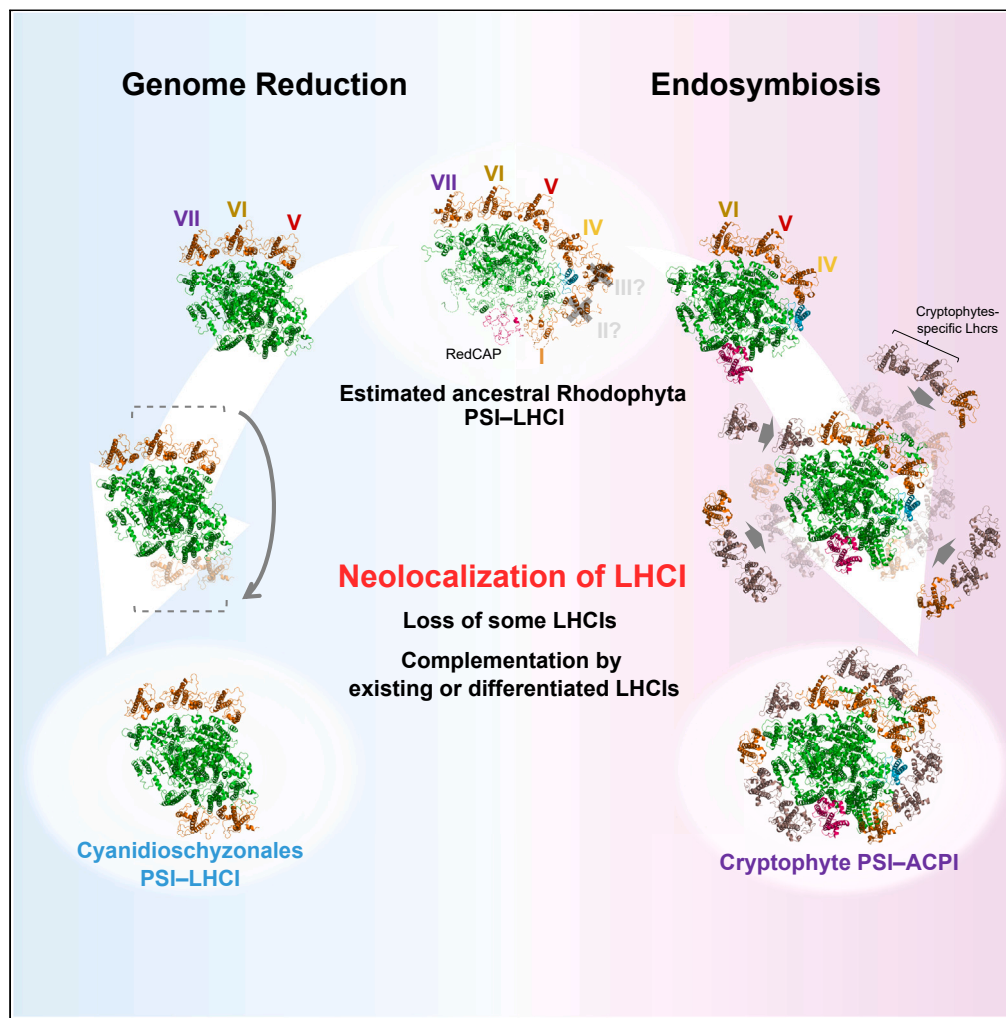**Article**

Unraveling the evolutionary trajectory of LHCI in red-lineage algae: Conservation, diversification, and neolocalization

Minoru Kumazawa,
Kentaro Ifuku

ifuku.kentaro.2m@kyoto-u.ac.jp

Highlights

Comprehensive phylogenetic analysis of LHCs reveals orthology in red-lineage algae

Genome reduction would trigger to the loss of LHCs and complementation by existing LHCs

Endosymbiosis also leads to loss of LHCs and complementation by differentiated LHCs

"Neolocalication": complementation of lost components by existing/ differentiated ones

Article

Unraveling the evolutionary trajectory of LHCI in red-lineage algae: Conservation, diversification, and neolocalization

Minoru Kumazawa[1] and Kentaro Ifuku[1,2,*]

SUMMARY

Red algae and the secondary symbiotic algae that engulfed a red alga as an endosymbiont are called red-lineage algae. Several photosystem (PS) I–light-harvesting complex I (LHCI) structures have been reported from red-lineage algae—two red algae *Cyanidioschyzon merolae* (Cyanidiophyceae) and *Porphyridium purpureum* (Rhodophytina), a diatom, and a Cryptophyte. Here, we clarified the orthologous relation of LHCIs by combining a detailed phylogenetic analysis and the structural information of PSI–LHCI. We found that the seven Lhcr groups in LHCI are conserved in Rhodophytina; furthermore, during both genome reduction in Cyanidioschyzonales and endosymbiosis leading to Cryptophyta, some LHCIs were lost and replaced by existing or differentiated LHCIs. We denominate "neolocalization" to these examples of flexible reorganization of LHCIs. This study provides insights into the evolutionary process of LHCIs in red-lineage algae and clarifies the need for both molecular phylogeny and structural information to elucidate the plausible evolutionary history of LHCI.

INTRODUCTION

Oxygenic photosynthetic organisms, such as cyanobacteria, algae, and terrestrial plants, play an essential role in capturing sunlight, producing organic matter, and maintaining life on Earth both underwater and on land. Among the various eukaryotic photosynthetic organisms, algae and terrestrial plants acquired chloroplasts through endosymbiosis with cyanobacteria.[1] These photosynthetic organisms possess primary plastids form Archaeplastida and are divided into three groups, Rhodophyta (red algae), Viridiplantae (green algae and terrestrial plants), and Glaucophyta. Several secondary or tertial endosymbiotic events led to diversified eukaryotic algae. For instance, red-lineage secondary endosymbiotic algae acquired plastids derived from red algae and include key marine taxa including diatoms and Haptophytes, which dominate in modern oceans.[2]

To enable more efficient light capture, photosynthetic organisms possess peripheral light-harvesting antennas around the two photosystems. In eukaryotic photosynthetic organisms, these antennas are protein complexes holding light-harvesting pigments, which transfer the excitation energy acquired from the light to the photosystems through excitation energy transfer.[2] Red algae have a phycobilisome, a superficial light-harvesting antenna complex on the stromal side of photosystem (PS) II, and two-dimensionally coordinated transmembrane light-harvesting pigment-protein complexes (LHCs) associated with PSI.[3–6] LHCs bind various types of chlorophylls and carotenoids as light-harvesting pigments and serve as light-harvesting antennas in red and green algae, land plants, red-lineage secondary endosymbiotic algae, green-lineage secondary endosymbiotic algae, and dinoflagellates.[7,8]

Red algal LHCs contain chlorophyll *a* and zeaxanthin as a carotenoid, while most LHCs of red-lineage secondary endosymbiotic algae include chlorophyll *a* and *c*; carotenoids depend on the taxonomic group. In fact, LHCs are named after their binding carotenoids.[8] For example, diatoms and Haptophytes contain fucoxanthin or 19′-hexanoyloxy fucoxanthin as major carotenoid in their LHCs, thus their LHCs are called fucoxanthin chlorophyll *a/c*-binding proteins (FCPs). Among red-lineage secondary endosymbiotic algae, diatoms utilize FCPs as peripheral antennas for both PS I and II.[9–13] Similar light-harvesting systems probably exist in other Stramenopiles and Haptophytes. At least, Eustigmatophyceae, belonging to Stramenopiles, utilize LHCs for light harvesting for both PS.[14]

Based on molecular phylogeny, the LHCs of red-lineage algae are divided into six subfamilies: Lhcr, Lhcz, Lhcq, Lhcf, Lhcx, and CgLhcr9 homologs.[15] Some Stramenopiles, including Eustigmatophyceae and Phaeophyceae, as well as Chromera from Alveolate have another LHC subfamily called red-shifted *Chromera* light-harvesting proteins (Red-CLH).[14,16] Red algae only have the Lhcr subfamily, Stramenopiles and Haptophytes possess all six subfamilies of the red-lineage LHCs, while Cryptophytes only have Lhcr and Lhcz subfamilies.

Recently, the advancement of cryoelectron microscopy structural analysis allowed discerning the structures of the PS-peripheral light-harvesting antenna supercomplexes of red-lineage algae. In red algae, the *Cyanidioschyzon merolae* PSI–LHCI supercomplex,[3] and the

[1]Graduate School of Agriculture, Kyoto University, Kyoto, Kyoto 606-8502, Japan
[2]Lead contact
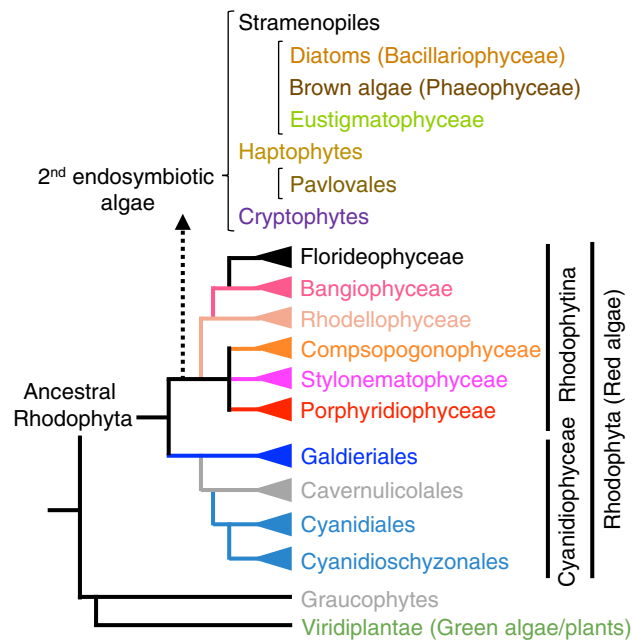*Correspondence: ifuku.kentaro.2m@kyoto-u.ac.jp

Figure 1. Simplified phylogeny of Rhodophyta (red algae) and classification of red-linage secondary endosymbiotic algae
The phylogeny of Rhodophyta is based on Yang et al. (2016) and Park et al. (2023). The secondary endosymbiosis node in the tree is based on Yoon et al. (2002) and Kim et al. (2017). The root of the tree is the ancestor of Archaeplastida. "Ancestral Rhodophyta" in the figure represents the last common ancestor of extant Rhodophyta.

