Unraveling the Evolutionary Trajectory of LHCI in Red-Lineage Algae: Conservation, Diversification, and Neolocalization

Minoru Kumazawa, Kentaro Ifuku*

Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan

*Corresponding Author:

Kentaro Ifuku: ifuku.kentaro.2m@kyoto-u.ac.jp

Abstract

Red algae and the secondary symbiotic algae that engulfed a red alga as an endosymbiont are called red-lineage algae. They comprise key marine taxa including diatoms, Haptophyta, and Cryptophyta. Several photosystem (PS) I–light-harvesting complex I (LHCI) structures have been reported from red-lineage algae —two red algae *Cyanidioschyzon merolae* (Cyanidiophyceae), *Porphyridium purpureum* (Rhodophytina), a diatom *Chaetoceros gracilis* and a Cryptophyte *Chroomonas placoidea*. Here, we clarified the orthologous relation of LHCIs in red-lineage algae by combining a detailed phylogenetic analysis of LHCIs 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 of red algae and endosymbiosis leading to Cryptophyta, some LHCIs were lost and replaced by existing or differentiated LHCIs. Especially in Cryptophyta, uniquely diversified Lhcrs form three sets of heterotrimers contributed to the expansion of the antenna size of PSI, supporting the modern ecological success of this taxon. We denominated "neolocalization" to these examples of flexible reorganization of LHCIs. This study provides new insights into the evolutionary process of LHCIs associated with PSI in the 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 (Delwiche, 1999). These photosynthetic organisms possessing 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 (Pierella Karlusich *et al.*, 2020).

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 (Croce and van Amerongen, 2020). 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 (Pi *et al.*, 2018; You *et al.*, 2023; Wolfe *et al.*, 1994; Marquardt and Rhiel, 1997). 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,

50 red-lineage secondary endosymbiotic algae, green lineage secondary endosymbiotic algae,
51 and dinoflagellates (Koziol *et al.*, 2007; Büchel, 2015).

52 Red algal LHCs contain chlorophyll *a* and zeaxanthin as a carotenoid, while most
53 LHCs of red-lineage secondary endosymbiotic algae include chlorophyll *a* and *c*; carotenoids
54 depend on the taxonomic group. In fact, LHCs are named after their binding carotenoids
55 (Büchel, 2015). For example, diatoms and Haptophytes contain fucoxanthin or 19'-
56 hexanoyloxy fucoxanthin as major carotenoid in their LHCs, thus their LHCs are called
57 fucoxanthin chlorophyll *a*/*c*-binding proteins (FCPs). Among red-lineage secondary
58 endosymbiotic algae, diatoms utilize FCPs as peripheral antennas for both PS I and II (Nagao
59 *et al.*, 2020; Nagao *et al.*, 2019; Nagao *et al.*, 2022; Xu *et al.*, 2020; Wang *et al.*, 2019).
60 Similar light-harvesting systems probably exist in other Stramenopiles and Haptophytes. At
61 least, Eustigmatophyceae, belonging to Stramenopiles, utilize LHCs for light-harvesting for
62 both PS (Umetani *et al.*, 2018).

63 Based on molecular phylogeny, the LHCs of red-lineage algae are divided into six
64 subfamilies: Lhcr, Lhcz, Lhcq, Lhcf, Lhcx, and CgLhcr9 homologs (Kumazawa *et al.*, 2022).
65 Some Stramenopiles, including Eustigmatophyceae and Phaeophyceae, as well as Chromera
66 from Alveolate have another LHC subfamily called red-shifted *Chromera* light-harvesting
67 proteins (Red-CLH) (Bína *et al.*, 2014; Umetani *et al.*, 2018). Red algae only have the Lhcr
68 subfamily, Stramenopiles and Haptophytes possess all six subfamilies of the red-lineage
69 LHCs, while Cryptophytes only have Lhcr and Lhcz subfamilies.

70 Recently, the advancement of cryoelectron microscopy structural analysis allowed
71 discerning the structures of the PS–peripheral light-harvesting antenna supercomplexes of

72 red-lineage algae. In red algae, the *Cyanidioschyzon merolae* PSI–LHCI supercomplex (Pi *et*
73 *al.*, 2018) and the *Porphyridium purpureum* phycobilisome–PSII–PSI–LHCI megacomplex
74 (You *et al.*, 2023) have been reported. In red-lineage secondary endosymbiotic algae,
75 molecular-level structural PS models have been reported in diatoms and a Cryptophyte. For
76 instance, the diatom *Chaetoceros gracilis* PSI–FCPI supercomplexes (Nagao *et al.*, 2020; Xu
77 *et al.*, 2020) and *C. gracilis* PSII–FCPII supercomplexes (Nagao *et al.*, 2019; Wang *et al.*,
78 2019; Nagao *et al.*, 2022) have been reported. In addition, PSII–FCPII structures from centric
79 diatoms *Thalassiosira pseudonana* and *Cyclotella meneghiniana* were recently reported (S.,
80 Zhao *et al.*, 2023; Feng *et al.*, 2023). The LHCs of Cryptophytes are called alloxanthin-
81 chlorophyll *a*/*c*-binding proteins (ACPs); the structure of the *Chroomonas placoidea* PSI–
82 ACPI has been recently reported (Zhao *et al.*, 2023).

83 With the identification of LHCs in the PSI–LHCI structural models recently reported,
84 it is now possible to evaluate the evolutionary process of the molecular assembly of LHCI
85 associated with PSI. The molecular assembly model of the red-lineage PSI–LHCI has been
86 discussed only based on spatial arrangements of the subunits present in the structures (Bai *et*
87 *al.*, 2021; L.,-S., Zhao *et al.*, 2023). However, an evolutionary model of the photosynthetic
88 supercomplex should comprise both molecular phylogeny and structural information. Such
89 an integrative understanding of the complex structures and molecular phylogenies of
90 primitive species has been attempted in the green lineage (Neilson and Durnford, 2010). Loss
91 and gain of LHC subfamilies during the evolutionary history of the red-lineage algae were
92 investigated through phylogenetic analysis of diatom LHCs (Kumazawa et al. 2022). A red-
93 lineage chlorophyll *a*/*b*-binding-like protein (RedCAP), a distinctive family of the LHC

94 superfamily, is conserved in PSI–LHCI of Rhodophytina red algae, Cryptophytes, and
95 diatoms (Engelken *et al.*, 2010; Sturm *et al.*, 2013; Xu *et al.*, 2020; You *et al.*, 2023; L.,-S.,
96 Zhao *et al.*, 2023). However, the evolutionary model of the red algal LHCI and RedCAP
97 remains incomplete because of the limited structural information, insufficient genome and
98 transcriptome information (until recently), and lack of detailed molecular phylogeny at the
99 ortholog level (Hoffman *et al.*, 2011; Dittami *et al.*, 2010).

100 In this study, we performed a molecular phylogenetic analysis to clarify the orthology
101 of LHCIs in red-lineage algae using recently reported genomes and transcriptomic data. The
102 detailed molecular phylogeny of the red-lineage LHCI, specifically Lhcrs, is combined with
103 new PSI–LHCI structural models from red and red-lineage algae. This has uncovered
104 conservation, diversification, and differentiation of the molecular assembly of LHCIs,
105 especially in red algae and Cryptophytes. Based on our analyses, we propose a new
106 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 (Engelken *et al.*, 2010; Sturm *et al.*, 2013; You *et al.*, 2023). 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 (Fig. 1) (Yoon *et al.*, 2002; Kim *et al.*, 2017). After secondary endosymbiosis, the earliest divergent event divided red algae into two major groups: Rhodophytina and Cyanidiophyceae (or Cyanidiophytina) (Yang *et al.*, 2016; Park *et al.*, 2023): The former includes classes such as Porphyridiophyceae, Stylonematophyceae, and Compsopogonophyceae with the subclades of Rhodellophyceae, Bangiophyceae, and Florideophyceae (Yang *et al.*, 2016; Borg *et al.*, 2023). The latter group, Cyanidiophyceae, contains the orders Galdieriales, Cavernulicolales, Cyanidiales, and Cyanidioschyzonales (Park *et al.*, 2023). 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

129 primary green lineage— (Li *et al.*, 2020).

