Duplications in the Chlorophyll *a* oxygenase (CAO) gene across the Viridiplantae

Mackenzie C. Poirier[1], Roberta Wright[1], Marina Cvetkovska[1]*

[1]Department of Biology, University of Ottawa, Ottawa, ON, Canada

***Corresponding author:**

Marina Cvetkovska
Department of Biology, University of Ottawa, 30 Marie-Curie Pr., Ottawa, ON, Canada, K1N 6N5
Email: mcvetkov@uottawa.ca
ORCID: 0000-0002-7080-8203

Statements and Declarations

Competing Interests: The authors report that there are no competing interests to declare

Acknowledgements: This project was supported by Natural Sciences and Engineering Research Council of Canada Discovery Grants (NSERC DG) awarded to M.C. The authors are grateful for the support from the Canada Foundation for Innovation (CFI) and University of Ottawa start-up funding. M.P. was supported by Ontario Graduate Scholarship (OGS), NSERC Graduate Scholarship, and Polar Knowledge Canada Antarctic Doctoral Scholarship.

Author's contributions: M. Poirier and M. Cvetkovska conceptualized the work and designed the experiments. Material preparation, data collection and analysis were performed by M. Poirier, R. Wright and M. Cvetkovska. The first draft of the manuscript was written by M. Poirier and all authors commented on all versions of the manuscript. All authors read and approved the final manuscript.

Abstract (150-250 words)

Viridiplantae, a diverse group of green plants and alga that have evolved from a common ancestor, are unified in their ability to produce and use two types of chlorophyll (chlorophyll *a* and chlorophyll *b*) to capture light energy. In addition to its role in light harvesting, appropriate levels of chlorophyll *b* are required for accumulation, assembly, and stability of light harvesting complexes within the photosynthetic apparatus. Chlorophyll *b* is synthesized from chlorophyll *a* by the enzyme chlorophyllide *a* oxygenase (CAO), a Rieske-type mononuclear non-heme iron oxygenase. A regulatory degron sequence, described in detail only in land plants, has been demonstrated to regulate the stability of CAO proteins based on the availability of chlorophyll *b*. Recent identification of CAO gene duplication in bryophyte and green algal species, combined with expanded availability of sequenced genomes within the Viridiplantae, prompted further investigation into the role of gene duplication in the evolution of chlorophyll *b* biosynthesis. Examination of genomes from 246 plant and algae species revealed independently occurring CAO duplications throughout the Viridiplantae, with a higher prevalence of duplication in land plants compared to their algal relatives. Additionally, we demonstrate that the degron sequence is absent in in chlorophytes, but first appears as a conserved sequence in charophytes, and is very highly conserved among the embryophytes. The evolutionary history of CAO throughout the Viridiplantae lineage is discussed based on these key observations, adding to our understanding of chlorophyll *b* biosynthesis and the role of CAO in photosynthetic species.

Keywords (4-6 keywords)

Chlorophyll biosynthesis, chlorophyll a oxygenase, gene duplication, degron, plants and algae

Introduction

Photosynthetic organisms form the base of all food chains and provide the organic carbon compounds that support life in our biosphere. The capture of light energy by chlorophyll is the first step in the process of photosynthesis that fixes of atmospheric CO_2 into stable organic compounds. Overwhelming evidence indicates that oxygenic eukaryotic photosynthesis that occurs in the chloroplasts of modern-day plants and algae traces its origins from an endosymbiotic event with cyanobacteria-like prokaryotes (Blankenship 2010; Cardona 2019; Sánchez-Baracaldo and Cardona 2020). Viridiplantae (or green plants), are primary endosymbionts that encompass two major clades: the chlorophytes and the streptophytes. Chlorophytes are a monophyletic group of marine, freshwater, and terrestrial green algae. With ~8,000 described species (Guiry 2024), this group encompasses a large diversity of adaptations, morphologies, and life histories (Leliaert et al. 2012). Streptophyta includes the charophytes, ~5,500 species of largely freshwater green algae (Guiry 2024) and the embryophytes, or land plants. Land plants are thought to have diverged from a single clade within the streptophytes (de Vries and Archibald 2018) and number ~430,000 species (Ruggiero et al. 2015). One characteristic that unifies members of the Viridiplantae is their ability to synthesize and use two types of chlorophylls: chlorophyll a (Chl *a*) and chlorophyll b (Chl *b*).