Porphyridium purpureum phycobilisome–PSII–PSI–LHCI megacomplex[4] have been reported. In red-lineage secondary endosymbiotic algae, molecular-level structural PS models have been reported in diatoms and a Cryptophyte. For instance, the diatom *Chaetoceros gracilis* PSI–FCPI supercomplexes[9,12] and *C. gracilis* PSII–FCPII supercomplexes[10,11,13] have been reported. In addition, PSII–FCPII structures from centric diatoms *Thalassiosira pseudonana* and *Cyclotella meneghiniana* were recently reported.[17,18] The LHCs of Cryptophytes are called alloxanthin-chlorophyll *a*/*c*-binding proteins (ACPs); the structure of the *Chroomonas placoidea* PSI–ACPI has been recently reported.[19]

With the identification of LHCs in the PSI–LHCI structural models recently reported, it is now possible to evaluate the evolutionary process of the molecular assembly of LHCI associated with PSI. The molecular assembly model of the red-lineage PSI–LHCI has been discussed only based on spatial arrangements of the subunits present in the structures.[19,20] However, an evolutionary model of the photosynthetic super-complex should comprise both molecular phylogeny and structural information. Such an integrative understanding of the complex structures and molecular phylogenies of primitive species has been attempted in the green lineage.[21] Loss and gain of LHC subfamilies during the evolutionary history of the red-lineage algae were investigated through phylogenetic analysis of diatom LHCs.[15] A red-lineage chlorophyll *a*/*b*-binding-like protein (RedCAP), a distinctive family of the LHC superfamily, is conserved in PSI–LHCI of Rhodophytina red algae, Cryptophytes, and diatoms.[4,12,19,22,23] However, the evolutionary model of the red algal LHCI and RedCAP remains incomplete because of the limited structural information, insufficient genome and transcriptome information (until recently), and lack of detailed molecular phylogeny at the ortholog level.[24,25]

In this study, we performed a molecular phylogenetic analysis to clarify the orthology of LHCIs in red-lineage algae using recently reported genomes and transcriptomic data. The detailed molecular phylogeny of the red-lineage LHCI, specifically Lhcrs, is combined with PSI–LHCI structural models recently reported from red and red-lineage algae. This has uncovered the conservation, diversification, and differentiation of the molecular assembly of LHCIs, especially in red algae and Cryptophytes. Based on our analyses, we propose a plausible evolutionary trajectory of LHCI proteins associated with PSI in red-lineage algae.

RESULTS

Molecular phylogeny of red algal LHCI

Red algae possess two types of membrane-spanning light-harvesting pigment-protein complexes in PSI; LHCs that belong to the Lhcr subfamily and a RedCAP, part of the LHC superfamily.[4,22,23] In contrast, in PSII, they have a large membrane-peripheral light-harvesting protein supercomplex, known as phycobilisome. To elucidate the molecular phylogeny of the LHC family in PSI, Lhcr sequences from a broad lineage of red algae were collected minimizing as much as possible any taxonomic biases.

Ancient Rhodophytina is the original endosymbiont of red-lineage secondary endosymbiotic algae (Figure 1).[26,27] After secondary endosymbiosis, the earliest divergent event divided red algae into two major groups: Rhodophytina and Cyanidiophyceae (or Cyanidiophytina).[28,29] The former includes classes such as Porphyridiophyceae, Stylonematophyceae, and Compsopogonophyceae with the subclades
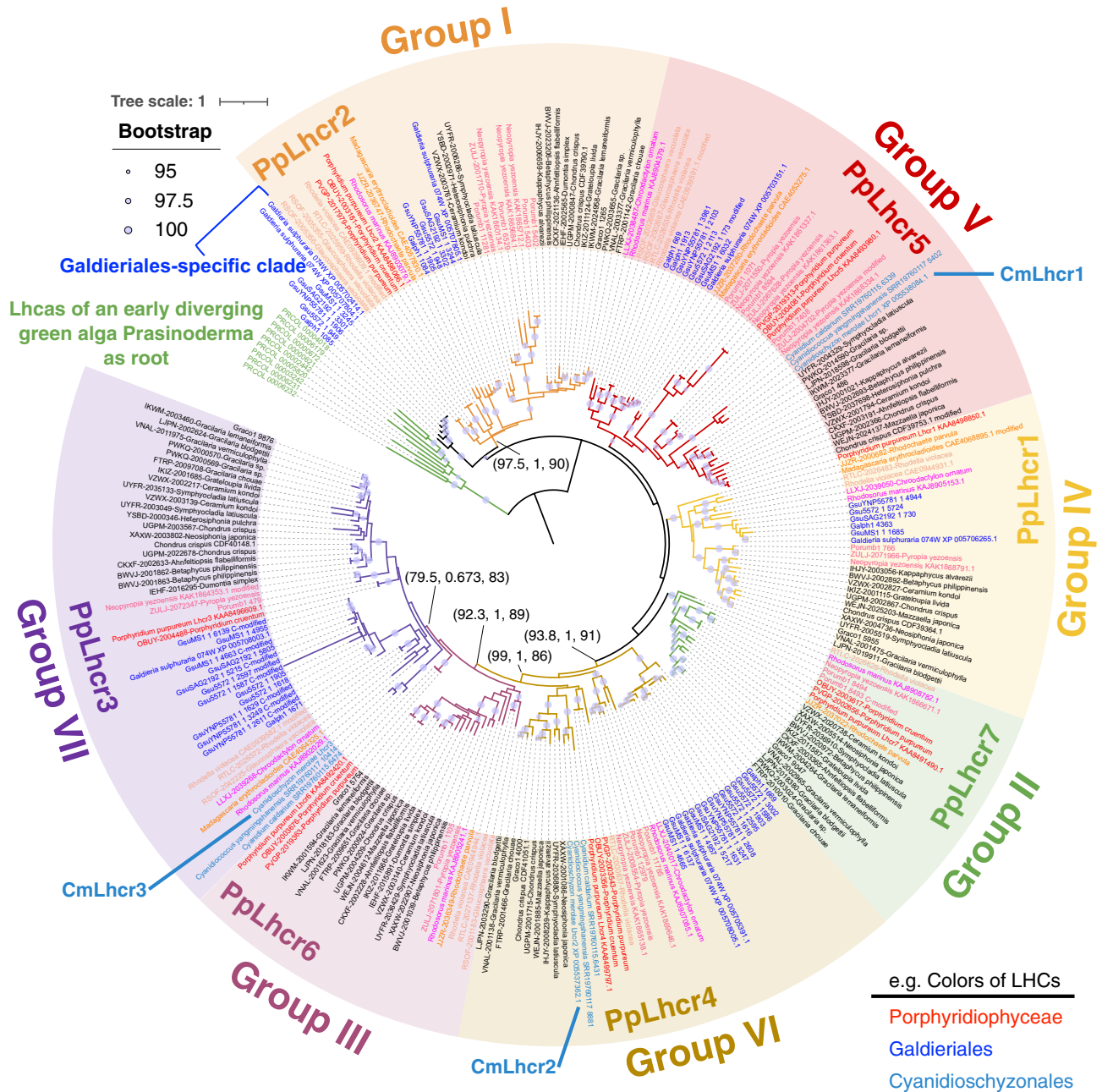
Figure 2. Molecular phylogenetic tree of Lhcr in red algae
The phylogenetic tree was constructed using IQ-TREE2 and rooted at the outgroup of Lhcas of an early diverging green alga, *Prasinoderma coloniale.* A total of 261 sequences with 341 amino acid sites were used, and the Q.pfam+F+I + R7 model was selected according to Bayesian information criterion scores. Circles on the node indicate ultrafast bootstrap support (≥95%). Numbers in parentheses are SH-aLRT support (%)/aBayes support/ultrafast bootstrap support (%). LHC colors correspond to the taxonomy in Figure 1. Short names of molecular species are described in Table S1. For example, *C. gracilis* Lhcr1 is shown as CgLhcr1.

of Rhodellophyceae, Bangiophyceae, and Florideophyceae.[28,30] The latter group, Cyanidiophyceae, contains the orders Galdieriales, Cavernulicolales, Cyanidiales, and Cyanidioschyzonales.[29] Among them, Galdieriales is considered as the earliest diverged order. Genomes or transcriptomes are available for all orders except Cavernulicolales; further, LHC sequences could be obtained through homology searches. Additionally, we obtained the sequences of putative LHCI (Lhca) associated with PSI in *Prasinoderma coloniale*—a member of the Prasinodermophyta class representing the earliest divergence within the primary green lineage.[31]

Next, a molecular phylogenetic tree was constructed using the obtained Lhcr sequences from red algae and Lhca sequences from the green alga *P. coloniale* (Figure 2). This is the first comprehensive report of the molecular phylogeny of red algal Lhcrs. In Figure 2, there
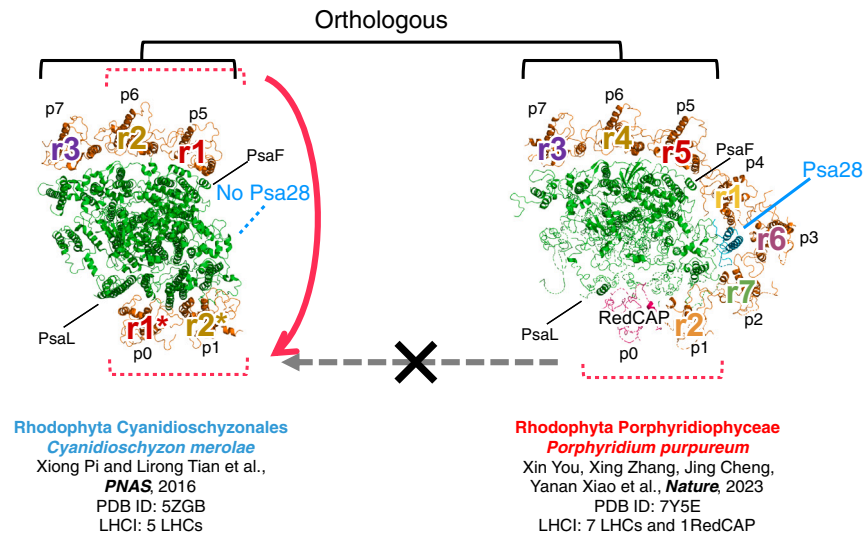
Figure 3. Estimated scheme of PSI–LHCI evolution in *Cyanidioschyzon*

The left PSI–LHCI structure corresponds to the Cyanidioschyzonales *Cyanidioschyzon merolae* and the right PSI–LHCI to the Porphyridiophyceae *Porphyridium purpureum*. The names of LHCIs are adapted from the original papers (Xiong Pi and Lirong Tian et al., 2016; Xin You, Xing Zhang, Jing Cheng, Yanan Xiao et al., 2023) and colored according to Figure 2.

are seven groups of Lhcrs (group I–VII) from red algae belonging to Rhodophytina, including *Porphyridium purpureum*. Each group contains one *P. purpureum* Lhcr; group VII, III, and VI containing PpLhcr3, PpLhcr4, and PpLhcr6 show monophyly, and group III and VI are sister groups to group VII. However, the distribution of Cyanidiophyceae Lhcrs is different: Galdieriales, belonging to Cyanidiophyceae, has Lhcrs classified into six including group I and IV–VII in addition to the Galdieriales-specific clade. This unique clade is a sister clade to group I. Galdieriales lacks group II and III LHCs. Cyanidiales and Cyanidioschyzonales have only three orthologs to those of Rhodophytina, and these Lhcrs— *Cyanidioschyzon merolae* Lhcr1 (hereafter CmLhcr1), CmLhcr2, and CmLhcr3— belong to group V, VI, and VII, respectively.

PSI-LHCI in "primitive red algae" cyanidioschyzonales

In a structural model of PSI–LHCI for *Cyanidioschyzon merolae* belonging to Cyanidioschyzonales,[3] five or three LHCs were attached to PSI. In *C. merolae* PSI–LHCI with five LHCIs, two copies of CmLhcr1 (r1 and r1*) and CmLhcr2 (r2 and r2*) in addition to one CmLhcr3 (r3) are associated with PSI (Figure 3). In PSI–LHCI with three LHCIs, CmLhcr1*, and CmLhcr2* are absent.[3,32] This may suggest that *C. merolae* would have 5LHCIs and 3LHCIs around PSI *in vivo*. The structure of the phycobilisome–PSII–PSI–LHCI megacomplex (Figure 3) has been reported in another red alga, *P. purpureum* belonging to Porphyridiophyceae[4]; it has eight LHCIs (PpLhcr1–7 and RedCAP) around PSI.