130 Next, a molecular phylogenetic tree was constructed using the obtained Lhcr
131 sequences from red algae and Lhca sequences from the green alga *P. coloniale* (Fig. 2). There
132 are seven groups of Lhcrs (group I–VII) from red algae belonging to Rhodophytina, including
133 *Porphyridium purpureum*. Each group contains one *P. purpureum* Lhcr; Group VII, III, and
134 VI containing PpLhcr3, PpLhcr4, and PpLhcr6 show monophyly, and Group III and VI are
135 sister groups to group VII. However, the distribution of Cyanidiophyceae Lhcrs is different:
136 Galdieriales, belonging to Cyanidiophyceae, has Lhcrs classified into six including group I
137 and IV–VII in addition to the Galdieriales-specific clade. This unique clade is a sister clade
138 to group I. Galdieriales lacks group II and -III LHCs. Cyanidiales and Cyanidioschyzonales
139 have only three orthologs to those of Rhodophytina, and these Lhcrs –*Cyanidioschyzon*
140 *merolae* Lhcr1 (hereafter CmLhcr1), CmLhcr2, and CmLhcr3– which belong to group V, VI
141 and VII, respectively.

142

143 **PSI–LHCI in "Primitive Red Algae" Cyanidioschyzonales**

144 In a previous structural model of PSI–LHCI for *Cyanidioschyzon merolae* belonging to
145 Cyanidioschyzonales (Pi *et al.*, 2018), there were five LHCs attached to PSI, including two
146 copies of CmLhcr1 (r1 and r1*) and CmLhcr2 (r2 and r2*) in addition to one CmLhcr3 (r3)
147 (Fig. 3). The structure of the phycobilisome–PSII–PSI–LHCI megacomplex (Fig. 3) has been
148 reported in another red alga, *P. purpureum* belonging to Porphyridiophyceae (You *et al.*,
149 2023); it has eight LHCIs (PpLhcr1–7 and RedCAP) around PSI.

150 From the stromal side, in the *P. purpureum* PSI–LHCI part, positions of eight molecular

151 LHCI species (r1–r7, and RedCAP) are labeled counterclockwise as positions 0–7 (p0–p7)
152 (Fig. 3). Accordingly, the positions of LHCI in *C. merolae* is labeled as p0, p1, and p5–p7 for
153 two CmLhcr1 (r1 and r1*), two CmLhcr2 (r2 and r2*), and CmLhcr3 (r3) (Pi *et al.*, 2018).
154 *C. merolae* PSI–LHCI does not have an LHC at positions p2–p4. The loss of Psa28 (also
155 called PsaR) in the genome of *C. merolae* implies that Psa28 would be crucial for LHCI
156 binding at p2–p4 (You *et al.*, 2023).

157 The Lhcrs at p5–p7 of both species seem to be conserved from their common ancestor.
158 Our phylogenetic analysis suggests that CmLhcr1–3 at p5–p7 in *C. merolae* PSI–LHCI
159 belong to groups V, VI, and VII, respectively, and that PpLhcr5, PpLhcr4, and PpLhcr3 at
160 p5–p7 in *P. purpureum* PSI–LHCI belong to group V, VI and VII, which are orthologs to
161 CmLhcr1–3 (Fig. 2 and 3). In contrast, significant changes are observed between the two
162 species in Lhcrs at p0 and p1 (Fig. 3): *P. purpureum* PSI binds RedCAP at p0, while that of
163 *C. merolae* binds CmLhcr1 at p0 in addition to p5; *P. purpureum* binds PpLhcr2 (group I
164 Lhcr) at p1, while *C. merolae* binds CmLhcr2 (group VI Lhcr) at p1 as well as p6. Importantly,
165 PpLhcr2 and CmLhcr2 are not orthologs in the phylogenetic tree, although they bind at the
166 same position; instead, CmLhcr2 and PpLhcr4 at p6 belong to group VI and they show an
167 orthologous relationship in the tree. Given that RedCAP is conserved in Galdieriales which
168 is an early diverged taxa of Cyanidiophyceae, RedCAP is lost in *C. merolae* (Engelken *et al.*,
169 2010; Sturm *et al.*, 2013). *C. merolae* lost RedCAP, group I–IV LHCIs, and a related PSI
170 subunit, and complemented positions p0 and p1 with group V and VI LHCIs to, at least
171 partially, maintain the antenna size for PSI.

172

173 **Molecular Phylogeny of Lhcrs in Red-Lineage Algae**

174 Cryptophytes do not possess the Lhcf subfamily or phycobilisome, but they do
175 contain phycobiliproteins in the thylakoid lumen and LHCs in the thylakoid membrane
176 (Spear-Bernstein and Miller, 1989). Cryptophyte LHC is called chlorophyll *a*/*c* (CAC)
177 proteins or alloxanthin-chlorophyll *a*/*c* proteins (ACPs) named after their pigments and
178 bound to both PSI and PSII (Kereïche *et al.*, 2008; Kaňa *et al.*, 2012; L.,-S., Zhao *et al.*,
179 2023). The PSI–LHCI and PSII–LHCII of diatoms and Haptophytes are called PSI–FCPI and
180 PSII–FCPII, respectively, because of the bound pigments in LHCs. The classification of
181 LHCs in red-lineage secondary endosymbiotic algae was described in our previous research
182 (Kumazawa *et al.*, 2022). The diatom FCPI comprises Lhcrs as well as the Lhcqs, CgLhcr9,
183 which is distinct from the Lhcr subfamily, presumably some Lhcfs, and a RedCAP (Nagao
184 *et al.*, 2020; Xu *et al.*, 2020; Kumazawa *et al.*, 2022; Calvaruso *et al.*, 2020). Unlike red algal
185 PSII with phycobilisome, diatoms predominantly have the Lhcfs located around PSII, and
186 one unique Lhcr (CgLhcr17 homolog) closely associated with the PSII core (Nagao *et al.*,
187 2019; Nagao *et al.*, 2022; Wang *et al.*, 2019; Kumazawa *et al.*, 2022; Calvaruso *et al.*, 2020).
188 Molecular phylogenetic analysis suggests that Haptophytes also possess LHC subfamily
189 compositions similar to diatoms, implying that they may have analogous PSI–FCPI and
190 PSII–FCPII (Kumazawa *et al.*, 2022). To elucidate how the Lhcrs in these algae were
191 generated from red algae during secondary endosymbiosis, LHC sequences were obtained
192 from a wide variety of Cryptophytes, Stramenopiles, and Haptophytes, and we performed

193 molecular phylogenetic analysis on the obtained Lhcr sequences using *Prasinoderma* Lhcas
194 as root of the tree (Fig. 4).

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

205 Stramenopiles, including the diatom *Chaetoceros gracilis*, have CgLhcr1 in group I
206 and CgLhcr5 in group V in addition to LHCIs in two Stramenopiles/Haptophyta-specific
207 groups. The clade including CgLhcr1 is a sister clade of the group I clade of red algae,
208 suggesting that this clade can be included in group I. One Stramenopiles/Haptophyta-specific
209 group contains CgLhcr2, r3, r6, r8, and r10, while the other includes CgLhcr7 and r17.
210 CgLhcr17 is a monomeric FCPII directly associated with PSII (Kumazawa *et al.*, 2022;
211 Nagao *et al.*, 2022). The Lhcrs of other Stramenopiles showed the same distribution pattern
212 in the groups as that of diatom Lhcrs. Group I includes Pavlovales Lhcrs from Haptophytes,
213 whereas Stramenopiles/Haptophyta-specific groups possess Lhcrs from other taxa of

214 Haptophyta. Altogether, Lhcrs of red-lineage secondary endosymbiotic algae are present not
215 only in red algal Lhcr groups but also in unique clades in the phylogenetic tree.
216
217 **Cryptophyte PSI–ACPI and diatom PSI–FCPI showing LHCI rearrangement**
218 Conservation and replacement of LHCIs (ACPIs) from ancestral red algae can be
219 estimated when we combined the structure of Cryptophyte *Chroomonas placoidea* PSI–
220 ACPI with the phylogenetic analysis. *C. placoidea* PSI–ACPI contains one RedCAP
221 molecule and 13 Lhcrs (Fig. 5) (Zhao *et al.*, 2023). It also has an unknown subunit and ACPI-
222 S, which appears to interconnect ACPIs and is not found in red algal PSI–LHCI and diatom
223 PSI–FCPI. From the stromal side, the binding positions of LHCI (ACPIs) can be assigned as
224 p0–p8 similar to the red algal PSI–LHCI (Fig. 3). RedCAP locates at p0 as observed in *P.*
225 *purpureum*. ACPI-4, 3, and 2 at p4, p5, and p6 belong to groups IV, V, and VI, respectively,
226 as did LHCIs in *P. purpureum* PSI–LHCI. In contrast, ACPI-7 at p1 belongs to group V in *C.*
227 *placoidea,* while p1 is occupied by group I PpLhcr2 in *P. purpureum* PSI–LHCI. Furthermore,
228 p2, p3, and p7 are occupied by ACPI-6, -5, and -1 belonging to the Cryptophyte-specific
229 clade.
230 Interestingly, *Chroomonas placoidea* PSI–ACPI has three sets of adjacent three
231 ACPIs (heterotrimer): ACPI-7–6–5, ACPI-11–10/13–9, and 14–10/13–12 (Fig. 5). ACPI-7,
232 11, and 14 in each heterotrimer belong to group V. ACPI-3 belonging to group-V ACPIs
233 would be rather ancestral LHCI because it binds at p3, meaning a true ortholog of group-V
234 LHCI in red algae (ex. PpLhcr5). ACPI-7, 11, and 14 belonging to group V should be
235 derivatives of ancestral ACPI-3. Other two ACPIs in each heterotrimer belong to the unique