The photosynthetic apparatus in the Viridiplatae is derived from the ancestral cyanobacterial endosymbiont, which used chlorophyll as the main pigment in their photosystems (Xiong and Bauer 2002; Sánchez-Baracaldo and Cardona 2020). Chl *a* is ubiquitously present in all oxygenic photosynthetic eukaryotes, and functions in both energy capture in the antenna light harvesting complexes (LHCs) and in driving electron transfer in the photosystem II (PSII) and photosystem I (PSI) reaction centers. Chl *b* is specific to the Viridiplantae although it has also been detected in prochlorophytes and *Acaryochloris*, unique photosynthetic prokaryotic groups that lack typical cyanobacterial phycobilin light-harvesting pigments (Palenik and Haselkorn 1992; Roche et al. 1996; Tomitani et al. 1999; Partensky et al. 2018). Chl *b* resides predominantly in the peripheral LHCs (Neilson and Durnford 2010) although it has been detected in the core complexes of some deep water marine chlorophytes (Kunugi et al. 2016).

Strict Chl *a/b* stoichiometry is required for optimal energy transfer during photosynthesis. Maintaining the correct Chl *a/b* ratios is a dynamic process and indicative of adaptation to different light conditions (Tanaka and Tanaka 2007). In addition to its role in light harvesting, appropriate levels of Chl *b* are required for accumulation, assembly, and stability of LHCs in algae (Bujaldon et al. 2017), bryophytes (Zhang et al. 2023), and angiosperms (Król et al. 1995; Reinbothe et al. 2006; Kim et al. 2009; Nick et al. 2013). Under low light conditions, the amount of Chl *b* relative to Chl *a* increases, leading to larger LHC antenna size to maximizes the surface area for light absorption (Kunugi et al. 2016; Kume et al. 2018; Ueno et al. 2019). Thus, the amount of Chl *b* is a major regulator of light-harvesting capacity.

Chlorophyll biosynthesis and turnover occurs through a series of complex multistep pathways (reviewed in Brzezowski et al. 2015; Willows 2020; Schumacher et al. 2022). Chl *b* is synthesized from Chl *a*, via the intermediate 7-hydromethyl chlorophyll *a*, by the action of a single enzyme: chlorophyllide *a* oxygenase (CAO) (Tanaka et al. 1998; Espineda et al. 1999; Oster et al. 2000; Mueller et al. 2012). Chl *b* can be converted back into Chl *a* through the action of chlorophyll *b* reductase (CBR) and 7-hydroxymethyl reductase (HCAR) to complete the cycle (Figure 1; reviewed in detail in Tanaka and Tanaka, 2019). This chlorophyll cycle plays an essential role in many

biological processes, including biogenesis of LHCs, antenna size regulation during light acclimation, and chlorophyll degradation during senescence. While many of the intricacies of how this pathway is regulated are still not fully understood (Tanaka and Tanaka 2019), it has been suggested that CAO levels and activity are key determinants of Chl *b* accumulation.