From the stromal side, in the *P. purpureum* PSI–LHCI part, positions of eight molecular LHCI species (r1–r7, and RedCAP) are labeled counterclockwise as positions 0–7 (p0–p7) (Figure 3). Accordingly, the positions of LHCI in *C. merolae* is labeled as p0, p1, and p5–p7 for two CmLhcr1 (r1 and r1*), two CmLhcr2 (r2 and r2*), and CmLhcr3 (r3).[3] *C. merolae* PSI–LHCI does not have an LHC at positions p2–p4. The loss of Psa28 (also called PsaR) in the genome of *C. merolae* implies that Psa28 would be crucial for LHCI binding at p2–p4.[4]

The Lhcrs at p5–p7 of both species seem to be conserved from their common ancestor. Our phylogenetic analysis suggests that CmLhcr1–3 at p5–p7 in *C. merolae* PSI–LHCI belong to groups V, VI, and VII, respectively, and that PpLhcr5, PpLhcr4, and PpLhcr3 at p5–p7 in *P. purpureum* PSI–LHCI belong to group V, VI, and VII, which are orthologs to CmLhcr1–3 (Figures 2 and 3). In contrast, significant changes are observed between the two species in Lhcrs at p0 and p1 (Figure 3): *P. purpureum* PSI binds RedCAP at p0, while that of *C. merolae* binds CmLhcr1 at p0 in addition to p5; *P. purpureum* binds PpLhcr2 (group I Lhcr) at p1, while *C. merolae* binds CmLhcr2 (group VI Lhcr) at p1 as well as p6. Importantly, PpLhcr2 and CmLhcr2 are not orthologs in the phylogenetic tree, although they bind at the same position; instead, CmLhcr2 and PpLhcr4 at p6 belong to group VI and they show an orthologous relationship in the tree. Given that RedCAP is conserved in Galdieriales which is an early diverged taxa of Cyanidiophyceae, RedCAP is lost in *C. merolae*.[22,23] *C. merolae* lost RedCAP, group I–IV LHCs, and a related PSI subunit, and complemented positions p0 and p1 with group V and VI LHCIs to, at least partially, maintain the antenna size for PSI.

Molecular phylogeny of Lhcrs in red-lineage algae

Cryptophytes do not possess the Lhcf subfamily or phycobilisome, but they do contain phycobiliproteins in the thylakoid lumen and LHCs in the thylakoid membrane.[33] Cryptophyte LHC is called chlorophyll *a/c* (CAC) proteins or alloxanthin-chlorophyll *a/c* proteins (ACPs) named after their pigments and bound to both PSI and PSII.[19,34,35] The PSI–LHCI and PSII–LHCII of diatoms and Haptophytes are called PSI–FCPI and PSII–FCPII, respectively, because of the bound pigments in LHCs. The classification of LHCs in red-lineage secondary endosymbiotic algae was described in our previous research.[15] The diatom FCPI comprises Lhcrs as well as the Lhcqs, CgLhcr9, which is distinct from the Lhcr
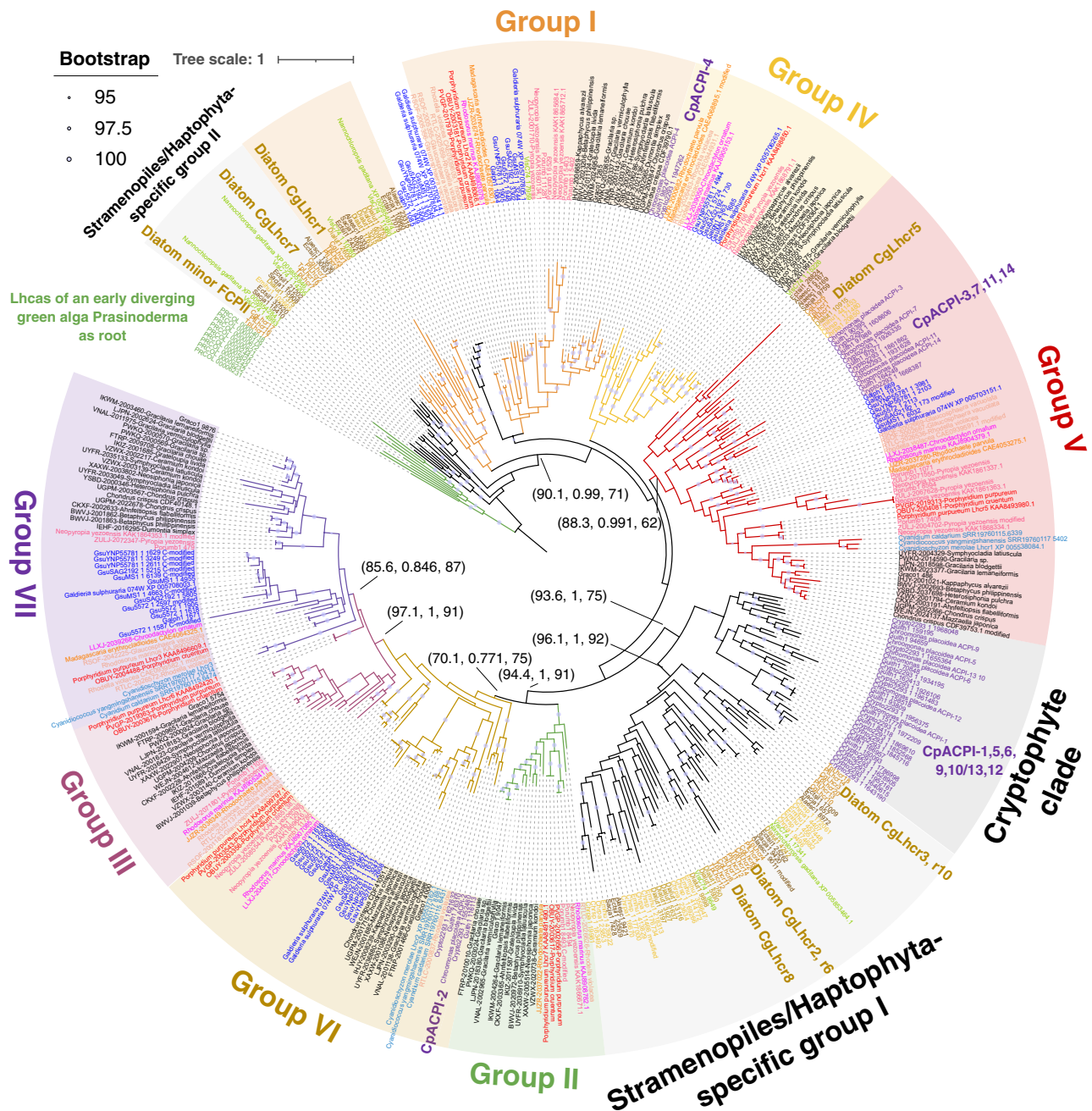
Figure 4. Molecular phylogenetic tree of Lhcr of red-lineage algae

The phylogenetic tree was inferred using IQ-TREE2 and rooted at the outgroup of Lhcas from an early diverging green alga, *Prasinoderma coloniale*. A total of 418 sequences with 489 amino acid sites were used, and the Q.pfam+F+R7 model was selected according to Bayesian information criterion scores. Circles on the node indicate ultrafast bootstrap support (≥95%). Numbers in parentheses are SH-aLRT support (%)/aBayes support/ultrafast bootstrap support (%). LHC colors correspond to the taxonomy in Figure 1. Short names of molecular species are described in Table S1.

subfamily, presumably some Lhcfs, and a RedCAP.[9,12,15,36] Unlike red algal PSII with phycobilisome, diatoms predominantly have the Lhcfs located around PSII, and one unique Lhcr (CgLhcr17 homolog) closely associated with the PSII core.[10,11,13,15,36] Molecular phylogenetic analysis suggests that Haptophytes also possess LHC subfamily compositions similar to diatoms, implying that they may have analogous PSI–FCPI and PSII–FCPII.[15] To elucidate how the Lhcrs in these algae were generated from red algae during secondary endosymbiosis, LHC sequences were obtained from a wide variety of Cryptophytes, Stramenopiles, and Haptophytes in addition to red algae. The molecular phylogenetic tree was inferred using these red-lineage Lhcrs in addition to red algal Lhcrs with *Prasinoderma* Lhcas as root of the tree (Figure 4).

LHCIs of Cryptophytes and Stramenopiles show different conservation patterns in the phylogenetic tree, suggesting different evolutionary processes of LHCIs (Figure 4). The phylogenetic tree of red-lineage LHCIs includes group I–VII in addition to one Cryptophyte-specific clade and two Stramenopiles/Haptophyta-specific clades. Cryptophytes, including *Chroomonas placoidea*, have ACPIs in groups IV–VI in addition to the Cryptophyte-specific clade. Group IV and VI contain CpACPI-4 and CpACPI-2, respectively. Group V contains not only CpACPI-3 but also CpACPI-7, 11, and 14. Groups I–III and VII are absent in Cryptophytes. The Cryptophyte-specific clade contains CpACPI-1, 5, 6, 9, 10/13, and 12; ACPI-6 and ACPI-10/13 are closely related and form one group, whereas CpACPI-1, 5, 9 forms the other.

Stramenopiles, including the diatom *Chaetoceros gracilis*, have CgLhcr1 in group I and CgLhcr5 in group V in addition to LHCIs in two Stramenopiles/Haptophyta-specific groups. The clade including CgLhcr1 is a sister clade of the group I clade of red algae, suggesting that this clade can be included in group I. One Stramenopiles/Haptophyta-specific group contains CgLhcr2, r3, r6, r8, and r10, while the other includes CgLhcr7 and r17. CgLhcr17 is a monomeric FCPII directly associated with PSII.[11,15] The Lhcrs of other Stramenopiles showed the same distribution pattern in the groups as that of diatom Lhcrs. Group I includes Pavlovales Lhcrs from Haptophytes, whereas Stramenopiles/Haptophyta-specific groups possess Lhcrs from other taxa of Haptophyta. Altogether, Lhcrs of red-lineage secondary endosymbiotic algae are present not only in red algal Lhcr groups but also in unique clades in the phylogenetic tree.