236 Cryptophyte-specific clade, wherein *C. placoidea* ACPIs can be divided into two groups by
237 focusing on monophyly: ACPI-6, 10/13, and ACPI-5, 9, 12 (Fig. 4). Thus, three sets of
238 heterotrimers comprise three ACPIs from the respective groups, which would result from
239 gene duplication.

240 According to the above analyses, the following model can be proposed for the
241 establishment of Cryptophyte PSI–ACPI (Fig. 6): After losing many Lhcr genes during
242 secondary endosymbiosis of red algae, Cryptophytes only preserved three molecular species
243 of Lhcr (ACPI-2, ACPI-3, and ACPI-4) and a RedCAP; at this point, the group V LHCI—an
244 ancestral ACPI-3 at p5—diversified forming a Cryptophyte-specific clade in the Lhcr
245 subfamily: one Cryptophyte-specific Lhcrs binds at p7 as ACPI-1; three sets of heterotrimers
246 of ACPIs, including one group V Lhcr and two Cryptophyte-specific Lhcrs, bind to restore
247 the antenna size of PSI. This molecular evolutionary model of *C. placoidea* PSI–ACPI (Fig.
248 6) contradicts the current model of PSI–LHCI complex evolution in the red-lineage, solely
249 based on LHCI compositions (Zhao et al., 2023). That is, Cryptophyte PSI–ACPI is not an
250 evolutional intermediate between red algal PSI–LHCI and diatom PSI–FCPI.

251 Not all Lhcrs in a diatom PSI–FCPI directly descend from red algal LHCI. The diatom
252 *Chaetoceros gracilis* PSI–FCPI possesses either none or one RedCAP molecule and 16 or 23
253 FCPs including eight Lhcrs (Nagao *et al.*, 2020; Xu *et al.*, 2020). In the structure of *C. gracilis*
254 PSI–FCPI, positions p0–p7 are occupied by CgRedCAP, and CgLhcr1–r7. CgRedCAP at p0
255 is homologous to PpRedCAP at p0 of *P. purpureum* PSI–LHCI. CgLhcr1 and CgLhcr5,
256 belonging to groups I and V, respectively, are positioned in p1 and p5, consistent with LHCIs
257 belonging to groups I and V at p1 and p5 in red algal PSI–LHCI. However, all other positions

258 except for p4 are occupied by Lhcrs belonging to Stramenopiles/Haptophyta-specific groups.
259 Furthermore, CgLhcr8 and CgLhcr10, belonging to the Stramenopiles/Haptophyta-specific
260 group I, are positioned counterclockwise adjacent to CgLhcr7 (positions p8 and p9,
261 respectively). The FCPI assigned at p4 in *C. gracilis* PSI–FCPI is CgLhcr4, belonging to the
262 Lhcq subfamily (Nagao *et al.*, 2020; Kumazawa *et al.*, 2022). Altogether, the diatom PSI–
263 FCPI shares orthologous LHCIs only at p0, p1, and p5 with the red algal PSI–LHCI. This
264 suggests that diatoms have lost many LHCIs from red algae and replaced them with
265 diversified FCPIs during endosymbiosis of red algae.
266
267
268 **Discussion**
269 The molecular assembly model of the red-lineage PSI–LHCI has been discussed only
270 based on spatial arrangements of the subunits in the structures (Bai *et al.*, 2021; Zhao *et al.*,
271 2023). However, this study made it clear that it is necessary to consider an evolutionary model
272 of the photosynthetic supercomplex using both molecular phylogeny and structural
273 information. Here, we would like to propose new evolutionary trajectory of LHCI proteins
274 associated with PSI in red-lineage algae.
275
276 **Putative composition of LHCIs in the common ancestor of primitive red algae**
277 The smaller number of LHCI in *C. merolae* PSI can be due to genome reduction in
278 Cyanidioschyzonales, including *C. merolae*, and Cyanidiales, resulting in small genome sizes
279 and numerous gene deletions (Cho *et al.*, 2023). Rhodophytina and Cyanidiophyceae
280 diverged from the common ancestor of red algae. Rhodophytina had a conserved set of Lhcrs:

281 groups I–VII. However, Cyanidiophyceae, Cyanidioschyzonales, and Cyanidiales, have only
282 three LHCIs which belong to groups V–VII. Galdieriales have six Lhcr molecular species
283 belonging to groups I, IV–VII, and the Galdieriales-specific clade. Furthermore,
284 Rhodophytina and Galdieriales possess RedCAP in their genomes, whereas
285 Cyanidioschyzonales and Cyanidiales do not (Engelken *et al.*, 2010; Sturm *et al.*, 2013). This
286 suggests that only a few Lhcr genes remained after genome reduction in Cyanidioschyzonales
287 and Cyanidiales, and that they would ensure PSI antenna size by repurposing the same
288 genetic products, CmLhcr1 and CmLhcr2, to different position around PSI.

289 Considering that Galdieriales is the earliest order diverged from others in
290 Cyanidiophyceae, it is reasonable to deduce the LHCI composition of the last common
291 ancestor of red algae from those of Rhodophytina and Galdieriales. Since both Rhodophytina
292 and Galdieriales possess RedCAP, the ancestral Rhodophyta likely had RedCAP at p0.
293 Galdieriales has Lhcrs of group I, IV–VII, which are orthologous Lhcrs in *P. purpureum* PSI–
294 LHCI bound at p1, p4–p7, while it lacks group II and III Lhcrs. Galdieriales conserves the
295 Psa28 (XP_005707993.1, called PsaR, synonymously) subunit, which was suggested to
296 stabilize LHCIs at p2–p4 in *P. purpureum* PSI–LHCI (You *et al.*, 2023). In Galdieriales,
297 Psa28 may help the association of at least LHCI at p4. Thus, the ancestral Rhodophyta should
298 have a conserved fundamental Lhcrs–RedCAP composition with RedCAP at p0 and Lhcrs at
299 least at p1, and p4–p7. This also suggests that *C. merolae* PSI–LHCI is not a "primitive" PSI–
300 LHCI.

301

302 **Evolutionary path of LHCI from red algae to Stramenopiles/Haptophyte through**

303 **Cryptophyta**

304 The evolutionary model of Cryptophyta PSI–ACPI introduced in Figure 6 suggest its
305 progression from the ancestral red algae PSI–LHCI. Cryptophyta acquired a plastid from red
306 algae, having its endosymbiotic nucleus, the nucleomorph, derived from red algae (Van Der
307 Auwera *et al.*, 1998; Archibald *et al.*, 2001). This supports the direct descendance of
308 Cryptophyte PSI–ACPI from red algae PSI–LHCI, without additional endosymbiosis. The
309 ancestral red algae PSI–LHCI, featuring LHCI at p0, 1, and p4–p7, shares positions p0, p4–
310 p5 with Cryptophyte PSI–ACPI. Despite losing group I Lhcr, Cryptophyta has diversified its
311 Lhcrs not only to fill position p7 but also to expand its antenna size by creating three
312 heterotrimers. This research meticulously details the evolution of LHCI from primitive red
313 algae to Cryptophytes, presenting Cryptophytes as a prime example of extensive LHCI
314 rearrangement and antenna enlargement of PSI through endosymbiosis.

