium* (Fig. 5). In the two kareniacean
792 species, EA-type proteins, together with a few LA-type proteins, found to dominate the Chl *a*
793 biosynthesis, albeit no VI-type protein was detected. In sharp contrast, VI-type proteins persist in
794 5-6 out of the 7 steps required for the non-mevalonate pathway for the IPP biosynthesis, albeit the
795 contributions of EA-type and LA-type proteins may not be negligible. The evolutionary chimerism
796 is most advanced in the heme biosynthesis, in which all the three protein types are identified (Fig.
797 5).

798 Based on the difference in degree of evolutionary chimerism among the three pathways
799 discussed in the previous section, we here explore the early evolution of kareniacean species.
800 Patron et al. (2006) and Nosenko et al. (2006) hypothesized that the ancestral kareniacean species
801 possessed a non-photosynthetic plastid prior to the haptophyte endosymbiosis, although both
802 studies assessed a restricted number of plastid-localized proteins. The above proposal is plausible,
803 as secondarily non-photosynthetic eukaryotes often possess plastids with no photosynthetic
804 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; Janouskovec et al, 2017). Noteworthy, *Karenia* and *Karlodinium* are closely related to a kleptoplastic species found in the Ross Sea, Antarctica (Gast et al, 2006, 2007). The intimate phylogenetic affinity between the species bearing the haptophyte-derived non-canonical plastids and that leading a kleptoplastic lifestyle lends an additional support for the non-photosynthetic nature of their common ancestor. The hypothesis for the non-photosynthetic nature in the ancestral kareniacian species can explain well the elimination of VI-type proteins from the Chl *a* biosynthesis in both *Karenia* and *Karlodinium* (Fig. 5). 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. 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 *Karenia* and *Karlodinium* 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 we observed both EA- and LA-type proteins in the current pathways (Fig. 5).

We 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 *Karenia* and *Karlodinium*. For instance, the contribution of VI-type proteins to the heme biosynthesis seemingly differs between *Karenia* and *Karlodinium*, which retain 6 and 3 VI-type proteins for the 9 steps in the heme biosynthesis, respectively (Fig. 5). Coincidently, the significance of LA-type proteins in the particular pathway seems to be expanded

in the *Karlodinium* pathway comparing to the *Karenia* 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, we need to assess carefully whether the differences between the *Karenia* and *Karlodinium* pathways observed in our comparisons stemmed from the incomplete coverages of gene repertories in the two kareniacean species in future studies.

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 *Lepidodinium* (Fig. 5). 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 we 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 *Lepidodinium* and *Karenia*/*Karlodinium* (Fig. 5). Thus, as discussed the putative ancestral state of kareniacean species (see above), we 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.

We unexpectedly revealed a large contribution of chlorarachniophyte genes to the Chl *a* biosynthesis in *Lepidodinium* (Fig. 5). In the organismal tree of eukaryotes, chlorarachniophytes and dinoflagellates belong to two distantly related taxonomic assemblages, Rhizaria and Alveolata,

853 respectively. Likewise, the current plastids in chlorarachniophytes and *Lepidodinium* were derived
854 from distinct green algal groups, ulvophytes and pedinophytes, respectively (Suzuki et al, 2016;
855 Kamikawa et al. 2015a). Thus, the relationship between their host lineages or that between their
856 endosymbiont lineages (plastids) can provide no ground for the presence of chlorarachniophyte
857 genes in the *Lepidodinium* genome. If *Lepidodinium* feeds on chlorarachniophytes in the natural
858 environment, such predator-prey relationship led to the genetic influx from the prey
859 (chlorarachniophyte) genome to the predator (*Lepidodinium*) genome. Nevertheless, under the
860 circumstance postulated above, gene transferred from chlorarachniophytes could not have been
861 restricted to a single metabolic pathway. To understand the biological reasons for the genetic
862 contribution from chlorarachniophytes to the Chl *a* biosynthesis in *Lepidodinium*, we need to
863 explore (i) potential interaction between dinoflagellates and chlorarachniophytes in the
864 environment and (ii) the biochemical and/or physiological commonality in the proteins involved
865 in the Chl *a* biosynthesis between chlorarachniophytes and *Lepidodinium*.

866 We also noticed a clear difference in contribution of EA-type proteins to the three
867 pathways between *Lepidodinium* and *Karenia*/*Karlodinium* (Fig. 5). EA-type proteins are most
868 likely indispensable for the heme, Chl *a* and IPP biosyntheses in *Karenia*/*Karlodinium*. On the
869 other hand, only MgPMT in the Chl *a* biosynthetic pathway appeared to be of green algal origin
870 in *Lepidodinium*. One potential factor, which could introduce the marked difference between the
871 *Lepidodinium* and *Karenia*/*Karlodinium* pathways, is the difference in plastid-targeting signal of
872 nucleus-encoded plastid proteins between their endosymbionts. In both eukaryotes bearing
873 primary plastids (e.g., green algae) and those bearing complex plastids (e.g., haptophytes and
874 dinoflagellates), the vast majority of plastid-related proteins are nucleus-encoded, which are
875 synthesized in the cytosol and localized to the plastid. In green algae, nucleus-encoded plastid-
876 related proteins are synthesized with N-terminal extensions (so-called transit peptides or TP),

877 which act as the "tags" to pass through the two membranes surrounding their plastids (Patron &
878 Waller, 2007). On the other hand, the "tag" sequences, which enable nucleus-encoded proteins to
879 localize in complex plastids surrounded by three or four membranes, are more complex than green
880 algal plastids (Bolte et al, 2009). Patron et al, (2005) revealed that dinoflagellates with peridinin-
881 containing plastids and haptophytes appeared to share a bipartite structure of plastid-targeting
882 signal, which is composed of signal peptide and the TP-like region. Consequently, without any
883 substantial modification on plastid-targeting signals, the ancestral kareniacean species could have
884 targeted the proteins encoded by endosymbiotically transferred genes back to the haptophyte-
885 derived plastid. In contrast, nucleus-encoded plastid-related proteins in the green algal
886 endosymbiont engulfed by the ancestral *Lepidodinium* unlikely possessed bipartite plastid-
887 targeting signals. Thus, in the ancestral *Lepidodinium*, the proteins encoded by endosymbiotically
888 transferred genes needed to acquire bipartite plastid-targeting signals to be localized in the green
889 alga-derived plastid surrounded by four membranes. Altogether, we here propose the initial
890 presence/absence of bipartite plastid-targeting signals was one of the major factors affecting the
891 EGT in dinoflagellates bearing non-canonical plastids. As anticipated from the above scenario,
892 many LA-type proteins acquired from diverse eukaryotes bearing complex plastids (e.g.,
893 stramenopiles and chlorarachniophytes) were identified in the heme, Chl *a* and IPP biosynthetic
894 pathways in *Lepidodinium*. Nevertheless, the factor discussed above may not be dominant enough
895 to exclude EA-type proteins from a green alga (e.g., MgPMT; Fig. 3C) and LA-type proteins from
896 bacteria (e.g., GSAT; Fig. 2B) from the plastid proteome in *Lepidodinium*.
897

898 **Conclusion**

899

900 We 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 one is the descendants from an ancestral species established a haptophyte-derived plastid (i.e. *Karenia* and *Karlodinium*), and *Lepidodinium* 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. We 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 *Karenia*/*Karlodinium* and *Lepidodinium*, EGT appeared to contribute to the pathways in the former lineage much larger than those in the latter lineage. We 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 *Karenia* and *Karlodinium* 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).