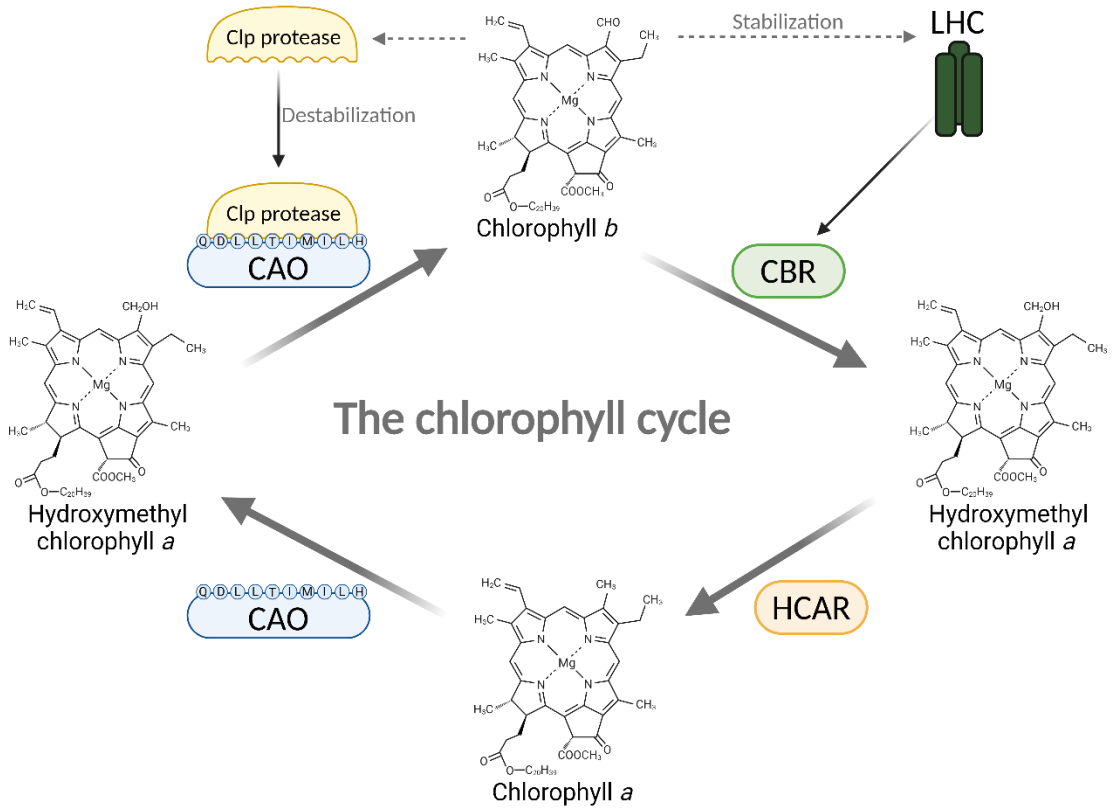

Figure 1. The chlorophyll cycle in land plants. Chlorophyll *a* is converted to chlorophyll *b* via the intermediate 7-hydromethyl chlorophyll *a*, by the action of the enzyme chlorophyllide *a* oxygenase (CAO). In a negative feedback loop, high levels of chlorophyll *b* trigger the binding of a stromal Clp protease to the degron motif (QDLLTIMILH) and the subsequent destabilization of CAO. Chlorophyll b is necessary for the stabilization of Light Harvesting Complexes (LHC) in plants. Excessive amounts of LHC or accumulation of energetically uncoupled non-functional LHCs (e.g., during senescence), triggers the accumulation of Chlorophyll b reductase (CBR) and the conversion of chlorophyll *b* to chlorophyll *a* via the action of CBR and 7-hydroxymethyl reductase (HCAR). This regulatory mechanism of CAO and CBR enables the fine- tuning and optimization of chlorophyll *b* and LHC levels.

CAO catalyzes the conversion of the methyl group of Chl *a* into a formyl group to form Chl *b* through two oxygenation reactions, likely using ferredoxin as a reductant (Oster et al. 2000). CAO is part of the Rieske-type mononuclear non-heme iron oxygenase group of proteins (Gray et al. 2004). CAO has three main regions: an N-terminal regulatory domain and a C-terminal catalytic domain, which are connected by a more variable linker region (Nagata et al. 2004). The catalytic domain contains a Rieske cluster and a mononuclear iron-binding domain, both of which are necessary for the enzymatic activity of CAO (Yamasato et al. 2005; Kunugi et al. 2013). The regulatory domain has

been characterized in detail only in land plants. In *Arabidopsis*, a conserved sequence (QDLLTIMILH) termed a degron was demonstrated to regulate Chl *b* accumulation by affecting CAO protein turnover. In a negative feedback loop, high levels of Chl *b* trigger the binding of a stromal Clp protease to the degron motif leading to destabilization of the CAO protein; this, in turn, halts Chl *b* biosynthesis (Sakuraba et al., 2009; Nakagawara et al., 2007).