Molecular phylogeny of RedCAPs

RedCAP is one of the families belonging to the LHC superfamily directly bound to red-lineage PSI.[22,23] RedCAP is assigned in the position 0 in PSI–LHCI of Rhodophytina *Porphyridium purpureum*, Cryptophyta *Chroomonas placoidea*, *Rhodomonas salina*, a diatom *Chaetoceros gracilis*.[4,12,19,37] To investigate the evolutionary history of RedCAP, we collected RedCAP sequences using BLASTP from various public databases and inferred the molecular phylogenetic tree of RedCAP (Figure 5). In red algae, only Galdieriales from Cyanidiophyceae and all orders from Rhodophytina conserved one copy of RedCAP, and other red-lineage algae—Cryptophyta, photosynthetic Stramenopiles (Heterokonts), and Haptophyta—conserved RedCAP. The RedCAPs of secondary endosymbiotic algae showed a monophyletic clade between the Galdieriales and the Rhodophytina clades. Within the clade of secondary endosymbiotic RedCAPs, the Cryptophyta clade occupies the most basal position. Haptophyta RedCAPs and Stramenopile RedCAPs form sister clades, though the bootstrap support for monophyly of Stramenopile RedCAPs was not high (45%). While Dinoflagellates have red-lineage PSI–LHCI, they do not possess RedCAPs in the position 0.[38,39]

Cryptophyte PSI–ACPI and diatom PSI–FCPI showing LHCI rearrangement

Conservation and replacement of LHCIs (ACPIs) from ancestral red algae can be estimated when we combined the structure of Cryptophyte *Chroomonas placoidea* PSI–ACPI with the phylogenetic analysis. *C. placoidea* PSI–ACPI contains one RedCAP molecule and 13 Lhcrs (Figure 6).[19] It also has Psa28, an unknown subunit and ACPI-S. Psa28 are conserved from red algae (Galdieriales and Rhodophytina) to secondary endosymbiotic algae including Cryptophyta, Stramenopiles, Haptophyta (e.g., *Diacronema lutheri* GenBank: KAG8457475.1 and *Gephyrocapsa oceanica* GenBank: CAE5900733.1), and even Dinoflagellates.[4,9,12,19,38,39] Psa28 in PSI–LHCI in red algae, Cryptophyta, diatoms, and Dinoflagellates, seems to interconnect PSI core and LHCI in p2–4 as discussed in *P. purpureum* PSI–LHCI.[4] ACPI-S appears to interconnect ACPIs and is not found in red algal PSI–LHCI and diatom PSI–FCPI. From the stromal side, the binding positions of LHCI (ACPIs) can be assigned as p0–p8 similar to the red algal PSI–LHCI (Figure 3). RedCAP locates at p0 as observed in *P. purpureum*. ACPI-4, 3, and 2 at p4, p5, and p6 belong to groups IV, V, and VI, respectively, as did LHCIs in *P. purpureum* PSI–LHCI. In contrast, ACPI-7 at p1 belongs to group V in *C. placoidea,* while p1 is occupied by group I PpLhcr2 in *P. purpureum* PSI–LHCI. Furthermore, p2, p3, and p7 are occupied by ACPI-6, -5, and -1 belonging to the Cryptophyte-specific clade.

Interestingly, *Chroomonas placoidea* PSI–ACPI has three sets of adjacent three ACPIs (heterotrimer): ACPI-7–6–5, ACPI-11–10/13–9, and 14–10/13–12 (Figure 6). ACPI-7, 11, and 14 in each heterotrimer belong to group V. ACPI-3 belonging to group-V ACPIs would be rather ancestral LHCI because it binds at p3, meaning a true ortholog of group-V LHCI in red algae (ex. PpLhcr5). ACPI-7, 11, and 14 belonging to group V should be derivatives of ancestral ACPI-3. Other two ACPIs in each heterotrimer belong to the unique Cryptophyte-specific clade, wherein *C. placoidea* ACPIs can be divided into two groups by focusing on monophyly: ACPI-6, 10/13, and ACPI-5, 9, 12 (Figure 4). Thus, three sets of heterotrimers comprise three ACPIs from the respective groups, which would result from gene duplication.

According to the aforementioned analyses, the following model can be proposed for the establishment of Cryptophyte PSI–ACPI (Figure 7): after losing many Lhcr genes during secondary endosymbiosis of red algae, Cryptophytes only preserved three molecular species of Lhcr (ACPI-2, ACPI-3, and ACPI-4) and a RedCAP; at this point, the group V LHCI—an ancestral ACPI-3 at p5—diversified forming a Cryptophyte-specific clade in the Lhcr subfamily: one Cryptophyte-specific Lhcrs binds at p7 as ACPI-1; three sets of heterotrimers of ACPIs, including one group V Lhcr and two Cryptophyte-specific Lhcrs, bind to restore the antenna size of PSI. This molecular evolutionary model of *C. placoidea* PSI–ACPI (Figure 7) contradicts the current model of PSI–LHCI complex evolution in the red-lineage, solely based on LHCI compositions.[19] That is, Cryptophyte PSI–ACPI is not an evolutional intermediate between red algal PSI–LHCI and diatom PSI–FCPI.

Not all Lhcrs in a diatom PSI–FCPI directly descend from red algal LHCI. The diatom *Chaetoceros gracilis* PSI–FCPI possesses either none or one RedCAP molecule and 16 or 23 FCPs including eight Lhcrs.[9,12] In the structure of *C. gracilis* PSI–FCPI, positions p0–p7 are occupied by CgRedCAP, and CgLhcr1–r7. CgRedCAP at p0 is homologous to PpRedCAP at p0 of *P. purpureum* PSI–LHCI. CgLhcr1 and CgLhcr5, belonging to groups I and V, respectively, are positioned in p1 and p5, consistent with LHCIs belonging to groups I and V at p1 and p5 in red algal PSI–LHCI. However, all other positions except for p4 are occupied by Lhcrs belonging to
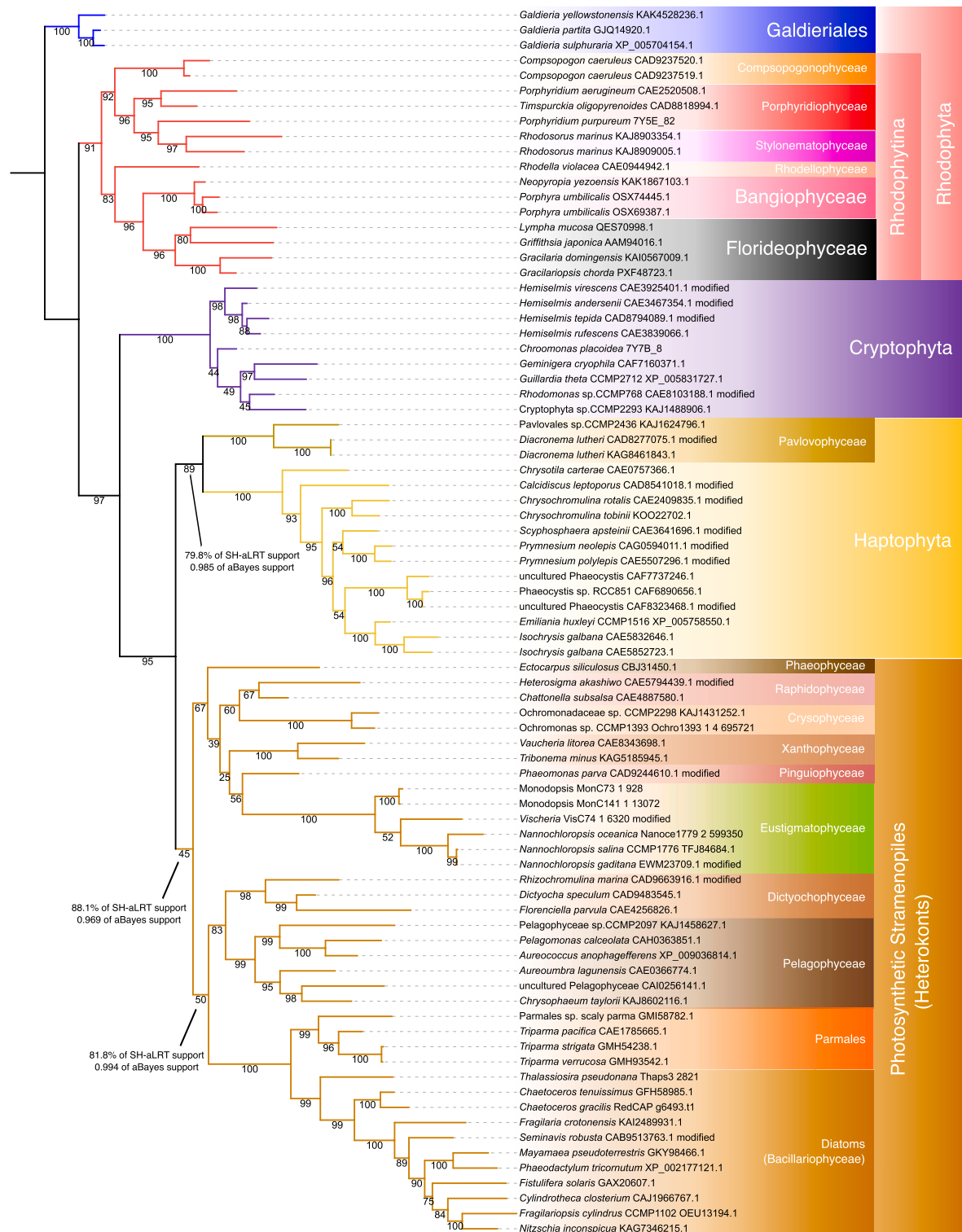
Tree scale: 1

Figure 5. Molecular Phylogenetic Tree of RedCAP

The phylogenetic tree was constructed using IQ-TREE2 and rooted at the outgroup of the Galdieriales RedCAP clade. Numbers on nodes are ultrafast bootstrap support (%).
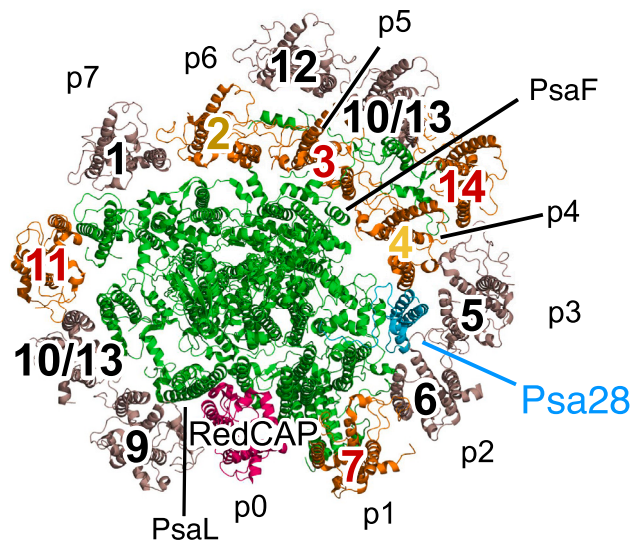
Figure 6. PSI–LHCI structures of Cryptophyte *Chroomonas placoidea* PSI–ACPI supercomplex
Top view of the supercomplex from the stromal side. The number on the structure corresponds to that of CpACPI proteins. The colors of ACPI numbers correspond to the groups in Figure 4. PSI core subunits except Psa28 (alternatively known as PsaR) are shown in green and Psa28 is cyan, respectively. RedCAP protein is shown in red and Lhcr proteins in LHCI orthologous/homologous to the red algal Lhcrs are colored orange. Lhcrs belonging to the Cryptophyte-specific clade are shown in brown. PsaL and PsaF subunits are indicated as PsaL and F, respectively.

Stramenopiles/Haptophyta-specific groups. Furthermore, CgLhcr8 and CgLhcr10, belonging to the Stramenopiles/Haptophyta-specific group I, are positioned counterclockwise adjacent to CgLhcr7 (positions p8 and p9, respectively). The FCPI assigned at p4 in *C. gracilis* PSI–FCPI is CgLhcr4, belonging to the Lhcq subfamily.[9,15] Altogether, the diatom PSI–FCPI shares orthologous LHCIs only at p0, p1, and p5 with the red algal PSI–LHCI. This suggests that diatoms have lost many LHCIs from red algae and replaced them with diversified FCPIs during endosymbiosis of red algae.