Acknowledgements

E. M. was supported by a research fellowship from the Japanese Society for Promotion of Sciences (JSPS) for Young Scientists (no. 15J00821). This work was supported in part by grants from the JSPS awarded to Y. I. (23117006 and 16H04826).

References

925 Beale, S.I. 1999. Enzymes of chlorophyll biosynthesis. *Photosynthesis Research* 60:43-73. DOI:
926 https://doi.org/10.1023/A:1006297731456.
927
928 Bentlage, B., Rogers, T. S., Bachvaroff, T. R. and Delwiche, C. F. 2016. Complex ancestries of
929 isoprenoid synthesis in dinoflagellates. *Journal of Eukaryotic Microbiology* 63:123-137.
930 DOI:10.1111/jeu.12261.
931
932 Bjørnland, T., Haxo, F.T. and Liaaen-Jensen, S. 2003. Carotenoids of the Florida red tide
933 dinoflagellate *Karenia brevis*. *Biochemical Systematics and Ecology* 31:1147-1162. DOI:
934 10.1016/S0305-1978(03)00044-9.
935
936 Bolte, K., Bullmann, L., Hempel, F., Bozarth, A., Zauner, S. and Maier, U. G. 2009. Protein
937 targeting into secondary plastids. *Journal of Eukaryotic Microbiology* 56:9-15. DOI:
938 10.1111/j.1550-7408.2008.00370.x.
939
940 Burki, F., Imanian, B., Hehenberger, E., Hirakawa, Y., Maruyama, S. and Keeling, P. J. 2014.
941 Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. *Eukaryotic Cell* 13:246-
942 255. DOI: 10.1128/EC.00299-13.
943
944 Chen, G. E., Canniffe, D. P. and Hunter, C. N. 2017. Three classes of oxygen-dependent cyclase
945 involved in chlorophyll and bacteriochlorophyll biosynthesis. *Proceedings of the National*
946 *Academy of Sciences of the United States of America* 114:6280-6285. DOI:
947 10.1073/pnas.1701687114.
948
949 Cihlář, J., Füssy, Z., Horák, A. and Oborník, M. 2016. Evolution of the tetrapyrrole biosynthetic
950 pathway in secondary algae: conservation, redundancy and replacement. *PLoS ONE* 11 e0166338.
951 DOI: 10.1371/journal.pone.0166338.
952
953 Darriba, D., Taboada, G.L., Doallo, R. and Posada, D. 2011. ProtTest 3: fast selection of best-fit
954 models of protein evolution. *Bioinformatics* 27:1164-1165. DOI: 10.1093/bioinformatics/btr088.
955
956 Dubey, V. S., Bhalla, R. and Luthra, R. 2003. An overview of the non-mevalonate pathway for
957 terpenoid biosynthesis in plants. *Journal of Biosciences* 28:637-646.
958
959 Eisenreich, W., Bacher, A., Arigoni, D. and Rohdich, F. 2004. Biosynthesis of isoprenoids via the
960 non-mevalonate pathway. *Cellular and Molecular Life Sciences* 61:1401-1426. DOI:

10.1007/s00018-004-3381-z.

Eschbach, S., Speth, V., Hansmann, P. and Sitte, P. 1990. Freeze-fracture study of the single membrane between host cell and endocytobiont in the dinoflagellates *Glenodinium foliaceum* and *Peridinium balticum*. *Journal of Phycology* 26:324-328. DOI: 10.1111/j.0022-3646.1990.00324.x.

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.

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.

Geider, R. and Gunter, P. A. 1989. Evidence for the presence of phycoerythrin in *Dinophysis norvegica*, a pink dinoflagellate. *British Phycological Journal* 24:195-198. DOI: 10.1080/00071618900650191.

Grauvogel, C., Reece, K. S., Brinkmann, H. and Petersen, J. 2007. Plastid isoprenoid metabolism in the oyster parasite *Perkinsus marinus* connects dinoflagellates and malaria pathogens—new impetus for studying alveolates. *Journal of Molecular Evolution* 65:725-729. DOI: 10.1007/s00239-007-9053-5.

Hewes, C. D., Mitchell, B. G., Moissan, T. A., Vernet, M. and Reid, F. M. H. 1998. The phycobilin signatures of chloroplasts from three dinoflagellate species: a microanalytical study of *Dinophysis caudata*, *D. fortii*, and *D. acuminata* (Dinophysiales, Dinophyceae). *Journal of Phycology* 34:945-

951. DOI: 10.1046/j.1529-8817.1998.340945.x.

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

Horiguchi, T. 2006. Algae and their chloroplasts with particular reference to the dinoflagellates. *Paleontological Research* 10:299-309. DOI: 10.2517/prpsj.10.299.

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.

Hehenberger, E., Burki, F., Kolisko, M. and Keeling, P. J. 2016. Functional relationship between a dinoflagellate host and its diatom endosymbiont. *Molecular Biology and Evolution*, 33(9): 2376-2390. DOI: 10.1093/molbev/msw109.

Hollingshead, S., Kopecná, J., Jackson, P. J., Canniffe, D. P., Davison, P. A., Dickman, M. J., Sobotka, R. and Hunter C. N. 2012. Conserved chloroplast open-reading frame *ycf54* is required for activity of the magnesium protoporphyrin monomethylester oxidative cyclase in *Synechocystis* PCC 6803. *The Journal of Biological Chemistry* 287:27823-27833. DOI: 10.1074/jbc.M112.352526.

Imanian, B., Pombert, J. F. and Keeling, P. J. 2010. The complete plastid genomes of the two 'dinotoms' *Durinskia baltica* and *Kryptoperidinium foliaceum*. *PLoS ONE* 5:e10711. DOI: 10.1371/journal.pone.0010711.

Inagaki, Y., Dacks, J. B., Doolittle, W. F., Watanabe, K. I. and Ohama, T. 2000. Evolutionary relationship between dinoflagellates bearing obligate diatom endosymbionts: insight into tertiary endosymbiosis. *International Journal of Systematic and Evolutionary Microbiology* 50:2075-2081. DOI: 10.1099/00207713-50-6-2075.

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.

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.

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.

Kořený, L., Oborník, M and Lukeš, J. 2013. Make it, take it, or leave it: heme metabolism of parasites. *PLoS Pathogens* 9: e1003088. DOI: 10.1371/journal.ppat.1003088.