315 In contrast, the evolutionary trajectory of Stramenopiles and Haptophyte LHCI is not
316 as straightforward as that of Cryptophyte ACPIs. When considering the evolutionary model
317 of PSI–LHCI based on molecular phylogeny, Stramenopiles, including diatoms and
318 Haptophytes, acquired the Lhcq subfamily in addition to the Lhcr subfamily (Kumazawa *et*
319 *al.*, 2022). In diatom FCPI, only group I and V Lhcrs (CgLhcr1 and CgLhcr5) at p1 and p5,
320 respectively, have conserved positions from red algae. Other Stramenopiles and at least
321 Pavlovales in Haptophyta share group-I and -V Lhcrs. In diatom PSI–FCPI, Lhcrs at p2, p3,
322 p6–p9 belong to the Stramenopiles/Haptophyta-specific groups(Nagao *et al.*, 2020;
323 Kumazawa *et al.*, 2022). CgLhcr4, which does not belong to the Lhcr subfamily but to the
324 Lhcq subfamily, is assigned to p4 in diatom PSI–FCPI. Moreover, the Lhcq subfamily in

325 PSI–FCPI not only binds directly to the PSI core but also forms the outer layer of LHCs in
326 PSI–FCPI, contributing to a larger antenna size of the diatom PSI (Nagao *et al.*, 2020; Xu *et*
327 *al.*, 2020; Kumazawa *et al.*, 2022). This indicates that Stramenopiles has experienced a
328 distinct process of LHCI re-acquisition around PSI, which would be largely different from
329 that of Cryptophytes.

330 The evolutionary history from red algae to red-lineage algae has been previously
331 described (Stiller *et al.*, 2014). Linear regression analysis on nuclear genomes suggested
332 serial endosymbiosis in red-lineage algae: Cryptophyta engulfed a red alga as secondary
333 (2nd) endosymbiosis; the photosynthetic Stramenopiles incorporated Cryptophyta as tertiary
334 (3rd) endosymbiosis; Haptophyta acquired many genes from Stramenopiles as quaternary
335 (4th) endosymbiosis. However, the plastid-encoded genes of Haptophyta and Cryptophyta
336 strongly support monophyly in the phylogenetic tree (Kim *et al.*, 2017). Haptophyta may
337 have acquired the ancestral Cryptophyta plastid before or after massive gene transfer from
338 Stramenopiles, reconciling the two seemingly contradictory phylogenetic trees of different
339 genomes. LHC genes are nuclear-encoded and should follow the history of nuclear-encoded
340 genes derived from the quaternary endosymbiosis (Stiller *et al.*, 2014). Consistently, the
341 composition of LHC subfamilies of Haptophyta is similar to that of Stramenopiles
342 (Kumazawa *et al.*, 2022). In contrast, most genes coding for the PS core are encoded in the
343 plastid genome originated from Cryptophyta. Based on these facts, we hypothesize that the
344 origin of the genes coding for the Haptophyte PS core complex and that for light-harvesting
345 antennas can be chimeric. Further genetic and structural analysis of the PSI–LHCI in
346 Haptophytes is required to confirm/dismiss this hypothesis.

347

348 **Molecular Evolution of PS supercomplexes through "neolocalization"**

349 Gene duplication and functional diversification, for instance in biosynthetic enzyme
350 families, are typically referred to as "neofunctionalization" (e.g. Hansen et al. 2021).
351 However, even after intensive diversification, the primary function of the LHC family and
352 RedCAP remains light-harvesting. Therefore, their diversification can be best considered as
353 that of structural relocalization rather than function. In our study on the red alga Cyanidiales
354 *Cyanidium caldarium* PSI–LHCI (Kato *et al.*, 2024), we proposed the term "neolocalization"
355 (Kato *et al.*, 2024). It was defined as a phenomenon where a structural defect caused by gene
356 loss is complemented or modified by the product of another existing gene. In this study, we
357 expand the definition of neolocalization to include modifications by the product of a gene
358 differentiated after duplication from an existing one. Phenomena matching neolocalization
359 are also observed in the green lineage. In green algae *Chlamydomonas reinhardtii*, green
360 algal Lhca7 and terrestrial plant Lhca2 do not share orthology, yet they bind at the same
361 position in green-lineage PSI–LHCIs (Neilson and Durnford, 2010; Suga *et al.*, 2019; Ben-
362 Shem *et al.*, 2003).

363 For the PS supercomplex, which has a long evolutionary timeframe, drastic events like
364 genome evolution and secondary endosymbiosis may primarily trigger neolocalization,
365 driving the molecular evolution of LHCs. To understand this process, a general model of
366 molecular evolution must consider both molecular phylogeny and structure. At present, the
367 complete evolutionary paths of LHC diversification and differentiation from Rhodophytina
368 red algae to Stramenopiles and Haptophyta through Cryptophyta remain to be elucidated. As

369 more genomic information and more structural models become available, the relationship
370 between these taxa will become clearer, allowing the construction of an entire evolutionary
371 model of PS supercomplexes.
372
373 **Methods**
374 **LHC Protein Sequence Acquisition**
375 LHC protein sequences were collected from genomes or transcriptomes of diverse
376 red-lineage species, including 37 Rhodophyta comprising 33 species, five diatoms, two
377 Eustigmatophyceae, four Haptophytes including one Pavlovophyceae, three Phaeophyceae
378 (Brown algae) and three Cryptophytes. Thirty-three species of red algae include one
379 Cyanidiales, two Cyanidioschyzonales, six Galdieriales, two Rhodellophyceae, two
380 Compsopogonophyceae, two Stylonematophyceae, two Porphyridiophyceae, two
381 Bangiophyceae, and 18 Florideophyceae. These genomic and transcriptomic datasets were
382 accessed from databases such as ChaetoBase v1.1, NCBI RefSeq, NCBI GenBank, NCBI
383 SRA, PDB, JGI Phycocosm and 1KP (https://db.cngb.org/onekp/,
384 https://ftp.cngb.org/pub/gigadb/pub/10.5524/100001_101000/100627/assemblies/). Specific
385 details of species and corresponding references are provided in Supplemental Table I. For
386 most diatoms, LHCs had been previously annotated, except for *Fistulifera solaris* JPCC
387 DA0580 (Kumazawa *et al.*, 2022). The protein sequence of *Thalassiosira pseudonana*
388 Lhcr18 was modified as the amino-acid sequence of g8189.t1 in the updated *T. pseudonana*
389 genome (https://doi.org/10.5683/SP2/ZDZQFE) (Filloramo *et al.*, 2021). A BLAST
390 similarity search was used to procure LHCs from various lineages, adapting techniques from

391 Kumazawa *et al.* (2022). Then, the reference sequences used in the BLASTP search were
392 replaced with the Lhcrs in *Porphyridium purpureum*. The transcriptomes of 1KP included 28
393 species of red algae, among which 23 species were selected for the next analyses based on
394 the quality of the LHC alignment (Leebens-Mack *et al.*, 2019). Contaminated sequences in
395 the 1KP dataset were identified and eliminated using molecular phylogenetic analysis and
396 BLASTP searches against the NCBI nr database. The IsoSeq transcriptomes of
397 *Cyanidiococcus yangmingshanensis* 8.1.23 F7 and *Cyanidium caldarium* DBV 063 E5 were
398 obtained from NCBI SRA and translated using TransDecoder
399 (https://github.com/TransDecoder/TransDecoder) and LHC proteins were identified using
400 BLASTP and clustered manually based on MAFFT alignment (Cho *et al.*, 2023; Katoh and
401 Standley, 2013). LHCs belonging to the Lhcr subfamily of secondary endosymbiotic algae
402 were obtained by preliminary phylogenetic analysis using muscle5 with super5 mode for
403 alignment, ClipKit v1.4.1 with kpic-smart-gap mode for trimming, and IQ-TREE v2.2.2.7 to
404 infer a phylogenetic tree (Edgar, 2022; Steenwyk, Buida, Li, X.,-X., Shen, *et al.*, 2020; Minh
405 *et al.*, 2020). All sequences for the following analyses were carefully curated for LHC
406 conserved domain; some were modified at their N-terminal region and C-terminal region. All
407 modified sequences are explicitly indicated in the figure 2 and 4.
408 Lhca sequences of *Prasinoderma coloniale* belonging to the earliest divergent taxa
409 of green algae were obtained for the root in the phylogenetic tree
410 (https://ftp.cngb.org/pub/CNSA/data2/CNP0000924/CNS0223647/CNA0013964/) (Li *et al.*,
411 2020).
412

413 **Phylogenetic Analysis of LHC**

414 LHC sequences were aligned using MAFFT E-INS-I v7.490 (Katoh and Standley,
415 2013), which is optimized for multi-domain proteins. Sequence alignments were then refined
416 using ClipKit v1.4.1 (Steenwyk, Buida, Li, X., X., Shen, *et al.*, 2020) with kpic-smart-gap
417 mode for the following tree inference. The molecular phylogenetic trees were inferred using
418 IQ-TREE v2.2.2.7 with extended model selection (-m MFP option) (Minh *et al.*, 2020;
419 Kalyaanamoorthy *et al.*, 2017). For exhaustive tree topological search, 500 initial parsimony
420 trees were constructed, the number of tree search iterations was extended to 1,000, and
421 perturbation strength was specified to 0.7 as IQ-TREE parameters. The inferred trees were
422 visualized using iTOL v6 (Letunic and Bork, 2021).