The evolutionary origins of CAO can be traced back to prochlorophytes, which encode a CAO gene homolog involved in Chl biosynthesis (Tomitani et al. 1999; Nagata et al. 2004). Notably, the prokaryotic CAO-like enzymes lack the regulatory N-terminal domain, suggesting that the Chl *b*-dependant mechanism for controlling CAO stability was acquired in eukaryotes. Recent work on several chlorophyte, charophyte, and embryophyte representative species suggested that the degron recognition sequence first appeared in charophytes as an adaptation to high-light environments in shallow waters (Kunugi et al. 2016). These conclusions, however, were based on the presence of a degron in the CAO sequence of a single charophyte species (*Klebsormidinium flaccidum*) and require further validation.

It has been postulated that gene duplication is the driver of plant evolution (Rensing 2014), but previous reports suggested that the CAO protein is encoded by a single gene in plants and algae (Kunugi et al. 2016; Schumacher et al. 2022). Two CAO genes were detected in rice (Lee et al. 2005) although only one of these homologs is necessary for Chl *b* synthesis (Jung et al. 2021). Recent work revealed CAO duplication in the bryophyte *Physcomitrium patens,* where both homologs contribute to Chl *b* accumulation (Zhang et al. 2023). CAO duplication was also reported in two Antarctic *Chlamydomonas* species , were the CAO duplicates appeared to be paralogous suggesting an independent, rather than ancestral, duplication event (Cvetkovska et al. 2019; Poirier et al. 2025).

These insights, combined with advances in plant and algal genome sequencing, prompted us to further examine the *CAO* gene content and the presence of the regulatory domain across the diverse plant and algal groups. In this work we identified and examined CAO sequence in diverse species across Viridiplantae and demonstrate a widespread occurrence of CAO duplicates in this group. Furthermore, we demonstrate that the degron sequence is absent in chlorophytes, but present in charophyte and embryophytes, suggesting a different mechanism for Chl *b* regulation in early aquatic lineages.

Materials and Methods

CAO sequence identification and analyses

To identify *CAO* sequences and duplication events with high accuracy, we used publicly available, high quality, and fully annotated genomes. Chlorophyte and streptophyte genomes were retrieved from Phycosom (Grigoriev et al. 2021), while embryophyte genomes were obtained from Phytozome v13 (Goodstein et al. 2012). Putative *CAO* genes were identified with a tBLASTn search with the *Chlamydomonas reinhardtii* (CrCAO; Cre01.g043350) and the *Arabidopsis thaliana* (AtCAO; AT1G44446) full-length CAO peptides (Merchant et al. 2007; Lamesch et al. 2012). A strict e-value cutoff of ≤e-30 was taken as a threshold, to avoid many unspecific hits due to the conserved nature of the Fe-binding and Rieske cluster domains present among diverse plant and algal proteins (Schmidt and Shaw 2001; Ferraro et al. 2005; Przybyla-Toscano et al. 2021). Putative

172 *CAO* sequences were manually curated, and truncated or low-quality sequences were excluded
173 from downstream analyses. Curated *CAO* sequences were aligned using MUSCLE (Edgar 2004).
174 The degron sequence (QDLLTUMILH) (Sakuraba et al. 2009) was identified by multiple sequence
175 alignment of the full-length CAO amino acid sequences and the regulatory N-terminal domain.
176 Sequence logos for each major group were generated by Geneious Prime 2024.11 (Dotmatics),
177 where the size of the letter reflects its frequence.