Kuzuyama, T. 2002. Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Bioscience, Biotechnology, and Biochemistry* 66:1619-1627. DOI: 10.1271/bbb.66.1619.

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.

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.

Minamizaki, K., Mizoguchi, T., Goto, T., Tamiaki, H. and Fujita, Y. 2008. Identification of two homologous genes, $chlA_I$ and $chlA_{II}$, that are differentially involved in isocyclic ring formation of chlorophyll *a* in the cyanobacterium *Synechocystis* sp. PCC 6803. *The Journal of Biological Chemistry* 283:2684-2692. DOI: 10.1074/jbc.M708954200.

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.

Nagata, N., Tanaka, R. and Tanaka, A. 2007. The major route for chlorophyll synthesis includes [3,8-divinyl]-chlorophyllide a reduction in *Arabidopsis thaliana*. *Plant and Cell Physiology*, 48: 1803-1808. DOI: 10.1093/pcp/pcm153.

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.

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.

Onuma, R. and Horiguchi, T. 2015. Kleptochloroplast enlargement, karyoklepty and the distribution of the cryptomonad nucleus in *Nusuttodinium* (= *Gymnodinium*) *aeruginosum* (Dinophyceae). *Protist* 166:177-195. DOI: 10.1016/j.protis.2015.01.004.

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.

Rohmer, M. 1999. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Natural Product Reports* 16:565-574. DOI: 10.1039/A709175C.

Schoefs, B. 2000. The light-dependent and light-independent reduction of protochlorophyllide *a* to chlorophyllide *a*. *Photosynthetica* 36:481-496. DOI: 10.1023/A:1007002101856.

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.

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.

Suzuki, J. Y. and Bauer, C. E. 1992. Light-independent chlorophyll biosynthesis: involvement of

1177 the chloroplast gene *chlL* (*frxC*). *The Plant Cell* 4:929-940. DOI: 10.1105/tpc.4.8.929.
1178
1179 Takishita, K., Ishida, K. and Maruyama, T. 2004. Phylogeny of nuclear-encoded plastid-targeted
1180 *gapdh* gene supports separate origins for the peridinin- and the fucoxanthin derivative-containing
1181 plastids of dinoflagellates. *Protist* 155:447-458. DOI: 10.1078/1434461042650325.
1182
1183 Taylor, F. J. R., Hoppenrath, M. and Saldarriaga, J. F. 2008. Dinoflagellate diversity and
1184 distribution. *Biodiversity and Conservation* 17:407-418. DOI: 10.1007/s10531-007-9258-3.
1185
1186 Tengs, T., Dahlberg, O. J., Shalchian-Tabrizi, K., Klaveness, D., Rudi, K., Delwiche, C. F. and
1187 Jakobsen, K. S. 2000. Phylogenetic analyses indicate that the 19' hexanoyloxy-fucoxanthin-
1188 containing dinoflagellates have tertiary plastids of haptophyte origin. *Molecular Biology and*
1189 *Evolution* 17:718-729. DOI: 10.1093/oxfordjournals.molbev.a026350.
1190
1191 Tomas, R. W. and Cox, E. R. 1973. Observations on the symbiosis of *Peridinium balticum* and its
1192 intracellular alga. *Journal of Phycology* 9:304-323. DOI: 10.1111/j.1529-8817.1973.tb04098.x.
1193
1194 Vavilin, D. V. and Vermaas, W. F. J. 2002. Regulation of the tetrapyrrole biosynthetic pathway
1195 leading to heme and chlorophyll in plants and cyanobacteria. *Physiologia Plantarum* 115:9-24.
1196 DOI: 10.1034/j.1399-3054.2002.1150102.x.
1197
1198 Watanabe, M. M., Takeda, Y., Sasa, T., Inouye, I., Suda, S., Sawaguchi, T. and Chihara, M. 1987.
1199 A green dinoflagellate with chlorophylls *a* and *b*: morphology, fine structure of the chloroplast and
1200 chlorophyll composition. *Journal of Phycology*,23(s2):382-389. DOI: 10.1111/j.1529-
1201 8817.1987.tb04148.x.
1202
1203 Watanabe, M. M., Suda, S., Inouye, I., Sawaguchi, T. and Chihara, M. 1990. *Lepidodinium viride*
1204 gen. et sp. nov. (Gymnodinaiales, Dinophyta), a green dinoflagellate with a chlorophyll *a*- and *b*-
1205 containing endosymbiont. *Journal of Phycology* 26:741-751. DOI: 10.1111/j.0022-
1206 3646.1990.00741.x.
1207
1208 van Wijk, K. J., and Baginsky, S. 2011. Plastid proteomics in higher plants: current state and future
1209 goals. *Plant Physiology* 155:1578-1588. DOI: 10.1104/pp.111.172932.
1210
1211 Yamanashi, K., Minamizaki, K. and Fujita, Y. 2015. Identification of the *chlE* gene encoding
1212 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.

Legends for Figures

Fig 1. Enzymes examined in this study, and their substrates and products.

Enzymes involved in C5 pathway for the heme biosynthesis, Chl *a* biosynthetic pathway and the non-mevalonate pathway for the IPP biosynthesis in cyanobacteria and/or land plant plastids are shown in red. C5 pathway 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.

Fig 2. Maximum-likelihood phylogenies of 9 proteins involved in 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*, *Karlodinium* and *Lepidodinium*. 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 as follows. Red, dinoflagellates including *Karenia*, *Karlodinium* and *Lepidodinium*; yellow, haptophytes; green, green plants (i.e. green algae plus land plants), yellowish green, chlorarachniophytes; brown, stramenopiles; light-blue, euglenids; dark blue, chlomerids; purple, cryptophytes; pink, red algae; light-green, glaucophytes;

blue-green, cyanobacteria; light-gray, apicomplexan parasites. Subtrees/branches of heterotrophic eukaryotes (except apicomplexan parasites) and bacteria (except cyanobacteria) are shown in dark-gray. 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). **E.** uroporphyrinogen III synthase (UROS). **F.** uroporphyrinogen decarboxylase (UROD). **G.** coproporphyrinogen oxidase (CPOX). **H.** protoporphyrinogen IX oxidase (PPOX) and **I.** ferrochelatase (FeCH). The identical ML trees with full sequence names and MLBPs ≥50% are provided as the supplementary materials.

Fig 3. Maximum-likelihood phylogenies of 7 proteins involved in the Chl *a* biosynthetic pathway. The details of this figure are same as those of Fig. 2. **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). **F.** light-dependent protochlorophyllide reductase (POR). **G.** chlorophyll synthase (CS). The identical ML trees with full sequence names and MLBPs ≥50% are provided as the supplementary materials. Note that we present no phylogeny of ChlI or MgPME cyclase, as the former is plastid-encoded, and the latter was not identified in dinoflagellates (including *Karenia*, *Karlodinium* or *Lepidodinium*), diatoms, cryptophytes or haptophytes.