423

424 **Visualization of PSI–LHCI structure**

425 All models of PSI − LHCI structures were obtained from RCSB PDB
426 (https://www.rcsb.org/). The following models were acquired: a *Cyanidioschyzon merolae*
427 PSI–LHCI (ID: 5ZGB), a diatom *Chaetoceros gracilis* PSI–FCPI (ID: 6LY5), a red alga
428 *Porphyridium purpureum* single-PBS-PSII-PSI-LHCs megacomplex (ID: 7Y5E), and a
429 Cryptophyte *Chroomonas placoidea* PSI–ACPI (ID: 7Y7B). The models of the complexes
430 were visualized using Open-Source PyMOL v2.5.0 (Schrodinger LLC, 2015).

431

432 **Acknowledgments**

433 This work was supported in part by JSPS KAKENHI grant Nos. JP22KJ2017 (M.K.),
434 JP23H02347 (K.I.), and a grant from Institute for Fermentation, Osaka, Japan (K.I.). We

would like to thank Dr. Atsushi Takabayashi of Hokkaido Univ. and Dr. Ryo Nagao of Shizuoka Univ. for their valuable discussions throughout this study.

Author Contributions

M.K. and K.I. conceived the project; M.K. performed all analyses; M.K. drafted the original manuscript; M.K. and K.I. revised the manuscript and wrote the final manuscript, and both authors joined the discussion of the results.

Declaration of Competing Interest

The authors declare no conflict of interest.

Additional information

Supplemental Table 1: Accessions of genomic or transcriptomic data.

Supplemental Data 1: LHC protein sequences used in the analyses.

Supplemental Data 2: Alignment to infer the phylogenetic tree of Figure 3.

Supplemental Data 3: Alignment to infer the phylogenetic tree of Figure 5.

Supplemental Data 4: Newick file of the phylogenetic tree of Figure 3.

Supplemental Data 5: Newick file of the phylogenetic tree of Figure 5.

455 **References**

456 **Archibald, J.M., Cavalier-Smith, T., Maier, U. and Douglas, S.** (2001) Molecular
457 chaperones encoded by a reduced nucleus: The cryptomonad nucleomorph. *J Mol Evol*,
458 **52**, 490–501.

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

463 **Bai, T., Guo, L., Xu, M. and Tian, L.** (2021) Structural Diversity of Photosystem I and
464 Its Light-Harvesting System in Eukaryotic Algae and Plants. *Front Plant Sci*, **12**.
465 Available at: https://www.frontiersin.org/articles/10.3389/fpls.2021.781035/full.

466 **Ben-Shem, A., Frolow, F. and Nelson, N.** (2003) Crystal structure of plant photosystem
467 I. *Nature*, **426**, 630–635. Available at: https://www.nature.com/articles/nature02200.

468 **Bína, D., Gardian, Z., Herbstová, M., Kotabová, E., Koník, P., Litvín, R., Prášil, O.,**
469 **Tichý, J. and Vácha, F.** (2014) Novel type of red-shifted chlorophyll a antenna complex
470 from Chromera velia: II. Biochemistry and spectroscopy. *Biochim Biophys Acta*
471 *Bioenerg*, **1837**, 802–810.

472 **Borg, M., Krueger-Hadfield, S.A., Destombe, C., Collén, J., Lipinska, A. and Coelho,**
473 **S.M.** (2023) Red macroalgae in the genomic era. *New Phytologist*, **240**, 471–488.
474 Available at: https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19211.

475 **Büchel, C.** (2015) Evolution and function of light harvesting proteins. *J Plant Physiol*,
476 **172**, 62–75. Available at: http://dx.doi.org/10.1016/j.jplph.2014.04.018.

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. Available at: https://academic.oup.com/plphys/article/183/1/67-79/6116362.

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. Available at: https://www.nature.com/articles/s41467-022-35566-x.

Croce, R. and Amerongen, H. van (2020) Light harvesting in oxygenic photosynthesis: Structural biology meets spectroscopy. *Science (1979)*, **369**, eaay2058. Available at: https://www.sciencemag.org/lookup/doi/10.1126/science.aay2058.

Delwiche, C.F. (1999) Tracing the Thread of Plastid Diversity through the Tapestry of Life. *Am Nat*, **154**, S164–S177. Available at: https://www.journals.uchicago.edu/doi/10.1086/303291.

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. Available at: http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-10-365.

Edgar, R.C. (2022) Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat Commun*, **13**, 6968. Available at: https://www.nature.com/articles/s41467-022-34630-w.

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. Available at: https://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-10-233.

Feng, Y., Li, Z., Li, X., et al. (2023) *Structure of a diatom photosystem II supercomplex containing a member of Lhcx family and dimeric FCPII*, Available at: https://www.science.org.

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 Genomics*, **22**, 379. Available at: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-021-07666-3.

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. Available at: https://linkinghub.elsevier.com/retrieve/pii/S167420522100263X.

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. Available at: https://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-11-101.

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

Kaňa, R., Kotabová, E., Sobotka, R. and Prášil, O. (2012) Non-photochemical quenching in cryptophyte alga Rhodomonas salina is located in chlorophyll a/c antennae.

521 *PLoS One*, **7**.

522 **Kato, K., Hamaguchi, T., Kumazawa, M., et al.** (2024) The structure of PSI-LHCI from
523 *Cyanidium caldarium* provides evolutionary insights into conservation and diversity of
524 red-lineage LHCs. *Proceedings of the National Academy of Sciences*, **121**. Available at:
525 https://pnas.org/doi/10.1073/pnas.2319658121.

526 **Katoh, K. and Standley, D.M.** (2013) MAFFT Multiple Sequence Alignment Software
527 Version 7: Improvements in Performance and Usability. *Mol Biol Evol*, **30**, 772–780.
528 Available at: https://academic.oup.com/mbe/article-
529 lookup/doi/10.1093/molbev/mst010.

530 **Kereïche, S., Kouřil, R., Oostergetel, G.T., Fusetti, F., Boekema, E.J., Doust, A.B.,**
531 **Weij-de Wit, C.D. van der and Dekker, J.P.** (2008) Association of chlorophyll a/c2
532 complexes to photosystem I and photosystem II in the cryptophyte Rhodomonas CS24.
533 *Biochim Biophys Acta Bioenerg*, **1777**, 1122–1128.

534 **Kim, J.I., Moore, C.E., Archibald, J.M., Bhattacharya, D., Yi, G., Yoon, H.S. and**
535 **Shin, W.** (2017) Evolutionary Dynamics of Cryptophyte Plastid Genomes. *Genome Biol*
536 *Evol*, **9**, 1859–1872. Available at: https://academic.oup.com/gbe/article-
537 lookup/doi/10.1093/gbe/evx123.

538 **Koziol, A.G., Borza, T., Ishida, K.I., Keeling, P., Lee, R.W. and Durnford, D.G.**
539 (2007) Tracing the evolution of the light-harvesting antennae in chlorophyll a/b-
540 containing organisms. *Plant Physiol*, **143**, 1802–1816.