DISCUSSION

The molecular assembly model of the red-lineage PSI–LHCI has been discussed only based on spatial arrangements of the subunits in the structures.[19,20] However, this study made it clear that it is necessary to consider an evolutionary model of the photosynthetic supercomplex using both molecular phylogeny and structural information. Here, we would like to propose an evolutionary trajectory of LHCI proteins associated with PSI in red-lineage algae.

Putative composition of LHCIs in the common ancestor of primitive red algae

The smaller number of LHCI in *C. merolae* PSI can be due to genome reduction in Cyanidioschyzonales, including *C. merolae*, and Cyanidiales, resulting in small genome sizes and numerous gene deletions.[40] Rhodophytina and Cyanidiophyceae diverged from the common ancestor of red algae. Rhodophytina had a conserved set of Lhcrs: groups I–VII. However, Cyanidiophyceae, Cyanidioschyzonales, and Cyanidiales, have only three LHCIs which belong to groups V–VII. Galdieriales have six Lhcr molecular species belonging to groups I, IV–VII, and the Galdieriales-specific clade. Furthermore, Rhodophytina and Galdieriales possess RedCAP in their genomes, whereas Cyanidioschyzonales and Cyanidiales do not.[22,23] This suggests that only a few Lhcr genes remained after genome reduction in Cyanidioschyzonales and Cyanidiales, and that they would ensure PSI antenna size by repurposing the same genetic products, CmLhcr1 and CmLhcr2, to different position around PSI.

Considering that Galdieriales is the earliest order diverged from others in Cyanidiophyceae, it is reasonable to deduce the LHCI composition of the last common ancestor of red algae from those of Rhodophytina and Galdieriales. Since both Rhodophytina and Galdieriales possess RedCAP, the ancestral Rhodophyta likely had RedCAP at p0. Galdieriales has Lhcrs of group I, IV–VII, which are orthologous Lhcrs in *P. purpureum* PSI–LHCI bound at p1, p4–p7, while it lacks group II and III Lhcrs. Galdieriales conserves the Psa28 (RefSeq: XP_005707993.1, called PsaR, synonymously) subunit, which was suggested to stabilize LHCIs at p2–p4 in *P. purpureum* PSI–LHCI.[4] In Galdieriales, Psa28 may help the association of at least LHCI at p4. Thus, the ancestral Rhodophyta should have a conserved fundamental Lhcrs–RedCAP composition with RedCAP at p0 and Lhcrs at least at p1, and p4–p7 (Figure 8). This also suggests that *C. merolae* PSI–LHCI is not a "primitive" PSI–LHCI.

Evolutionary path of LHCI from red algae to stramenopiles/haptophyte through Cryptophyta

The evolutionary model of Cryptophyta PSI–ACPI introduced in Figure 7 suggests its progression from the ancestral red algae PSI–LHCI. Cryptophyta acquired a plastid from red algae, having its endosymbiotic nucleus, the nucleomorph, derived from red algae.[41,42] This supports
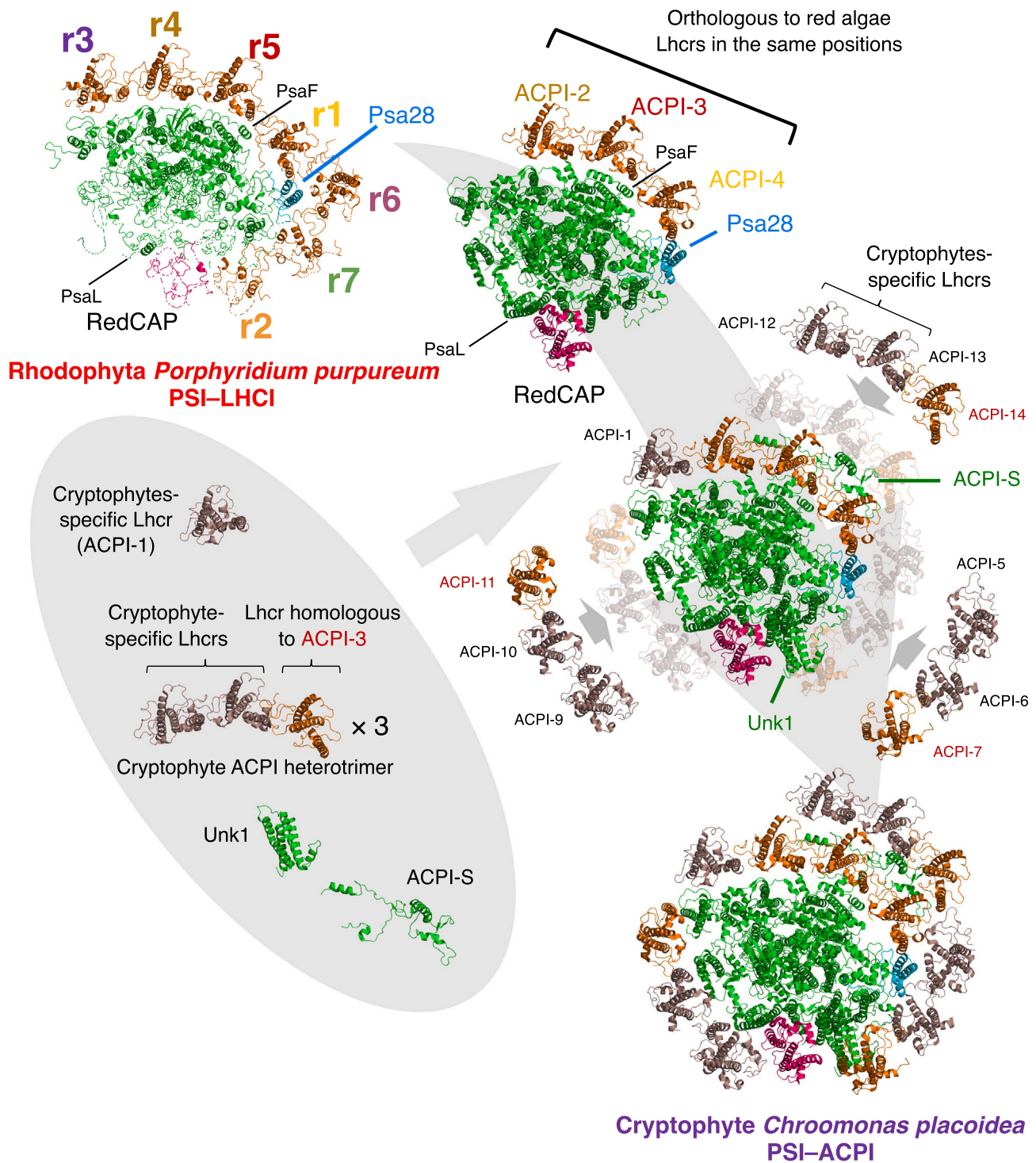
Figure 7. Estimated trajectory of PSI–LHCI evolution of Cryptophytes after endosymbiosis with red alga based on the molecular phylogeny of LHCI

The PSI–ACPI of Cryptophytes *Chroomonas placoidea* has one RedCAP and 13 LHCs, including five Lhcrs homologous to red algal Lhcrs and eight Cryptophyte-specific Lhcrs. Color code is the same as in Figure 6. The cryptophyte PSI–ACPI conserved the PSI core including Psa28 with RedCAP and ACP-4, 3, and 2 at positions p0, p4–p6, respectively, as the putative PSI–LHCI in ancestral red alga. ACP-2–4 are orthologous to PpLhcr1, r5, and r4, respectively. During the endosymbiosis event, an unknown subunit (Unk1) and ACPI-S were added to the conserved ancestral set of PSI–ACPI. In addition, a Cryptophyte-specific Lhcr at ACPI-1 and three ACPI heterotrimers were attached to three different sides of PSI–ACPI.
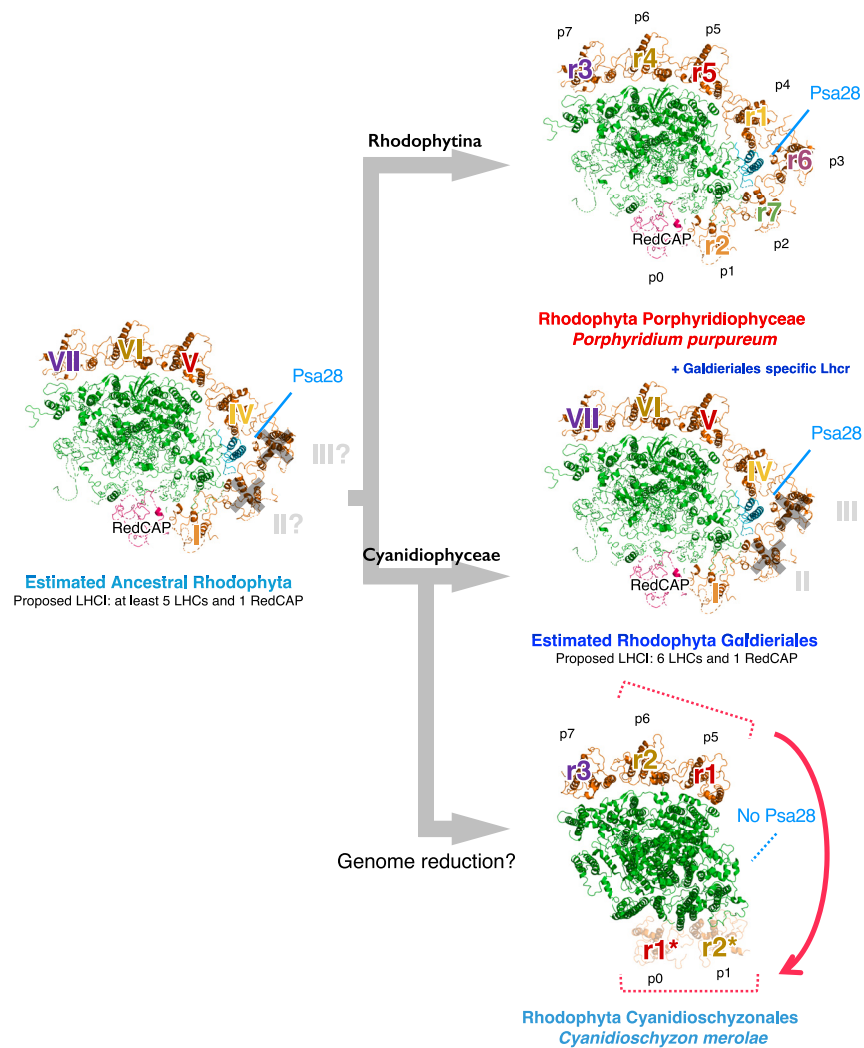
Figure 8. LHCI evolutionary model of red algae based on phylogeny and structures

The estimated ancestral Rhodophyta and Galdieriales PSI–LHCI are depicted based on the *P. purpureum* PSI–LHCI structural model. PSI core subunits except Psa28 (alternatively known as PsaR) are shown in green and Psa28 is cyan. RedCAP protein is shown in red and Lhcr proteins in LHCI are colored orange. The numbers of Lhcr groups are indicated on the respective estimated models. The PSI–LHCI models for *P. purpureum* and *C. merolae* are consistent with those shown in Figure 2.

the direct descendance of Cryptophyte PSI–ACPI from red algae PSI–LHCI, without additional endosymbiosis. The ancestral red algae PSI–LHCI, featuring LHCI at p0, 1, and p4–p7, shares positions p0, p4–p5 with Cryptophyte PSI–ACPI. Despite losing group I Lhcr, Cryptophyta has diversified its Lhcrs not only to fill position p7 but also to expand its antenna size by creating three heterotrimers. This research meticulously details the evolution of LHCI from primitive red algae to Cryptophytes, presenting Cryptophytes as a prime example of extensive LHCI rearrangement and antenna enlargement of PSI through endosymbiosis.