Fig 4. Maximum-likelihood phylogenies of 7 proteins involved in the non-mevalonate pathway for IPP biosynthesis.

The details of this figure are same as those of Fig. 2. **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). The identical ML trees with full sequence names and MLBPs ≥50% are provided as the supplementary materials.

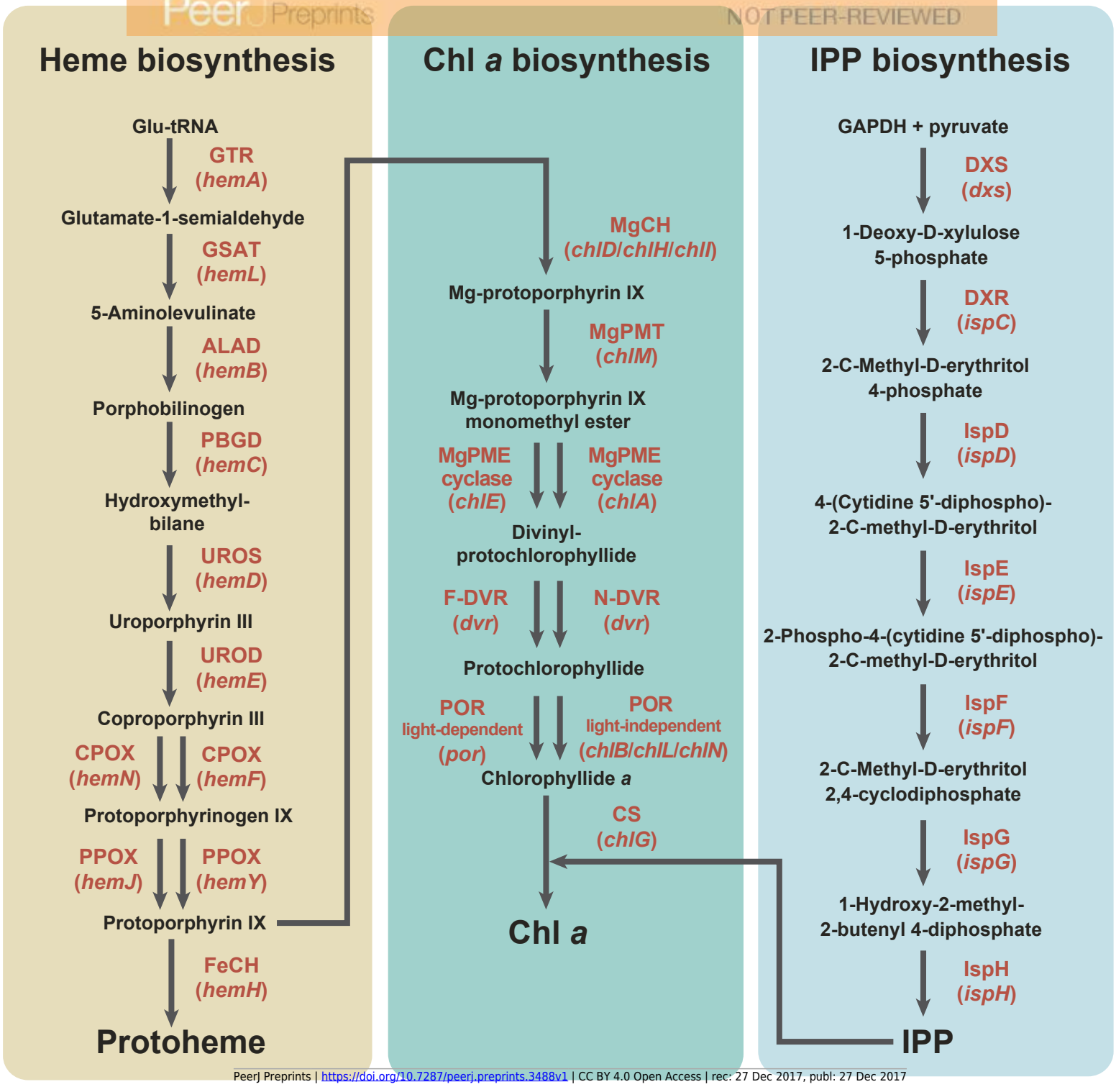
Fig 5. Overview of the origins of proteins involved in three plastid-localized biosynthetic pathways in *Karenia*, *Karlodinium* and *Lepidodinium*.

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. For DVR involved in the Chl *a* biosynthetic pathway, we distinguish N-DVR and F-DVR by labeling "N" and "F," respectively. The squares in the fourth lows labeled with question marks represent the sequences of which origins remain uncertain.

Figure 1(on next page)

Fig 1. Enzymes examined in this study, and their substrates and products.

Enzymes involved in C5 pathway for the heme biosynthesis, Chl *a* biosynthetic pathway and the non-mevalonate pathway for the IPP biosynthesis in cyanobacteria and/or land plant plastids are shown in red. C5 pathway 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.