541 **Kumazawa, M., Nishide, H., Nagao, R., Inoue-Kashino, N., Shen, J., Nakano, T.,**
542 **Uchiyama, I., Kashino, Y. and Ifuku, K.** (2022) Molecular phylogeny of fucoxanthin-

543 chlorophyll *a* / *c* proteins from *Chaetoceros gracilis* and Lhcq/Lhcf diversity. *Physiol*
544 *Plant*, **174**, e13598. Available at: https://onlinelibrary.wiley.com/doi/10.1111/ppl.13598.
545 **Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., et al.** (2019) One thousand plant
546 transcriptomes and the phylogenomics of green plants. *Nature*, **574**, 679–685. Available
547 at: https://www.nature.com/articles/s41586-019-1693-2.
548 **Letunic, I. and Bork, P.** (2021) Interactive Tree Of Life (iTOL) v5: an online tool for
549 phylogenetic tree display and annotation. *Nucleic Acids Res*, **49**, W293–W296.
550 Available at: https://academic.oup.com/nar/article/49/W1/W293/6246398.
551 **Li, L., Wang, S., Wang, H., et al.** (2020) The genome of Prasinoderma coloniale unveils
552 the existence of a third phylum within green plants. *Nat Ecol Evol*, **4**, 1220–1231.
553 Available at: https://www.nature.com/articles/s41559-020-1221-7.
554 **Marquardt, J. and Rhiel, E.** (1997) The membrane-intrinsic light-harvesting complex
555 of the red alga Galdieria sulphuraria (formerly Cyanidium caldarium): biochemical and
556 immunochemical characterization. *Biochimica et Biophysica Acta (BBA) -*
557 *Bioenergetics*, **1320**, 153–164. Available at:
558 https://linkinghub.elsevier.com/retrieve/pii/S0005272897000200.
559 **Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D.,**
560 **Haeseler, A. von and Lanfear, R.** (2020) IQ-TREE 2: New Models and Efficient
561 Methods for Phylogenetic Inference in the Genomic Era E. Teeling, ed. *Mol Biol Evol*,
562 **37**, 1530–1534. Available at: https://academic.oup.com/mbe/article/37/5/1530/5721363.
563 **Nagao, R., Kato, K., Ifuku, K., et al.** (2020) Structural basis for assembly and function
564 of a diatom photosystem I-light-harvesting supercomplex. *Nat Commun*, **11**, 2481.

565 Available at: https://www.nature.com/articles/s41467-020-16324-3.

566 **Nagao, R., Kato, K., Kumazawa, M., et al.** (2022) Structural basis for different types
567 of hetero-tetrameric light-harvesting complexes in a diatom PSII-FCPII supercomplex.
568 *Nat Commun*, **13**, 1764. Available at: https://www.nature.com/articles/s41467-022-
569 29294-5.

570 **Nagao, R., Kato, K., Suzuki, T., et al.** (2019) Structural basis for energy harvesting and
571 dissipation in a diatom PSII–FCPII supercomplex. *Nat Plants*, **5**, 890–901. Available at:
572 https://www.nature.com/articles/s41477-019-0477-x.

573 **Neilson, J.A.D. and Durnford, D.G.** (2010) Structural and functional diversification of
574 the light-harvesting complexes in photosynthetic eukaryotes. *Photosynth Res*, **106**, 57–
575 71.

576 **Park, S.I., Cho, C.H., Ciniglia, C., et al.** (2023) Revised classification of the
577 Cyanidiophyceae based on plastid genome data with descriptions of the
578 Cavernulicolales ord. nov. and Galdieriales ord. nov. (Rhodophyta). *J Phycol*, **59**, 444–
579 466. Available at: https://onlinelibrary.wiley.com/doi/10.1111/jpy.13322.

580 **Pi, X., Tian, L., Dai, H.-E., Qin, X., Cheng, L., Kuang, T., Sui, S.-F. and Shen, J.-R.**
581 (2018) Unique organization of photosystem I–light-harvesting supercomplex revealed
582 by cryo-EM from a red alga. *Proceedings of the National Academy of Sciences*, **115**,
583 4423–4428. Available at: https://pnas.org/doi/full/10.1073/pnas.1722482115.

584 **Pierella Karlusich, J.J., Ibarbalz, F.M. and Bowler, C.** (2020) Phytoplankton in the
585 *Tara* Ocean. *Ann Rev Mar Sci*, **12**, 233–265. Available at:
586 https://www.annualreviews.org/doi/10.1146/annurev-marine-010419-010706.

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

588 **Spear-Bernstein, L. and Miller, K.R.** (1989) Unique location of the phycobiliprotein
589 light-harvesting pigment in the cryptophyceae. *J Phycol*, **25**, 412–419. Available at:
590 https://onlinelibrary.wiley.com/doi/10.1111/j.1529-8817.1989.tb00245.x.

591 **Steenwyk, J.L., Buida, T.J., Li, Y., Shen, X.-X. and Rokas, A.** (2020) ClipKIT: A
592 multiple sequence alignment trimming software for accurate phylogenomic inference A.
593 Hejnol, ed. *PLoS Biol*, **18**, e3001007. Available at:
594 https://dx.plos.org/10.1371/journal.pbio.3001007.

595 **Steenwyk, J.L., Buida, T.J., Li, Y., Shen, X.X. and Rokas, A.** (2020) ClipKIT: A
596 multiple sequence alignment-trimming algorithm for accurate phylogenomic inference.
597 *bioRxiv*, 1–17. Available at: http://dx.doi.org/10.1371/journal.pbio.3001007.

598 **Stiller, J.W., Schreiber, J., Yue, J., Guo, H., Ding, Q. and Huang, J.** (2014) The
599 evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat*
600 *Commun*, **5**, 1–7. Available at: http://dx.doi.org/10.1038/ncomms6764.

601 **Sturm, S., Engelken, J., Gruber, A., Vugrinec, S., G Kroth, P., Adamska, I. and**
602 **Lavaud, J.** (2013) A novel type of light-harvesting antenna protein of red algal origin
603 in algae with secondary plastids. *BMC Evol Biol*, **13**, 159. Available at:
604 https://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-13-159.

605 **Suga, M., Ozawa, S.-I., Yoshida-Motomura, K., Akita, F., Miyazaki, N. and**
606 **Takahashi, Y.** (2019) Structure of the green algal photosystem I supercomplex with a
607 decameric light-harvesting complex I. *Nat Plants*, **5**, 626–636. Available at:
608 https://www.nature.com/articles/s41477-019-0438-4.

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.

Wang, W., Yu, L.-J., Xu, C., et al. (2019) Structural basis for blue-green light harvesting and energy dissipation in diatoms. *Science (1979)*, **363**. Available at: https://www.science.org/doi/10.1126/science.aav0365.

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. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **1188**, 357–366. Available at: https://linkinghub.elsevier.com/retrieve/pii/0005272894900566.

Xu, C., Pi, X., Huang, Y., et al. (2020) Structural basis for energy transfer in a huge diatom PSI-FCPI supercomplex. *Nat Commun*, **11**, 5081. Available at: https://www.nature.com/articles/s41467-020-18867-x.

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. Available at: https://www.nature.com/articles/srep21361.

Yoon, H.S., Hackett, J.D., Pinto, G. and Bhattacharya, D. (2002) The single, ancient origin of chromist plastids. *Proceedings of the National Academy of Sciences*, **99**, 15507–15512. Available at: https://pnas.org/doi/full/10.1073/pnas.242379899.

You, X., Zhang, Xing, Cheng, J., Xiao, Y., Ma, J., Sun, S., Zhang, Xinzheng, Wang, H.-W. and Sui, S.-F. (2023) In situ structure of the red algal phycobilisome–PSII–PSI–

LHC megacomplex. *Nature*, **616**, 199–206. Available at: https://www.nature.com/articles/s41586-023-05831-0.

Zhao, L.-S., Wang, P., Li, K., et al. (2023) Structural basis and evolution of the photosystem I–light-harvesting supercomplex of cryptophyte algae. *Plant Cell*, **35**, 2449–2463. Available at: https://academic.oup.com/plcell/article/35/7/2449/7082805.

Zhao, S., Shen, L., Li, X., et al. (2023) Structural insights into photosystem II supercomplex and trimeric FCP antennae of a centric diatom Cyclotella meneghiniana. *Nat Commun*, **14**, 8164. Available at: https://www.nature.com/articles/s41467-023-44055-8.

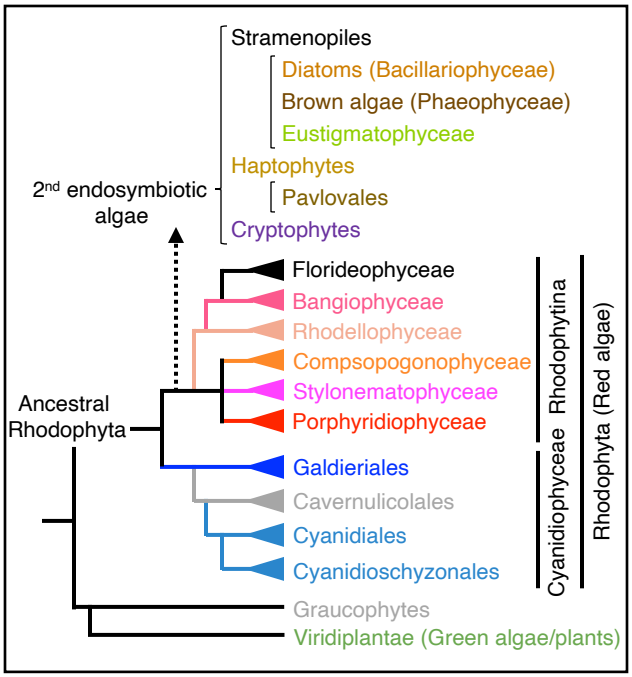

Figure 1 Simplified phylogeny of Rhodophyta (red algae) and classification of red-linage secondary endosymbiotic algae.