In contrast, the evolutionary trajectory of Stramenopiles and Haptophyte LHCI is not as straightforward as that of Cryptophyte ACPIs. When considering the evolutionary model of PSI–LHCI based on molecular phylogeny, Stramenopiles, including diatoms and Haptophytes, acquired the Lhcq subfamily in addition to the Lhcr subfamily.[15] In diatom FCPI, only group I and V Lhcrs (CgLhcr1 and CgLhcr5) at p1 and p5, respectively, have conserved positions from red algae. Other Stramenopiles and at least Pavlovales in Haptophyta share group I and V Lhcrs. In diatom PSI–FCPI, Lhcrs at p2, p3, and p6–p9 belong to the Stramenopiles/Haptophyta-specific groups.[9,15] CgLhcr4, which does not belong to the Lhcr subfamily but to the Lhcq subfamily, is assigned to p4 in diatom PSI–FCPI. Moreover, the Lhcq subfamily in PSI–FCPI not only binds directly to the PSI core but also forms the outer layer of LHCs in PSI–FCPI, contributing to a larger antenna size of the diatom PSI.[9,12,15] This indicates that Stramenopiles has experienced a distinct process of LHCI re-acquisition around PSI, which would be largely different from that of Cryptophytes.

The evolutionary history from red algae to red-lineage algae has been previously described.[43] Linear regression analysis on nuclear genomes suggested serial endosymbiosis in red-lineage algae: Cryptophyta engulfed a red alga as secondary (2nd) endosymbiosis; the

photosynthetic Stramenopiles incorporated Cryptophyta as tertiary (3rd) endosymbiosis; Haptophyta acquired many genes from Stramenopiles as quaternary (4th) endosymbiosis. However, the plastid-encoded genes of Haptophyta and Cryptophyta strongly support monophyly in the phylogenetic tree.[27] Haptophyta may have acquired the ancestral Cryptophyta plastid before or after massive gene transfer from Stramenopiles, reconciling the two seemingly contradictory phylogenetic trees of different genomes.[44–46] LHC genes are nuclear-encoded and should follow the history of nuclear-encoded genes derived from the quaternary endosymbiosis.[43,44] RedCAPs are nuclear-encoded and form a monophylic clade including Stramenopiles and Haptophytes, and Haptophyta RedCAPs are sister to the entire Stramenopile clade. This is consistent with the serial endosymbiosis hypothesis for Haptophytes. However, monophyly of Stramenopiles is weak. This makes it difficult to determine whether Haptophyta acquired RedCAP from the common ancestor of Stramenopiles or from a specific Stramenopile lineage. Consistently, the composition of LHC subfamilies of Haptophyta is similar to that of Stramenopiles.[15] In contrast, most genes coding for the PS core are encoded in the plastid genome originated from Cryptophyta. Based on these facts, we hypothesize that the origin of the genes coding for the Haptophyte PS core complex and that for light-harvesting antennas can be chimeric. Further genetic and structural analysis of the PSI–LHCI in Haptophytes is required to confirm/dismiss this hypothesis.

Molecular evolution of PS supercomplexes through "neolocalization"

Gene duplication and functional diversification, for instance in biosynthetic enzyme families, are typically referred to as "neofunctionalization" (e.g., Hansen et al. 2021[47]). However, even after intensive diversification, the primary function of the LHC family and RedCAP remains light-harvesting. Therefore, their diversification can be best considered as that of structural relocalization rather than function. In our study on the red alga Cyanidiales *Cyanidium caldarium* PSI–LHCI, we proposed the term "neolocalization".[48] It was defined as a phenomenon where a structural defect caused by gene loss is complemented or modified by the product of another existing gene. In this study, we expand the definition of neolocalization to include modifications by the product of a gene differentiated after duplication from an existing one. Phenomena matching neolocalization are also observed in the green lineage. In green algae *Chlamydomonas reinhardtii*, green algal Lhca7 and terrestrial plant Lhca2 do not share orthology, yet they bind at the same position in green-lineage PSI–LHCIs.[21,49–52] Similarly, it is reported that Arabidopsis can attach AtLhca1 and AtLhca4 to state-2 PSI in the position of Chlamydomonas Lhca9 and Lhca2 to PsaB/H/I/B/G side of PSI core.[53] These facts suggest that neolocalization primarily complements antenna size of PSI–LHCI upon gene loss triggered by genomic reduction, endosymbiotic events, or other reasons, while it may also contribute to light-environmental adaptation.

Neolacalization following to genomic events has made the evolutionary history of LHCI complicated. Therefore, simple conventional models solely based on compositions in LHCI may lead to incorrect conclusions about the evolutionary trajectory of LHCI. To understand this complex process, a general model of molecular evolution must consider both molecular phylogeny and structure. At present, the complete evolutionary paths of LHC diversification and differentiation from Rhodophytina red algae to Stramenopiles and Haptophyta through Cryptophyta remain to be elucidated. As more genomic information and more structural models become available, the relationship between these taxa will become clearer, allowing the construction of an entire evolutionary model of PS supercomplexes.

Limitations of the study

By combining molecular phylogenetic analyses with structural information of PSI–LHCI in red-lineage algae, we estimate the evolutionary trajectory of LHCI, especially within red algae and from red algae to Cryptophytes. However, considering that the structure of PSI–LHCI has not been reported for all major lineages of red algae, our estimations presented in this study require further verification in future research. In addition, this study is limited to the Lhcr subfamily and RedCAP, while other subfamilies, such as Lhcq and Lhcf, also comprise LHCI in diatoms and possibly in other Stramenopiles and Haptophytes. Further analysis of Lhcq and Lhcf would contribute to estimating LHCI evolution within Stramenopiles and Haptophytes.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Dr. Kentaro Ifuku (ifuku.kentaro.2m@kyoto-u.ac.jp).

Materials availability

Materials generated in this study are included in Supplemental Data.

Data and code availability

Data used in this study are from public databases. Accessions are listed in Table S1. This paper does not report the original code. Any additional information required to reanalyze the data reported in this study is available from the lead contact upon request.

ACKNOWLEDGMENTS

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - LHC protein sequence acquisition
  - Phylogenetic analysis of LHC
  - Phylogenetic analysis of RedCAP
  - Visualization of PSI–LHCI structure
- QUANTIFICATION AND STATISTICAL ANALYSIS

REFERENCES

1. Delwiche, C.F. (1999). Tracing the Thread of Plastid Diversity through the Tapestry of Life. Am. Nat. *154*, S164–S177. https://doi.org/10.1086/303291.

2. Croce, R., and van Amerongen, H. (2020). Light harvesting in oxygenic photosynthesis: Structural biology meets spectroscopy. Science *369*, eaay2058. https://doi.org/10.1126/science.aay2058.

3. Pi, X., Tian, L., Dai, H.E., Qin, X., Cheng, L., Kuang, T., Sui, S.F., and Shen, J.R. (2018). Unique organization of photosystem I–light-harvesting supercomplex revealed by cryo-EM from a red alga. Proc. Natl. Acad. Sci. USA *115*, 4423–4428. https://doi.org/10.1073/pnas.1722482115.

4. You, X., Zhang, X., Cheng, J., Xiao, Y., Ma, J., Sun, S., Zhang, X., Wang, H.-W., and Sui, S.-F. (2023). In situ structure of the red algal phycobilisome–PSII–PSI–LHC megacomplex. Nature *616*, 199–206. https://doi.org/10.1038/s41586-023-05831-0.

5. Wolfe, G.R., Cunningham, F.X., Grabowski, B., and Gantt, E. (1994). Isolation and characterization of Photosystems I and II from the red alga Porphyridium cruentum. Biochim. Biophys. Acta Bioenerg. *1188*, 357–366. https://doi.org/10.1016/0005-2728(94)90056-6.

6. Marquardt, J., and Rhiel, E. (1997). The membrane-intrinsic light-harvesting complex of the red alga Galdieria sulphuraria (formerly Cyanidium caldarium): biochemical and immunochemical characterization. Biochim. Biophys. Acta Bioenerg. *1320*, 153–164. https://doi.org/10.1016/S0005-2728(97)00020-0.

7. Koziol, A.G., Borza, T., Ishida, K.I., Keeling, P., Lee, R.W., and Durnford, D.G. (2007). Tracing the evolution of the light-harvesting antennae in chlorophyll a/b-containing organisms. Plant Physiol. *143*, 1802–1816. https://doi.org/10.1104/pp.106.092536.

8. Büchel, C. (2015). Evolution and function of light harvesting proteins. J. Plant Physiol. *172*, 62–75. https://doi.org/10.1016/j.jplph.2014.04.018.

9. Nagao, R., Kato, K., Ifuku, K., Suzuki, T., Kumazawa, M., Uchiyama, I., Kashino, Y., Dohmae, N., Akimoto, S., Shen, J.R., et al. (2020). Structural basis for assembly and function of a diatom photosystem I-light-harvesting supercomplex. Nat. Commun. *11*, 2481. https://doi.org/10.1038/s41467-020-16324-3.

10. Nagao, R., Kato, K., Suzuki, T., Ifuku, K., Uchiyama, I., Kashino, Y., Dohmae, N., Akimoto, S., Shen, J.-R., Miyazaki, N., and Akita, F. (2019). Structural basis for energy harvesting and dissipation in a diatom PSII–FCPII supercomplex. Nat. Plants *5*, 890–901. https://doi.org/10.1038/s41477-019-0477-x.

11. Nagao, R., Kato, K., Kumazawa, M., Ifuku, K., Yokono, M., Suzuki, T., Dohmae, N., Akita, F., Akimoto, S., Miyazaki, N., and Shen, J.R. (2022). Structural basis for different types of hetero-tetrameric light-harvesting complexes in a diatom PSII-FCPII supercomplex. Nat. Commun. *13*, 1764. https://doi.org/10.1038/s41467-022-29294-5.

12. Xu, C., Pi, X., Huang, Y., Han, G., Chen, X., Qin, X., Huang, G., Zhao, S., Yang, Y., Kuang, T., et al. (2020). Structural basis for energy transfer in a huge diatom PSI-FCPI supercomplex. Nat. Commun. *11*, 5081. https://doi.org/10.1038/s41467-020-18867-x.

13. Wang, W., Yu, L.J., Xu, C., Tomizaki, T., Zhao, S., Umena, Y., Chen, X., Qin, X., Xin, Y., Suga, M., et al. (2019). Structural basis for blue-green light harvesting and energy dissipation in diatoms. Science *363*, eaav0365. https://doi.org/10.1126/science.aav0365.

14. Umetani, I., Kunugi, M., Yokono, M., Takabayashi, A., and Tanaka, A. (2018). Evidence of the supercomplex organization of photosystem II and light-harvesting complexes in Nannochloropsis granulata. Photosynth. Res. *136*, 49–61. https://doi.org/10.1007/s11120-017-0438-z.