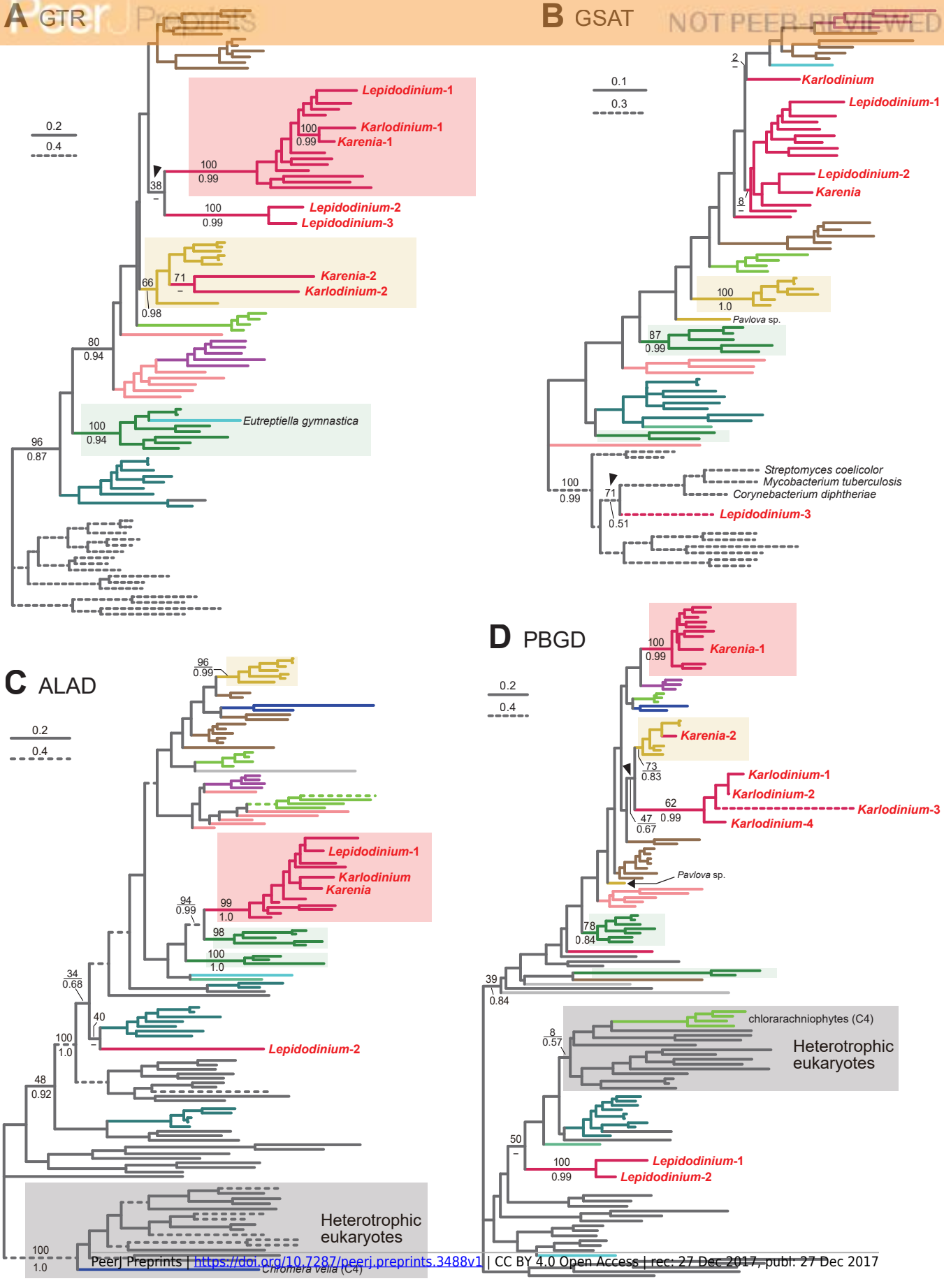


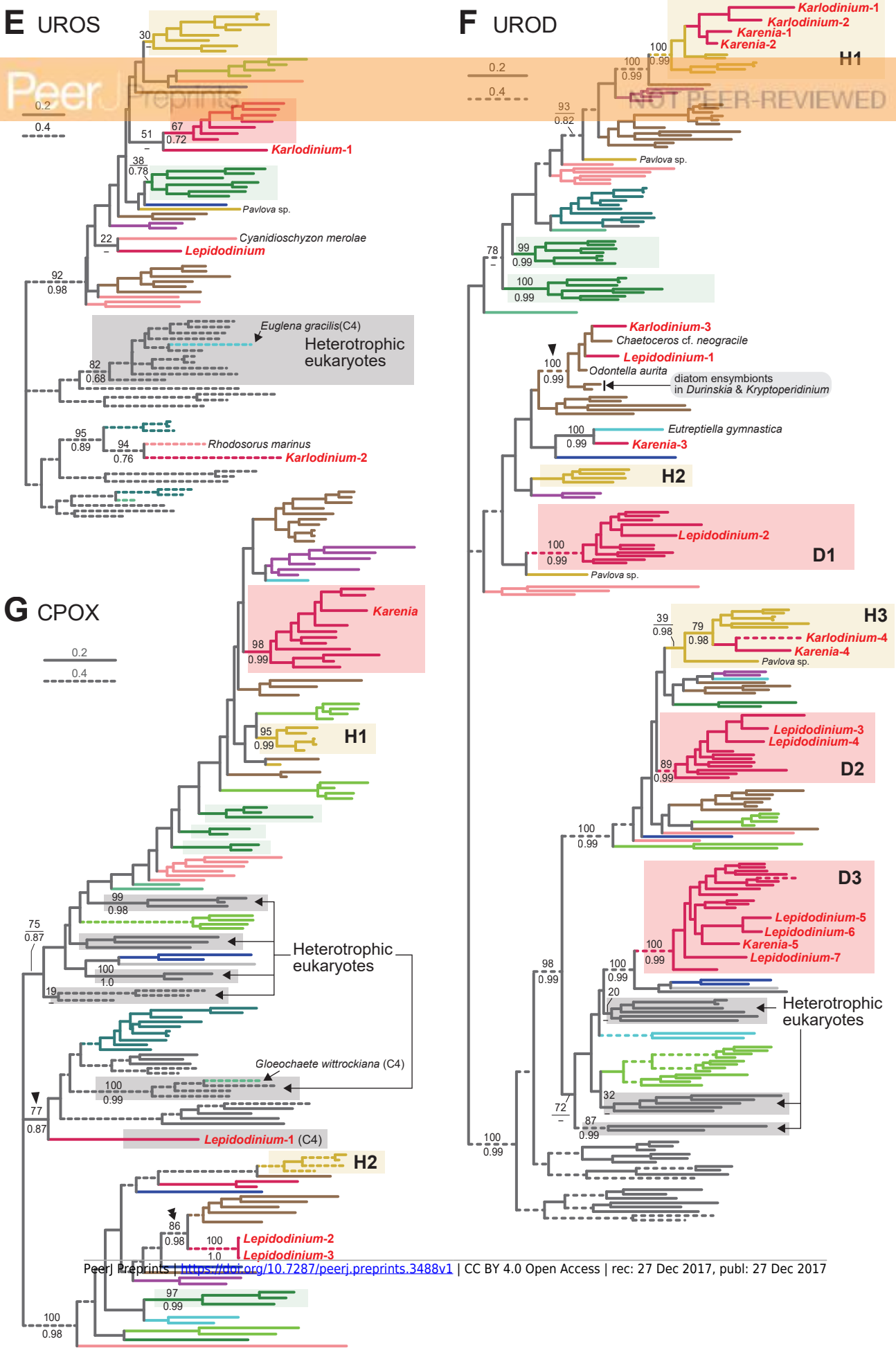
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3488v1 | CC BY 4.0 Open Access | rec: 27 Dec 2017, publ: 27 Dec 2017

Figure 2(on next page)