178 ***Phylogenetic Inference***

179 The analyses described above resulted in 374 individual CAO sequences (Supplementary Table 1).
180 To avoid inaccurate inference and over-representation, we manually removed heterodimeric CAO
181 proteins, where the Rieske domain and the Fe-binding domain are located on different genes
182 (Kunugi et al. 2013) (Supplementary Table S1). We also removed identical CAO sequences isolated
183 from algal cultures identified as the same species but belonging to different strains or culture
184 collections (*Scenedesmus obliquus*, *Auxenochlorella protothecoides*, *Chlorella sorokiniana*,
185 *Ostreococcus tauri*, *Mesostigma viride*, *Zygnema circumcarinatum*). This resulted in a 280 amino
186 acid alignment of 326 unique CAO sequences across 246 species.

187 These manually curated CAO sequences were aligned with MUSCLE and trimmed to remove gaps
188 and ambiguously aligned regions with Gblock (Castresana 2000; Talavera and Castresana 2007).
189 Maximum likelihood trees were inferred in the CIPRES Science Gateway (Miller et al. 2015) using
190 RAxML v8.0 (Stamatakis 2014) with 1000 bootstraps with the Whelan Goldman matrix for globular
191 proteins (WAG), a gamma shape parameter, and empirical estimation of invariable sites. Trees
192 were visualized in iTOL v6.8.2 (Letunic and Bork 2021). Previously inferred taxonomic information
193 for chlorophytes and streptophytes was obtained from AlgaeBase, University of Galway
194 (www.algaebase.org), and for land plants from the Plants Of the World Online, Royal Botanic
195 Gardens, Kew (https://powo.science.kew.org).

196 **Results and Discussion**

197 ***Occurrences of multiple CAO genes are widespread among the Viridiplantae***

198 Our analysis demonstrates a widespread occurrence of *CAO* gene duplication across the
199 Viridiplantae. Multiple gene copies are more prevalent among the embryophytes, where ~41% of
200 examined species encode for more than one *CAO* gene. In comparison, we detected two CAO gene
201 copies in only ~11% of chlorophyte and ~20% of charophyte species (Table 1; Table 2,
202 Supplementary Table 1; Supplementary Table 2). All 23 species where we detect more than two
203 *CAO* genes are members of the Embryophyta. The modern-day cultivar of sugarcane (*Saccharum*
204 sp. R570) encodes for 6 *CAO* copies, the largest number we report. This species is a result of inter-
205 specific hybridization (*Saccharum officinarum* x *spontaneum;* Dumont et al. 2022), which has
206 resulted in a very large and highly redundant genome (~0.3 Gb, ~48,000 genes) (Healey et al. 2024).

207 We also detected multiple occurrences of heterodimeric *CAO* gene, where the functional CAO
208 protein is encoded by two genes (Table 1, Supplementary Table 1). Previous work in the
209 prasinophyte *Micromonas pusila* demonstrated that in this species CAO was encoded by two non-
210 homologous genes, one that possesses the mononuclear iron-binding motif (*MpCAO1*) and a
211 second one that encodes for the Rieske cluster (*MpCAO2*). It was shown that a combination of
212 both gene products was necessary to form an active heterodimeric CAO protein (Kunugi et al.

2013; Dey et al. 2023). We expand on this work by showing the presence of a heterodimeric CAO encoded by two separate genes across the Mammiellophycea, a group within the prasinophytes that contains *M. pusila* and several of its relatives (Worden et al. 2009) , the ecologically important picoalgae *Ostreococcus tauri*, *Ostreococcus lucimarinus* and *Bathycoccus prasinos* (Palenik et al. 2007; Moreau et al. 2012; Blanc-Mathieu et al. 2014), as well as the related *Chloropicon primus* (Chloropicophyceae) (Lemieux et al. 2019).