15. Kumazawa, M., Nishide, H., Nagao, R., Inoue-Kashino, N., Shen, J.R., Nakano, T., Uchiyama, I., Kashino, Y., and Ifuku, K. (2022). Molecular phylogeny of fucoxanthin-chlorophyll a/c proteins from *Chaetoceros gracilis* and Lhcq/Lhcf diversity. Physiol. Plantarum *174*, e13598. https://doi.org/10.1111/ppl.13598.

16. Bína, D., Gardian, Z., Herbstová, M., Kotabová, E., Koník, P., Litvín, R., Prášil, O., Tichý, J., and Vácha, F. (2014). Novel type of red-shifted chlorophyll a antenna complex from Chromera velia: II. Biochemistry and spectroscopy. Biochim. Biophys. Acta Bioenerg. *1837*, 802–810. https://doi.org/10.1016/j.bbabio.2014.01.011.

17. Zhao, S., Shen, L., Li, X., Tao, Q., Li, Z., Xu, C., Zhou, C., Yang, Y., Sang, M., Han, G., et al. (2023). Structural insights into photosystem II supercomplex and trimeric FCP antennae of a centric diatom Cyclotella meneghiniana. Nat. Commun. *14*, 8164. https://doi.org/10.1038/s41467-023-44055-8.

18. Feng, Y., Li, Z., Li, X., Shen, L., Liu, X., Zhou, C., Zhang, J., Sang, M., Han, G., Yang, W., et al. (2023). Structure of a diatom photosystem II supercomplex containing a member of Lhcx family and dimeric FCPII. Sci. Adv. *9*. https://doi.org/10.1126/sciadv.adi8446.

19. Zhao, L.S., Wang, P., Li, K., Zhang, Q.B., He, F.Y., Li, C.Y., Su, H.N., Chen, X.L., Liu, L.N., and Zhang, Y.Z. (2023). Structural basis and evolution of the photosystem I–light-harvesting supercomplex of cryptophytes

algae. Plant Cell *35*, 2449–2463. https://doi.org/10.1093/plcell/koad087.

20. Bai, T., Guo, L., Xu, M., and Tian, L. (2021). Structural Diversity of Photosystem I and Its Light-Harvesting System in Eukaryotic Algae and Plants. Front. Plant Sci. *12*, 781035. https://doi.org/10.3389/fpls.2021.781035.

21. Neilson, J.A.D., and Durnford, D.G. (2010). Structural and Functional Diversification of the Light-Harvesting Complexes in Photosynthetic Eukaryotes (Kluwer Academic Publishers). https://doi.org/10.1007/s11120-010-9576-2.

22. Engelken, J., Brinkmann, H., and Adamska, I. (2010). Taxonomic distribution and origins of the extended LHC (light-harvesting complex) antenna protein superfamily. BMC Evol. Biol. *10*, 233. https://doi.org/10.1186/1471-2148-10-233.

23. Sturm, S., Engelken, J., Gruber, A., Vugrinec, S., Kroth, P.G., Adamska, I., and Lavaud, J. (2013). A novel type of light-harvesting antenna protein of red algal origin in algae with secondary plastids. BMC Evol. Biol. *13*, 159. https://doi.org/10.1186/1471-2148-13-159.

24. Hoffman, G.E., Sanchez-Puerta, M.V., and Delwiche, C.F. (2011). Evolution of light-harvesting complex proteins from Chl c-containing algae. BMC Evol. Biol. *11*, 101. https://doi.org/10.1186/1471-2148-11-101.

25. Dittami, S.M., Michel, G., Collén, J., Boyen, C., and Tonon, T. (2010). Chlorophyll-binding proteins revisited - a multigenic family of light-harvesting and stress proteins from a brown algal perspective. BMC Evol. Biol. *10*, 365. https://doi.org/10.1186/1471-2148-10-365.

26. Yoon, H.S., Hackett, J.D., Pinto, G., and Bhattacharya, D. (2002). The single, ancient origin of chromist plastids. Proc. Natl. Acad. Sci. USA *99*, 15507–15512. https://doi.org/10.1073/pnas.242379899.

27. Kim, J.I., Moore, C.E., Archibald, J.M., Bhattacharya, D., Yi, G., Yoon, H.S., and Shin, W. (2017). Evolutionary Dynamics of Cryptophyte Plastid Genomes. Genome Biol. Evol. *9*, 1859–1872. https://doi.org/10.1093/gbe/evx123.

28. Yang, E.C., Boo, S.M., Bhattacharya, D., Saunders, G.W., Knoll, A.H., Fredericq, S., Graf, L., and Yoon, H.S. (2016). Divergence time estimates and the evolution of major lineages in the florideophyte red algae. Sci. Rep. *6*, 21361. https://doi.org/10.1038/srep21361.

29. Park, S.I., Cho, C.H., Ciniglia, C., Huang, T.Y., Liu, S.L., Bustamante, D.E., Calderon, M.S., Mansilla, A., McDermott, T., Andersen, R.A., and Yoon, H.S. (2023). Revised classification of the Cyanidiophyceae based on plastid genome data with descriptions of the Cavernulicolales ord. nov. and Galdieriales ord. nov. (Rhodophyta). J. Phycol. *59*, 444–466. https://doi.org/10.1111/jpy.13322.

30. Borg, M., Krueger-Hadfield, S.A., Destombe, C., Collén, J., Lipinska, A., and Coelho, S.M. (2023). Red macroalgae in the genomic era. New Phytol. *240*, 471–488. https://doi.org/10.1111/nph.19211.

31. Li, L., Wang, S., Wang, H., Sahu, S.K., Marin, B., Li, H., Xu, Y., Liang, H., Li, Z., Cheng, S., et al. (2020). The genome of Prasinoderma coloniale unveils the existence of a third phylum within green plants. Nat. Ecol. Evol. *4*, 1220–1231. https://doi.org/10.1038/s41559-020-1221-7.

32. Antoshvili, M., Caspy, I., Hippler, M., and Nelson, N. (2019). Structure and function of photosystem I in Cyanidioschyzon merolae. Photosynth. Res. *139*, 499–508. https://doi.org/10.1007/s11120-018-0501-4.

33. Spear-Bernstein, L., and Miller, K.R. (1989). Unique location of the phycobiliprotein light-harvesting pigment in the cryptophyceae[1]. J. Phycol. *25*, 412–419. https://doi.org/10.1111/j.1529-8817.1989.tb00245.x.

34. Kereïche, S., Kouřil, R., Oostergetel, G.T., Fusetti, F., Boekema, E.J., Doust, A.B., van der Weij-de Wit, C.D., and Dekker, J.P. (2008). Association of chlorophyll a/c2 complexes to photosystem I and photosystem II in the cryptophytes Rhodomonas CS24. Biochim. Biophys. Acta Bioenerg. *1777*, 1122–1128. https://doi.org/10.1016/j.bbabio.2008.04.045.

35. Kaňa, R., Kotabová, E., Sobotka, R., and Prášil, O. (2012). Non-photochemical quenching in cryptophytes alga Rhodomonas salina is located in chlorophyll a/c antennae. PLoS One *7*, e29700. https://doi.org/10.1371/journal.pone.0029700.

36. Calvaruso, C., Rokka, A., Aro, E.M., and Büchel, C. (2020). Specific Lhc Proteins Are Bound to PSI or PSII Supercomplexes in the Diatom *Thalassiosira pseudonana*. Plant Physiol. *183*, 67–79. https://doi.org/10.1104/pp.20.00042.

37. Zhang, S., Si, L., Su, X., Zhao, X., An, X., and Li, M. (2024). Growth phase-dependent reorganization of cryptophytes photosystem I antennae. Commun. Biol. *7*, 560. https://doi.org/10.1038/s42003-024-06268-5.

38. Li, X., Li, Z., Wang, F., Zhao, S., Xu, C., Mao, Z., Duan, J., Feng, Y., Yang, Y., Shen, L., et al. (2024). Structures and organizations of PSI–AcpPCI supercomplexes from red tidal and coral symbiotic photosynthetic dinoflagellates. Proc. Natl. Acad. Sci. USA *121*, e2315476121. https://doi.org/10.1073/pnas.2315476121.

39. Zhao, L.S., Wang, N., Li, K., Li, C.Y., Guo, J.P., He, F.Y., Liu, G.M., Chen, X.L., Gao, J., Liu, L.N., and Zhang, Y.Z. (2024). Architecture of symbiotic dinoflagellate photosystem I–light-harvesting supercomplex in Symbiodinium. Nat. Commun. *15*, 2392. https://doi.org/10.1038/s41467-024-46791-x.

40. Cho, C.H., Park, S.I., Huang, T.Y., Lee, Y., Ciniglia, C., Yadavalli, H.C., Yang, S.W., Bhattacharya, D., and Yoon, H.S. (2023). Genome-wide signatures of adaptation to extreme environments in red algae. Nat. Commun. *14*, 10. https://doi.org/10.1038/s41467-022-35566-x.

41. Van Der Auwera, G., Hofmann, C.J.B., De Rijk, P., and De Wachter, R. (1998). The Origin of Red Algae and Cryptomonad Nucleomorphs: A Comparative Phylogeny Based on Small and Large Subunit rRNA Sequences of Palmaria palmata, Gracilaria Verrucosa, and the Guillardia Theta Nucleomorph.

42. Archibald, J.M., Cavalier-Smith, T., Maier, U., and Douglas, S. (2001). Molecular chaperones encoded by a reduced nucleus: The cryptomonad nucleomorph. J. Mol. Evol. *52*, 490–501. https://doi.org/10.1007/s002390010179.

43. Stiller, J.W., Schreiber, J., Yue, J., Guo, H., Ding, Q., and Huang, J. (2014). The evolution of photosynthesis in chromist algae through serial endosymbioses. Nat. Commun. *5*, 5764. https://doi.org/10.1038/ncomms6764.

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

45. Dorrell, R.G., Villain, A., Perez-Lamarque, B., Audren de Kerdrel, G., McCallum, G., Watson, A.K., Ait-Mohamed, O., Alberti, A., Corre, E., Frischkorn, K.R., et al. (2021). Phylogenomic fingerprinting of tempo and functions of horizontal gene transfer within ochrophytes. Proc. Natl. Acad. Sci. USA *118*, e2009974118. https://doi.org/10.1073/pnas.2009974118.

46. Penot, M., Dacks, J.B., Read, B., and Dorrell, R.G. (2022). Genomic and meta-genomic insights into the functions, diversity and global distribution of haptophyte algae. Appl. Phycol. *3*, 340–359. https://doi.org/10.1080/26388081.2022.2103732.

47. Hansen, C.C., Nelson, D.R., Møller, B.L., and Werck-Reichhart, D. (2021). Plant cytochrome P450 plasticity and evolution. Mol. Plant *14*, 1244–1265. https://doi.org/10.1016/j.molp.2021.06.028.

48. Kato, K., Hamaguchi, T., Kumazawa, M., Nakajima, Y., Ifuku, K., Hirooka, S., Hirose, Y., Miyagishima, S.Y., Suzuki, T., Kawakami, K., et al. (2024). The structure of PSI-LHCI from *Cyanidium caldarium* provides evolutionary insights into conservation and diversity of red-lineage LHCs. Proc. Natl. Acad. Sci. USA *121*, e2319658121. https://doi.org/10.1073/pnas.2319658121.