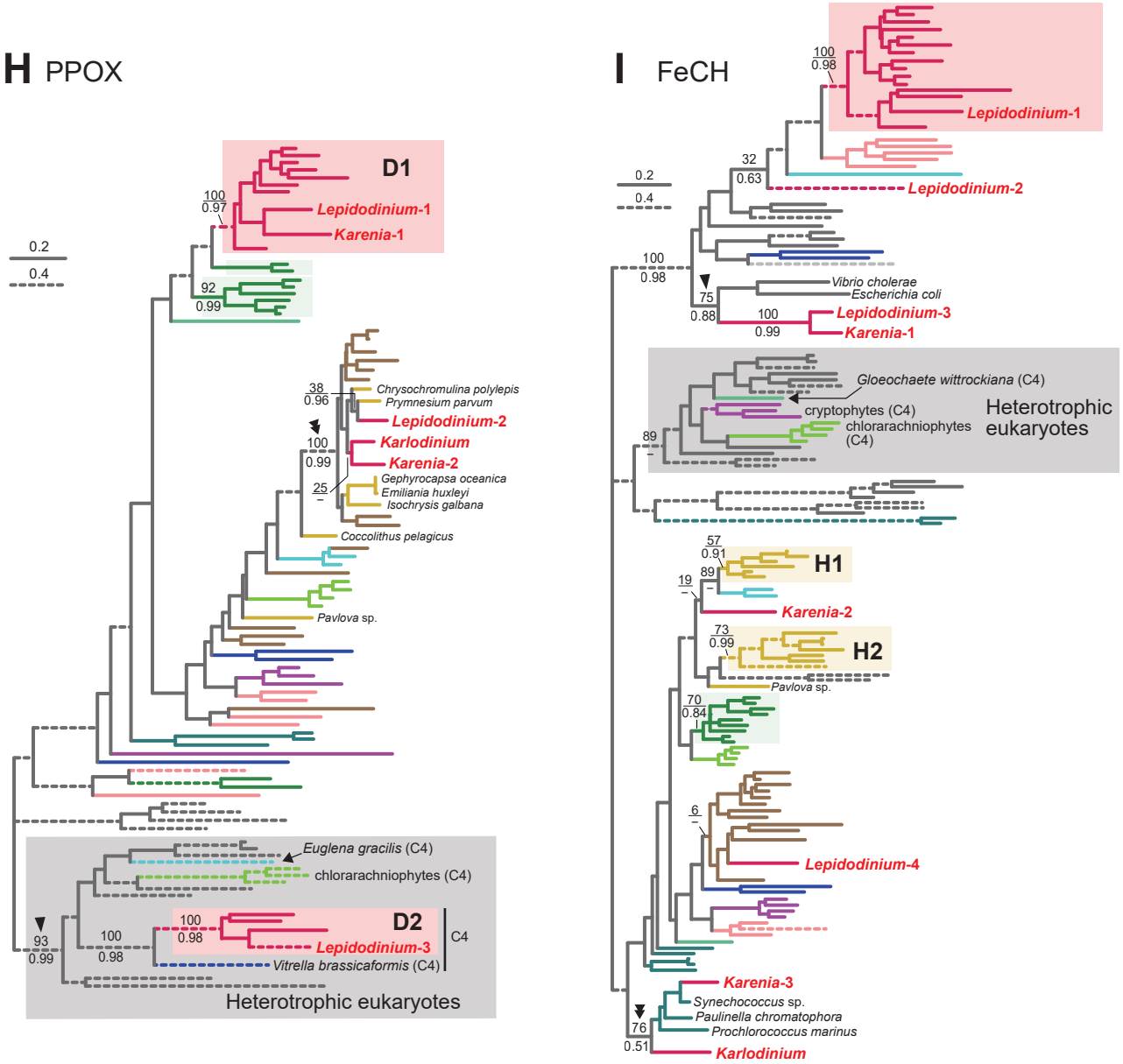
Fig 2. Maximum-likelihood phylogenies of 8 proteins involved in C5 pathway for the heme biosynthesis.

We provide the maximum-likelihood bootstrap values (MLBPs), as well as Bayesian posterior probabilities (BPPs), only for the selected nodes, which are important to inferred the origins of the proteins of *Karenia*, *Karlodinium* and *Lepidodinium*. 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 as follows. Red, dinoflagellates including *Karenia*, *Karlodinium* and *Lepidodinium*; yellow, haptophytes; green, green plants (i.e. green algae plus land plants), yellowish green, chlorarachniophytes; brown, stramenopiles; light-blue, euglenids; dark blue, chlomerids; purple, cryptophytes; pink, red algae; light-green, glaucophytes; blue-green, cyanobacteria; apicomplexan parasites, light-gray. Subtrees/branches of heterotrophic eukaryotes (except apicomplexan parasites) and bacteria (except cyanobacteria) are shown in dark-gray. 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). **E.** Uroporphyrinogen III synthase (UROS). **F.** uroporphyrinogen decarboxylase (UROD). **G.** coproporphyrinogen oxidase (CPOX). **H.** protoporphyrinogen IX oxidase (PPOX) and **I.** ferrochelatase (FeCH). The identical ML trees with full sequence names and MLBPs ≥50% are provided as the supplementary materials.





PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3488v1 | CC BY 4.0 Open Access | rec: 27 Dec 2017, publ: 27 Dec 2017