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.

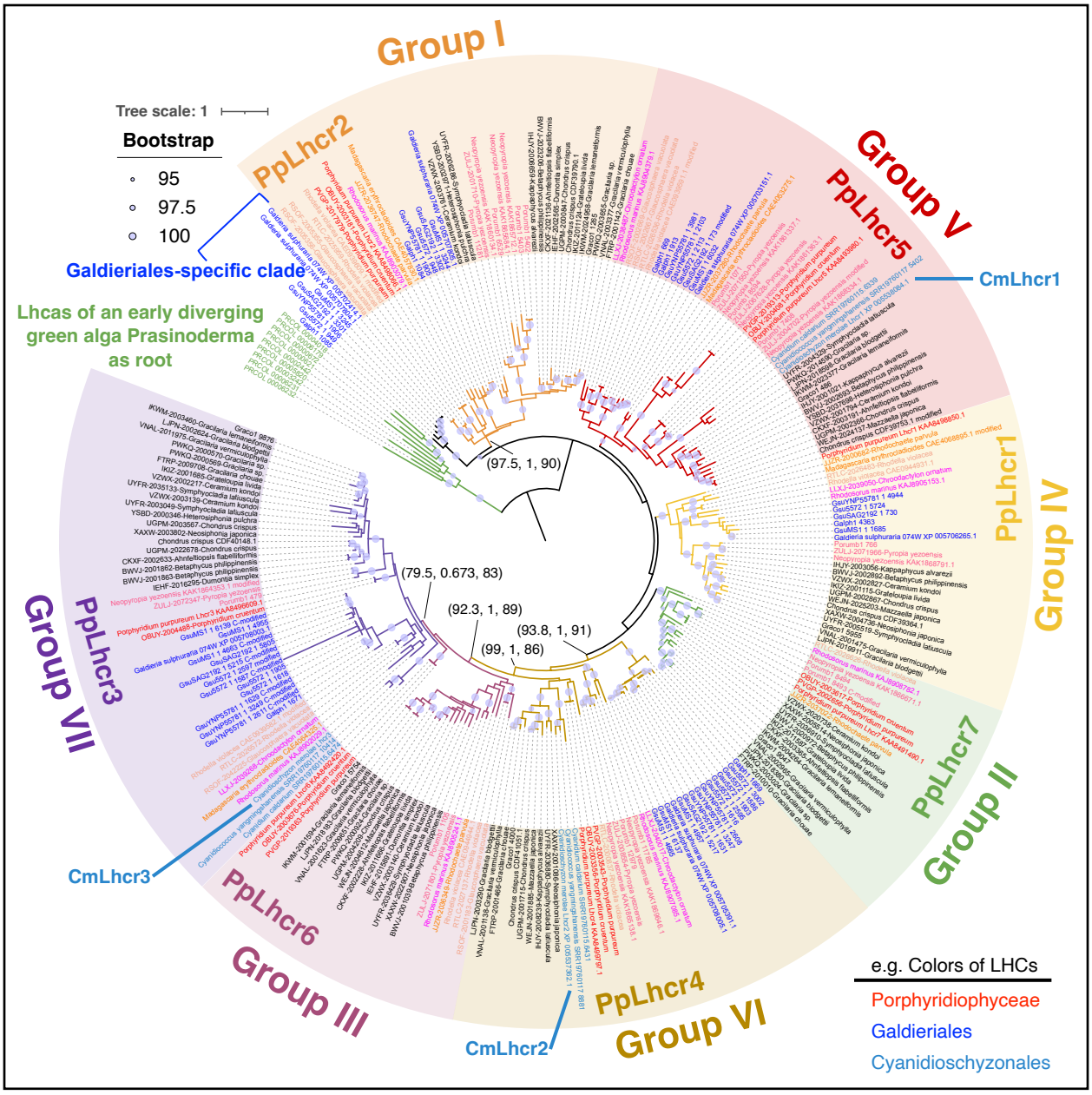

Figure 2 Molecular phylogenetic tree of Lhcr of red algae.

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 Supplemental Table 1. For example, *C. gracilis* Lhcr1 is shown as CgLhcr1.

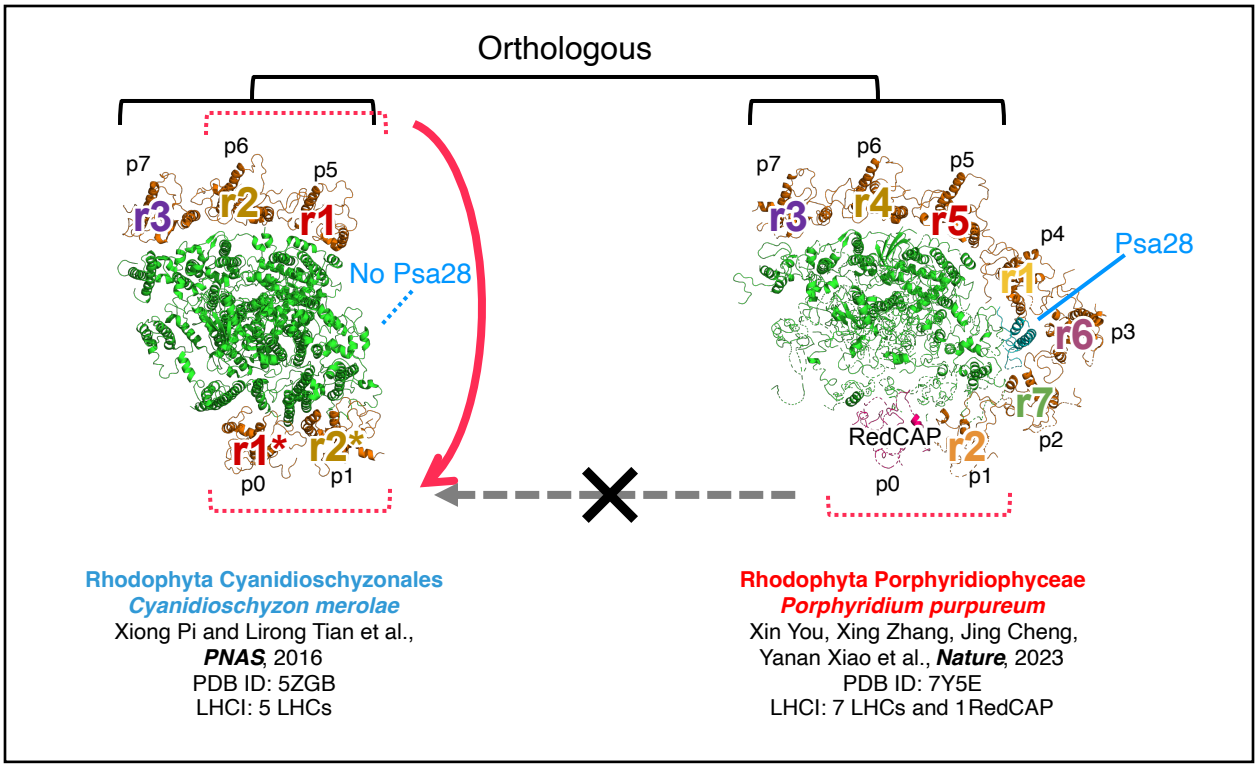

Figure 3 Estimated scheme of PSI-LHCI evolution in *Cyanidioschyzon.*

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.