49. Suga, M., Ozawa, S.I., Yoshida-Motomura, K., Akita, F., Miyazaki, N., and Takahashi, Y. (2019). Structure of the green algal photosystem I supercomplex with a decameric light-harvesting complex I. Nat. Plants *5*, 626–636. https://doi.org/10.1038/s41477-019-0438-4.

50. Ben-Shem, A., Frolow, F., and Nelson, N. (2003). Crystal structure of plant photosystem I. Nature *426*, 630–635. https://doi.org/10.1038/nature02200.

51. Ozawa, S.I., Bald, T., Onishi, T., Xue, H., Matsumura, T., Kubo, R., Takahashi, H., Hippler, M., and Takahashi, Y. (2018). Configuration of ten light-harvesting chlorophyll a/b complex I subunits in chlamydomonas reinhardtii photosystem I. Plant Physiol. *178*, 583–595. https://doi.org/10.1104/pp.18.00749.

52. Su, X., Ma, J., Pan, X., Zhao, X., Chang, W., Liu, Z., Zhang, X., and Li, M. (2019). Antenna arrangement and energy transfer pathways of a green algal photosystem-I–LHCI supercomplex. Nat. Plants *5*, 273–281. https://doi.org/10.1038/s41477-019-0380-5.

53. Crepin, A., Kučerová, Z., Kosta, A., Durand, E., and Caffarri, S. (2020). Isolation and characterization of a large photosystem I–light-harvesting complex II supercomplex with an additional Lhca1–a4 dimer in Arabidopsis. Plant J. *102*, 398–409. https://doi.org/10.1111/tpj.14634.

54. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

55. Edgar, R.C. (2022). Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. Nat. Commun. *13*, 6968. https://doi.org/10.1038/s41467-022-34630-w.

56. Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

57. Steenwyk, J.L., Buida, T.J., Li, Y., Shen, X.X., and Rokas, A. (2020). ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol. *18*, e3001007. https://doi.org/10.1371/journal.pbio.3001007.

58. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

59. Letunic, I., and Bork, P. (2024). Interactive Tree of Life (iTOL) v6: Recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. *52*, W78–W82. https://doi.org/10.1093/nar/gkae268.

60. Schrodinger LLC (2015). The PyMOL Molecular Graphics System, Version 1.8.

61. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745. https://doi.org/10.1093/nar/gkv1189.

62. Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2013). GenBank. Nucleic Acids Res. *41*, D36–D42. https://doi.org/10.1093/nar/gks1195.

63. Leinonen, R., Sugawara, H., and Shumway, M.; International Nucleotide Sequence Database Collaboration (2011). The sequence read archive. Nucleic Acids Res. *39*, D19–D21. https://doi.org/10.1093/nar/gkq1019.

64. Grigoriev, I.V., Hayes, R.D., Calhoun, S., Kamel, B., Wang, A., Ahrendt, S., Dusheyko, S., Nikitin, R., Mondo, S.J., Salamov, A., et al. (2021). PhycoCosm, a comparative algal genomics resource. Nucleic Acids Res. *49*, D1004–D1011. https://doi.org/10.1093/nar/gkaa898.

65. Steenwyk, J.L., Buida, T.J., Li, Y., Shen, X.X., and Rokas, A. (2020). ClipKIT: A multiple sequence alignment-trimming algorithm for accurate phylogenomic inference. Preprint at bioRxiv. https://doi.org/10.1101/2020.06.08.140384.

66. Filloramo, G.V., Curtis, B.A., Blanche, E., and Archibald, J.M. (2021). Re-examination of two diatom reference genomes using long-read sequencing. BMC Genom. *22*, 379. https://doi.org/10.1186/s12864-021-07666-3.

67. Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., Deyholos, M.K., Gitzendanner, M.A., Graham, S.W., Grosse, I., Li, Z., Melkonian, M., Mirarab, S., et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. Nature *574*, 679–685. https://doi.org/10.1038/s41586-019-1693-2.

68. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589. https://doi.org/10.1038/nmeth.4285.

69. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. *49*, W293–W296. https://doi.org/10.1093/nar/gkab301.

70. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321. https://doi.org/10.1093/sysbio/syq010.

71. Anisimova, M., Gil, M., Dufayard, J.F., Dessimoz, C., and Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. *60*, 685–699. https://doi.org/10.1093/sysbio/syr041.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Blastp	Altschul et al. (1990)[54]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Muscle5	Edgar (2022)[55]	https://www.drive5.com/muscle/
MAFFT E-INS-I v7.490 or v7.525	Katoh and Standley (2013)[56]	https://mafft.cbrc.jp/alignment/server/index.html
ClipKit v1.4.1	Steenwyk et al. (2020)[57]	https://jlsteenwyk.com/ClipKIT/advanced/index.html
IQ-TREE v2.2.2.7	Minh et al. (2020)[58]	http://www.iqtree.org/
iTOL v6	Letunic and Bork (2024)[59]	https://itol.embl.de/
Open-Source PyMOL v2.5.0	Schrodinger LLC (2015)[60]	https://github.com/schrodinger/pymol-open-source
Other		
NCBI RefSeq	O'Leary et al. (2016)[61]	https://www.ncbi.nlm.nih.gov/refseq/
NCBI GenBank	Benson et al. (2013)[62]	https://www.ncbi.nlm.nih.gov/genbank/
NCBI SRA	Leinonen, Sugawara and Shumway (2011)[63]	https://www.ncbi.nlm.nih.gov/sra
JGI Phycocosm	Grigoriev et al. (2021)[64]	https://phycocosm.jgi.doe.gov/phycocosm/home
1KP	Leebens-Mack et al. (2019)[65]	https://db.cngb.org/onekp/
ChaetoBase v1.1	Kumazawa et al. (2022)[15]	https://chaetoceros.nibb.ac.jp/
Thalassiosira pseudonana Lhcr18 (g8189.t1)	Filloramo et al. (2021)[66]	https://doi.org/10.5683/SP2/ZDZQFE
Lhca sequences of *Prasinoderma coloniale*	Li et al. (2020)[31]	https://ftp.cngb.org/pub/CNSA/data2/CNP0000924/CNS0223647/CNA0013964/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

All data used in this study were acquired from public databases cited in corresponding sections.

METHOD DETAILS

LHC protein sequence acquisition

LHC protein sequences were collected from genomes or transcriptomes of diverse red-lineage species, including 37 Rhodophyta comprising 33 species, five diatoms, two Eustigmatophyceae, four Haptophytes including one Pavlovophyceae, three Phaeophyceae (Brown algae) and three Cryptophytes. Thirty-three species of red algae include one Cyanidiales, two Cyanidioschyzonales, six Galdieriales, two Rhodellophyceae, two Compsopogonophyceae, two Stylonematophyceae, two Porphyridiophyceae, two Bangiophyceae, and 18 Florideophyceae. These genomic and transcriptomic datasets were accessed from databases such as ChaetoBase v1.1, NCBI RefSeq, NCBI GenBank, NCBI SRA, PDB, JGI Phycocosm and 1KP (https://db.cngb.org/onekp/). Specific details of species and corresponding references are provided in Table S1. For most diatoms, LHCs had been previously annotated, except for *Fistulifera solaris* JPCC DA0580.[15] The protein sequence of *Thalassiosira pseudonana* Lhcr18 was modified as the amino-acid sequence of g8189.t1 in the updated *T. pseudonana* genome (https://doi.org/10.5683/SP2/ZDZQFE).[66] A BLAST similarity search was used to procure LHCs from various lineages, adapting techniques from Kumazawa et al. (2022).[15] Then, the reference sequences used in the BLASTP search were replaced with the Lhcrs in *Porphyridium purpureum*. The transcriptomes of 1KP included 28 species of red algae, among which 23 species were selected for the next analyses based on the quality of the LHC alignment.[67] RedCAPs were identified by not only these genomes and transcriptomes but also by BLASTP against NCBI nr and tsa-nr database to identify them from wide variety of red-lineage algae. Contaminated sequences in the 1KP dataset were identified and eliminated using molecular phylogenetic analysis and BLASTP searches against the NCBI nr database. The IsoSeq transcriptomes of *Cyanidiococcus yangmingshanensis* 8.1.23 F7 and *Cyanidium caldarium* DBV 063 E5 were obtained from NCBI SRA and translated using TransDecoder (https://github.com/TransDecoder/TransDecoder) and LHC proteins were identified using BLASTP and clustered manually based on MAFFT alignment.[40,56] LHCs belonging to the Lhcr subfamily of secondary endosymbiotic algae were obtained by preliminary

phonetic analysis using muscle5 with super5 mode for alignment, ClipKit v1.4.1 with kpic-smart-gap mode for trimming, and IQ-TREE v2.2.2.7 to infer a phylogenetic tree.[55,57,58] All sequences for the following analyses were carefully curated for LHC conserved domain; some were modified at their N-terminal region and C-terminal region. All modified sequences are explicitly indicated in the Figures 2 and 4.

Lhca sequences of *Prasinoderma coloniale* belonging to the earliest divergent taxa of green algae were obtained for the root in the phylogenetic tree (https://ftp.cngb.org/pub/CNSA/data2/CNP0000924/CNS0223647/CNA0013964/).[31]

Phylogenetic analysis of LHC

LHC sequences (Data S1) were aligned using MAFFT E-INS-I v7.490,[56] which is optimized for multi-domain proteins (Data S2 and S3). Sequence alignments were then refined using ClipKit v1.4.1[65] with kpic-smart-gap mode for the following tree inference. The molecular phylogenetic trees were inferred using IQ-TREE v2.2.2.7 with extended model selection (-m MFP option) (Data S4 and S5).[58,68] For exhaustive tree topological search, 500 initial parsimony trees were constructed, the number of tree search iterations was extended to 1,000, and perturbation strength was specified to 0.7 as IQ-TREE parameters. The inferred trees were visualized using iTOL v6.[69]

Phylogenetic analysis of RedCAP

RedCAP sequences (Data S6) were aligned using MAFFT E-INS-I v7.525 (Data S7). The alignment was not trimmed and directly used to infer the phylogenetic tree by IQ-TREE v2.2.2.7 with the extended number of tree search iterations to 1,000. The Q.pfam+F+I+G4 model was selected for tree inference. The inferred trees (Data S8) were visualized using iTOL v6.

Visualization of PSI–LHCI structure

All models of PSI–LHCI structures were obtained from RCSB PDB (https://www.rcsb.org/). The following models were acquired: a *Cyanidioschyzon merolae* PSI–LHCI (ID: 5ZGB), a diatom *Chaetoceros gracilis* PSI–FCPI (ID: 6LY5), a red alga *Porphyridium purpureum* single-PBS–PSII–PSI–LHCs megacomplex (ID: 7Y5E), and a Cryptophyte *Chroomonas placoidea* PSI–ACPI (ID: 7Y7B). The models of the complexes were visualized using Open-Source PyMOL v2.5.0.[60]

QUANTIFICATION AND STATISTICAL ANALYSIS

For the molecular phylogenetic analysis, ultrafast bootstrap approximation (UFBoot) was employed to assess branch support, with 1000 replicates performed for each analysis. These values are displayed on the nodes of the phylogenetic trees presented in Figures 2, 4, and 5. Not only UFBoot but also the SH-like approximate likelihood ratio test[70] and the aBayes test[71] were displayed on some nodes of the trees and explained in each figure legend.