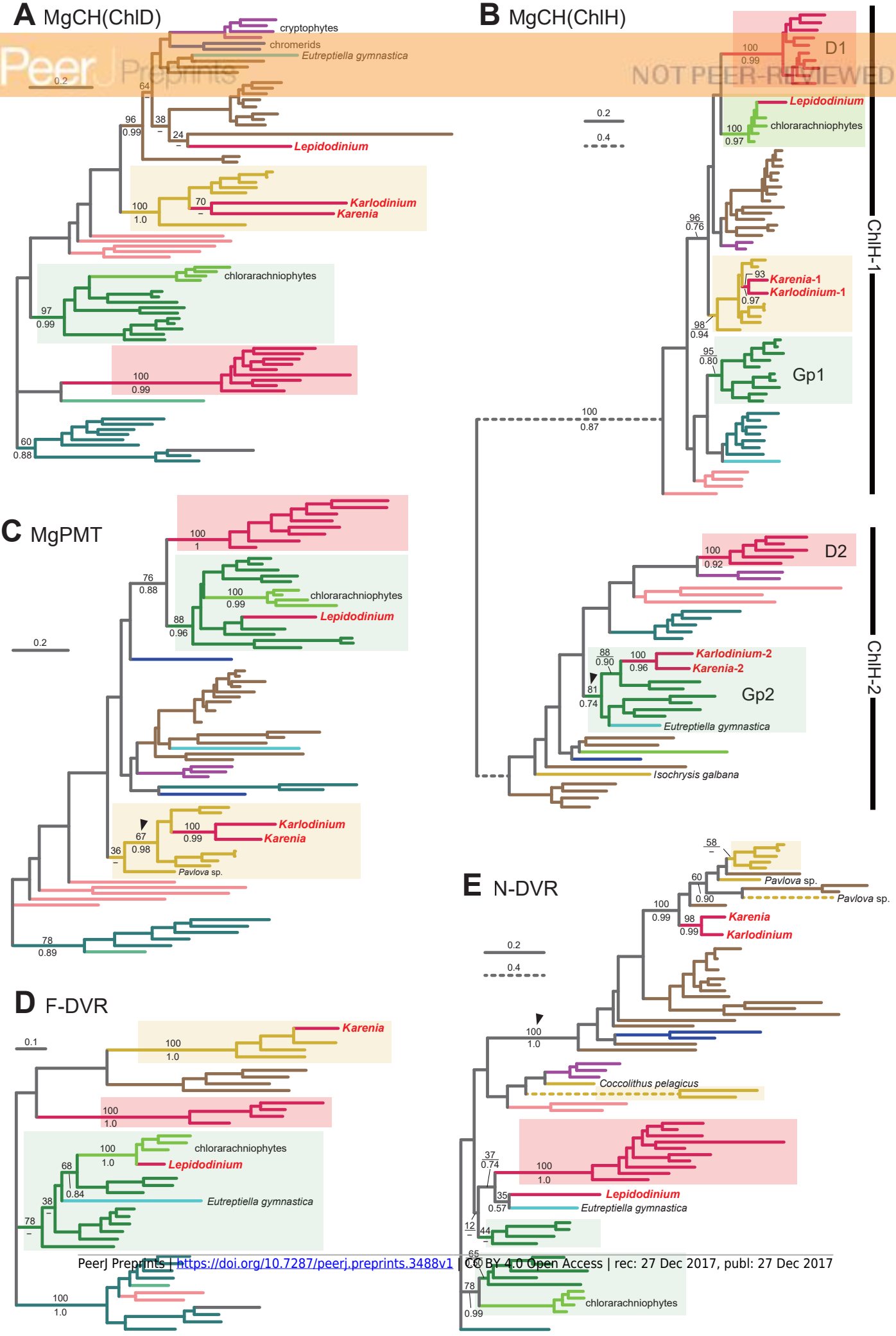


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3488v1 | CC BY 4.0 Open Access | rec: 27 Dec 2017, publ: 27 Dec 2017

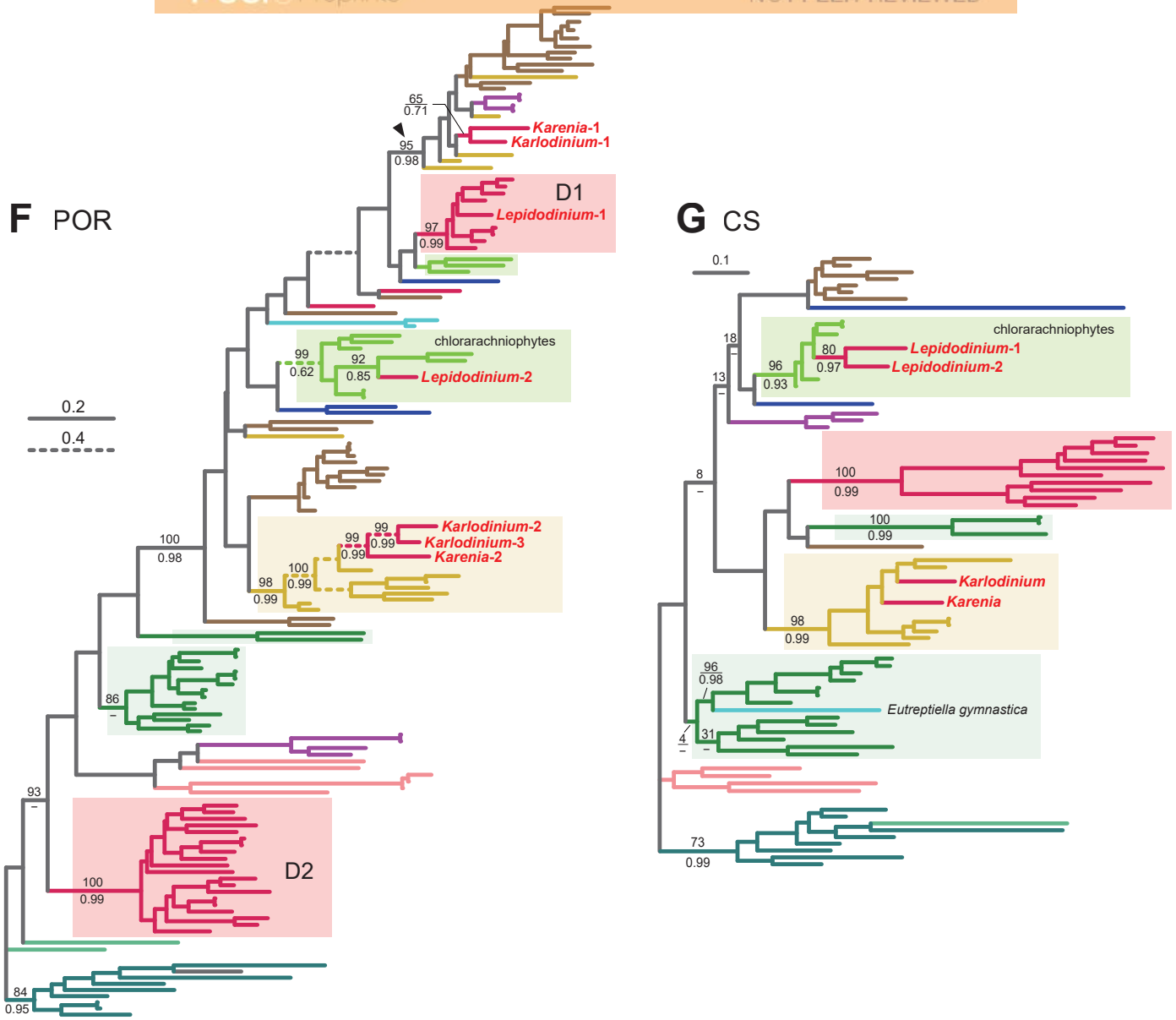
Figure 3(on next page)

Fig 3. Maximum-likelihood phylogenies of 7 proteins involved in the Chl *a* biosynthetic pathway.

The details of this figure are same as those of Fig. 3. **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). **F.** light-dependent protochlorophyllide reductase (POR). **G.** chlorophyll synthase (CS). The identical ML trees with full sequence names and MLBPs ≥50% are provided as the supplementary materials. Note that we present no phylogeny of ChlI or MgPMT cyclase, as the former is plastid-encoded, and the latter was not identified in dinoflagellates (including *Karenia*, *Karlodinium* or *Lepidodinium*), diatoms, cryptophytes or haptophytes.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3488v1 | CC BY 4.0 Open Access | rec: 27 Dec 2017, publ: 27 Dec 2017



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3488v1 | CC BY 4.0 Open Access | rec: 27 Dec 2017, publ: 27 Dec 2017

Figure 4(on next page)

Fig 4. Maximum-likelihood phylogenies of 7 proteins involved in the non-mevalonate pathway for IPP biosynthesis.

The details of this figure are same as those of Fig. 3. **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). The identical ML trees with full sequence names and MLBPs ≥50% are provided as the supplementary materials.