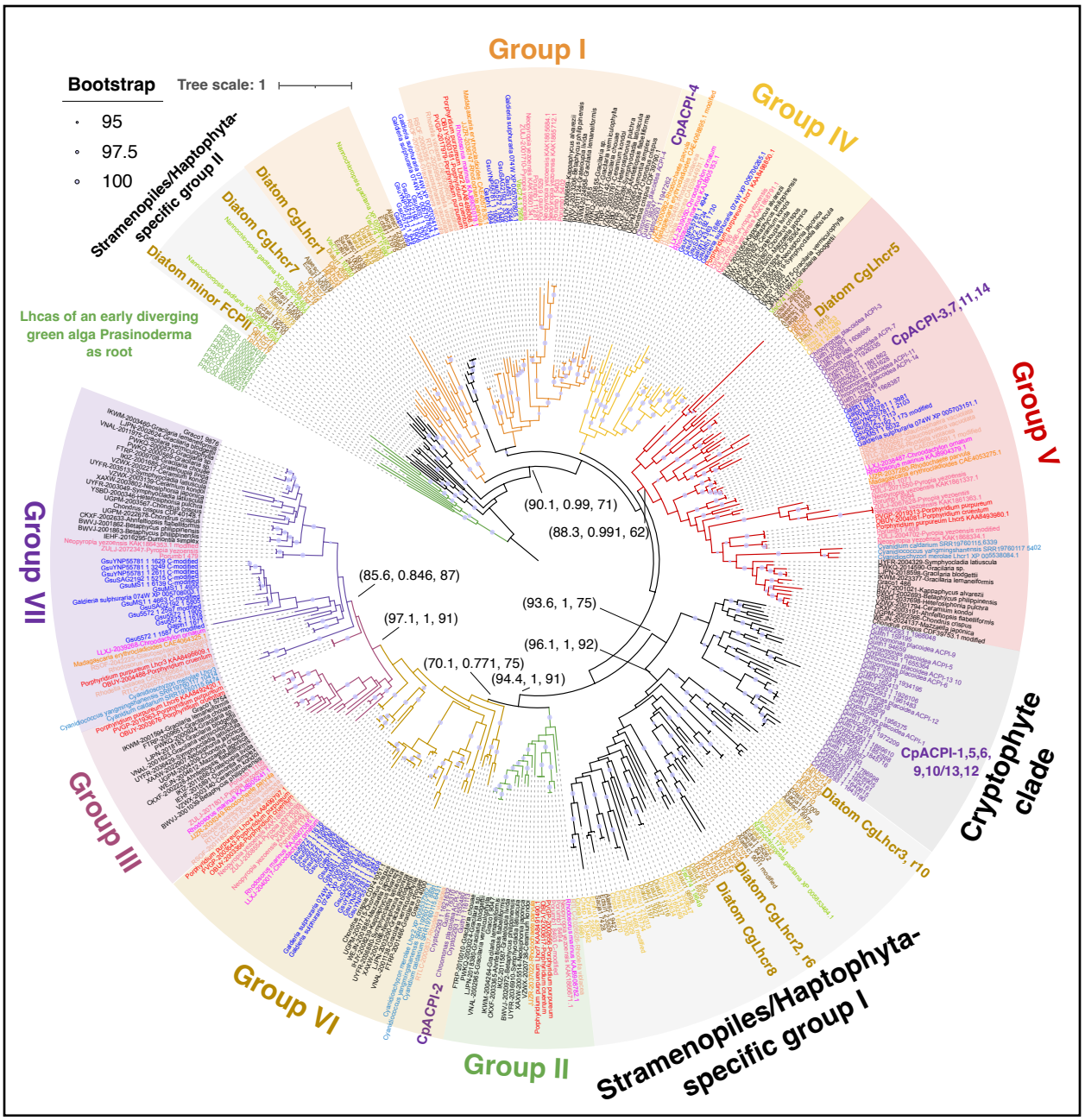

Figure 4 Molecular phylogenetic tree of Lhcr of red-lineage algae.

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 Supplemental Table 1.

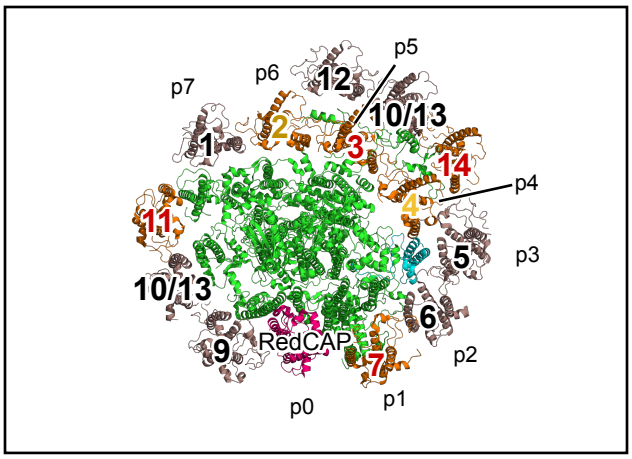

Figure 5 PSI–LHCI structures of Cryptophyte *Chroomonas placoidea* PSI–ACPI supercomplex.

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.

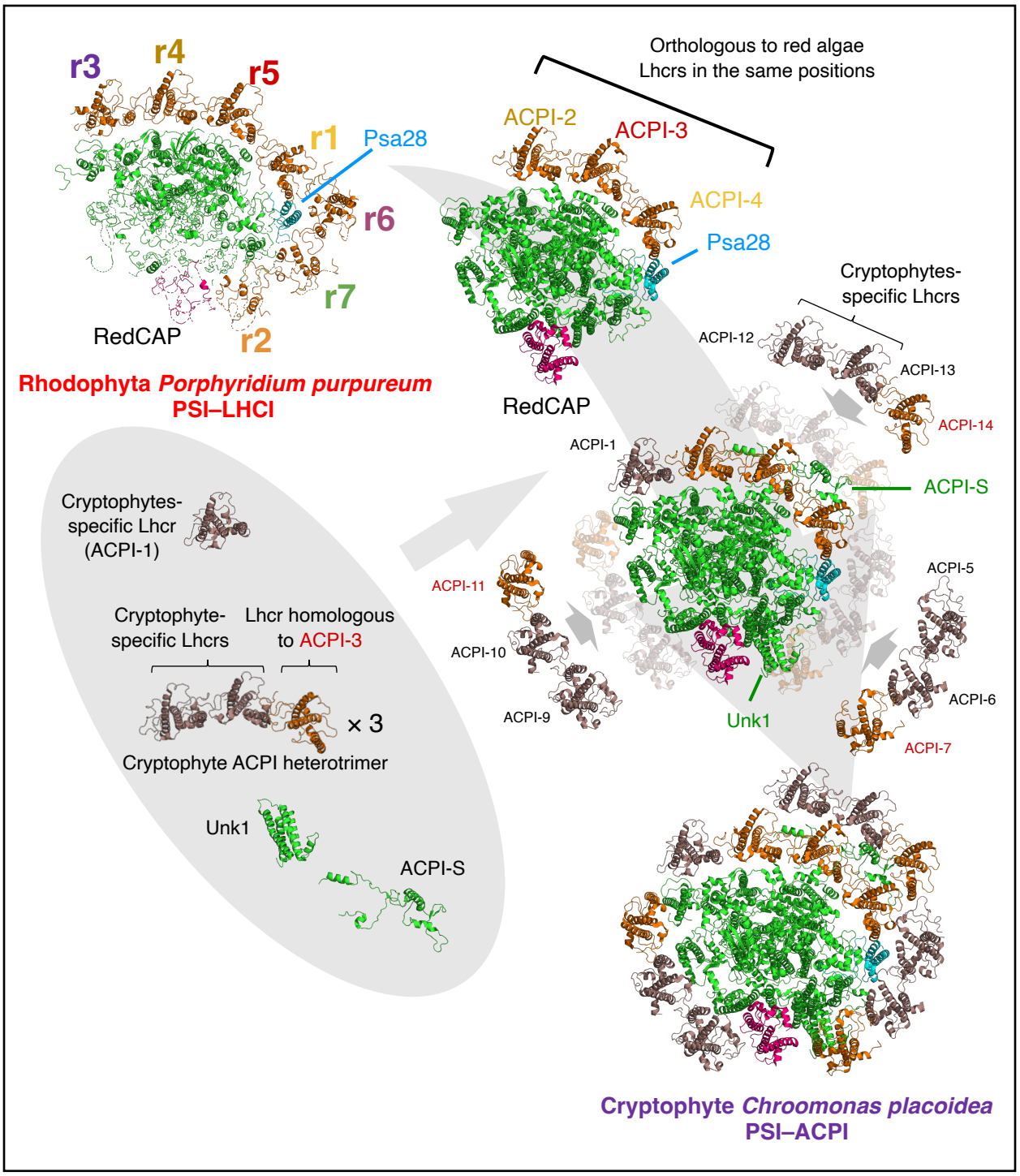

Figure 6 Estimated trajectory of PSI–LHCI evolution of Cryptophytes after endosymbiosis with red alga based on the molecular phylogeny of LHCI.

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 5. 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,