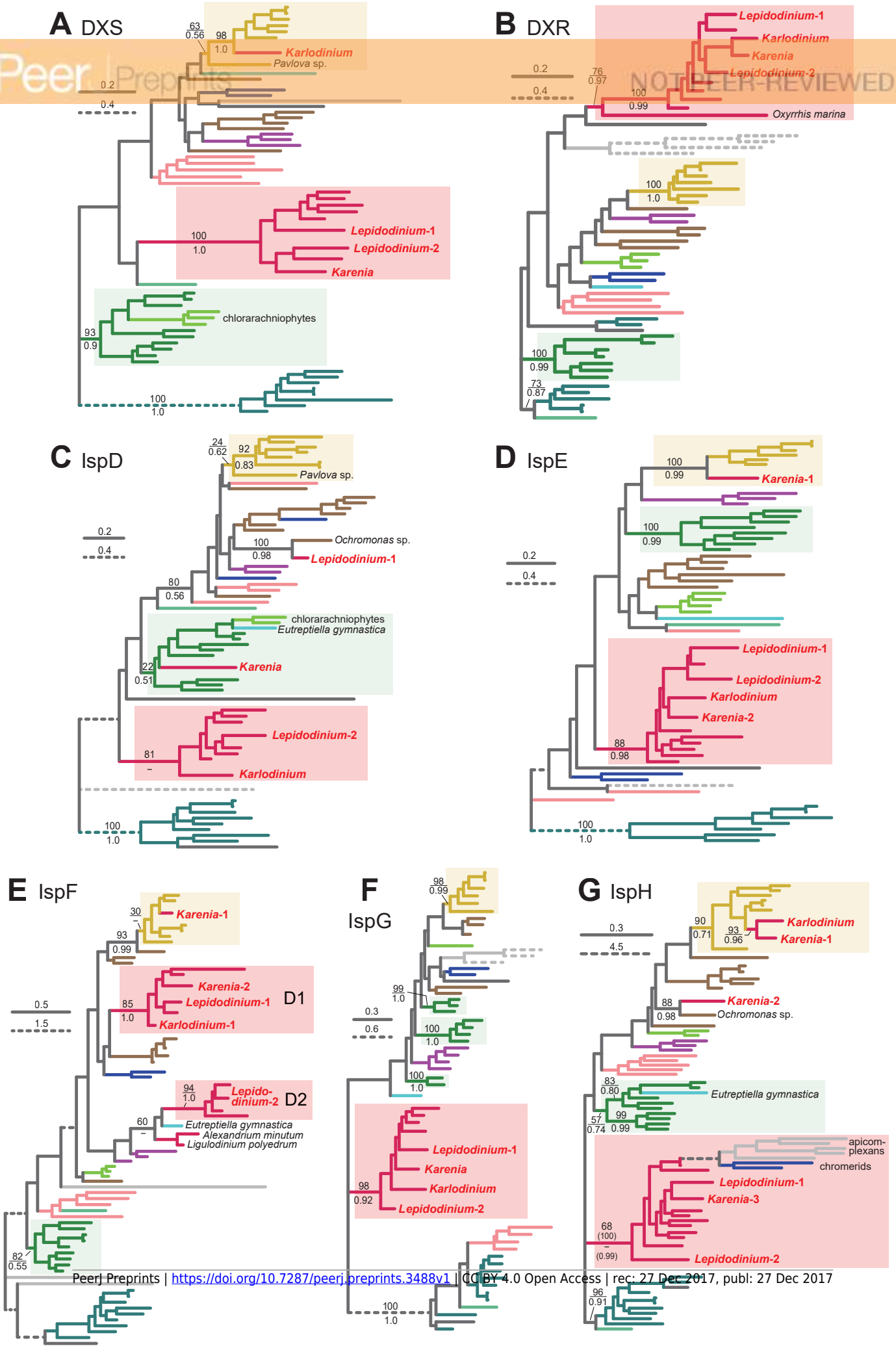
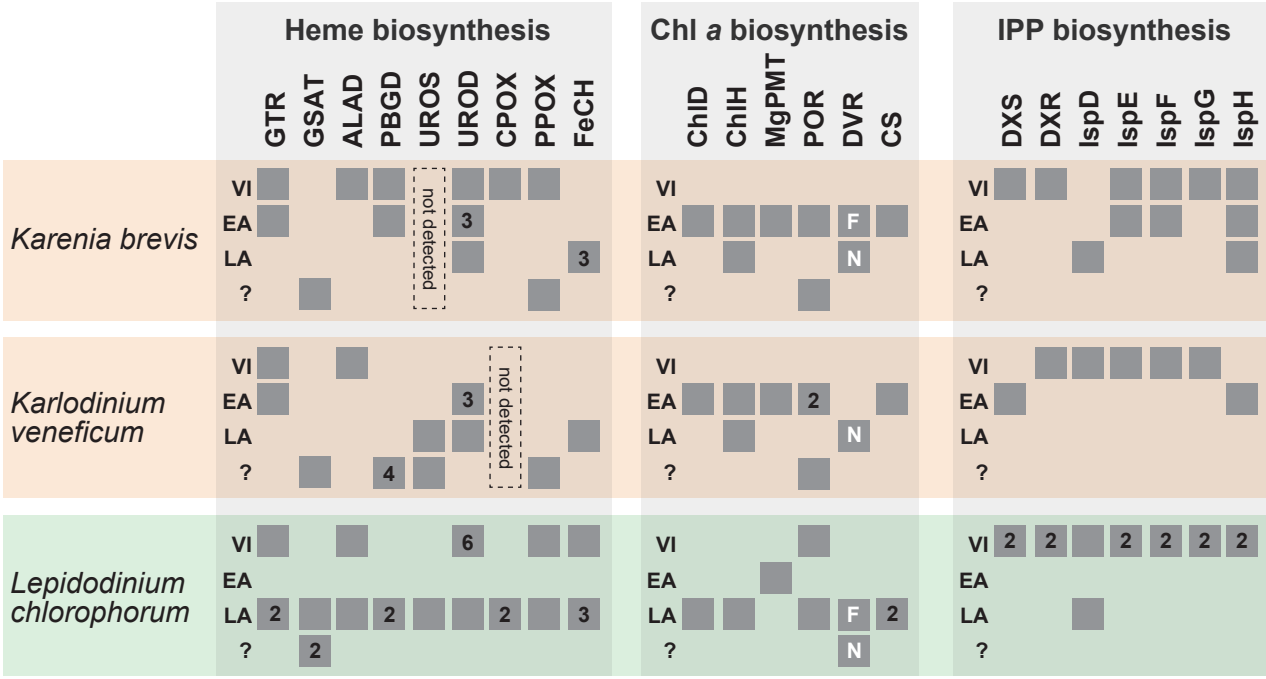


Figure 5(on next page)

Fig 5. Overview of the origins of proteins involved in three plastid-localized biosynthetic pathways in *Karenia*, *Karlodinium* and *Lepidodinium.*

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. For DVR involved in the Chl a biosynthetic pathway, we distinguish N-DVR and F-DVR by labeling “N” and “F,” respectively. The squares in the fourth lows labeled with question marks represent the sequences of which origins remain uncertain.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3488v1 | CC BY 4.0 Open Access | rec: 27 Dec 2017, publ: 27 Dec 2017