We also show that the heterodimeric CAO is not restricted only within the Prasinophytes. We detect a similar *CAO* gene arrangement in the genome of the multinucleate green alga *Caulerpa lentillifera* (Bryopsidales) (Arimoto et al. 2019) , as well as in in the genome of the secondary endosymbiont *Bigelowiella natans* (Chlorarachniophyceae) (Curtis et al. 2012). Phylogenetic analysis based on plastid-encoded protein sequences revealed that chlorarachiniophytes acquired their secondary plastids by the uptake of a filamentous green alga in the Bryopsidales (Suzuki et al. 2016), suggesting that a heterodimeric CAO has been inherited from the green algal endosymbiont. Considering that CAO in prochlorophytes is encoded by a single gene that contains both the Rieske and iron-binding domain (Nagata et al. 2004), it is likely that the occurrence of two separate genes is not an ancestral condition but rather occurred after the primary endosymbiotic event. Indeed, it has been suggested that acquisition of two separate genes encoding for key functional domains may not be an evolutionarily difficult process. CAO likely functions as a trimer (Kunugi et al. 2013; Liu et al. 2022), and the electron transfer from the Rieske to the mononuclear iron occurs between two neighbouring CAO proteins. A functionally similar process likely occurs in the CAO enzyme encoded by two separate genes.

Table 1: Number of species across the Viridiplantae with fully sequenced and annotated genomes available through Phytozome (Total), compared to the number of species that encode for multiple CAO genes (Multiple CAO) and number of species that encode for two *CAO* genes that form heterodimeric complex (Heterodimeric CAO). The number in brackets represents the percentage of species within the group that encode multiple CAO genes or heterodimeric *CAO* genes

Group	Total	Heterodimeric CAO	Multiple CAO
Chlorophytes*	70	11 (15.7%)	8 (11.5%)
Charophytes	10	0	2 (20%)
Non-vascular plants	6	0	3 (50%)
Lycophytes	3	0	2 (66%)
Gymnosperms	1	0	0
Angiosperms	151	0	62 (41.1%)

*The Chlorophyte group also includes the green algal secondary endosymbiont within the Chlorarachniophyceae (*Bigelowiella natans*)

All examined streptophytes encode for at least one *CAO* gene (Supplementary Table 1) signifying the importance of Chl *b* for plant physiology, but we were not able to identify a homolog of *CAO* in several chlorophyte algal species. Not surprisingly, we did not detect a *CAO* homolog in the invertebrate parasite *Helicosporidium* sp., a colorless protist that belongs within the trebouxiophycean green algal clade (Tartar et al. 2003). As a consequence of its parasitic lifestyle, *Helicosporidium* has lost nearly all genes associated with light harvesting, photosynthesis and chlorophyll biogenesis (Pombert et al. 2014), including CAO.

250 We were also not able to conclusively identify a *CAO* gene in the genomes of the endolithic coral
251 holobiont *Ostreobium quekettii* (Bryopsidales) (Iha et al. 2021), the aquaculture species
252 *Tetraselmis striata* (Chlorodendrophyceae) (Steadman Tyler et al. 2019) and the halotolerant
253 *Picocystis* sp. ML (Picocystophyceae) (Junkins et al. 2019) . All identified sequences within the
254 genomes of these species had e-value higher than the cut-off (≥e-15) used in this work and high
255 similarity to pheophorbide *a* oxygenase (PAO), a Rieske-type protein involved in chlorophyll
256 degradation (Pružinská et al. 2003). The lack of *CAO* genes in these algae is surprising and could be
257 an artifact of genome assembly, as Chl *b* has been experimentally detected in all three species
258 (Roesler et al. 2002; Massé et al. 2020; Conlon et al. 2024).

259 It is tempting to speculate that these species may have a different mechanism for Chl *b*
260 biosynthesis or a CAO sequence significantly different than most algae and plants. For instance,
261 *Ostreobium* is an endolithic coral holobiont adapted to extreme shading, and has lost many genes
262 associated with photoprotection and light perception due to evolution in a light-limited
263 environment (Iha et al. 2021). This alga is also the only known eukaryote that has lost the gene
264 encoding light-dependent protochlorophyllide oxidoreductase (LPOR), a key enzyme that catalyzes
265 the conversion of protochlorophyllide to chlorophyllide *a*, a necessary step towards the synthesis
266 of Chl *a* (Reinbothe et al. 2010). Instead, this alga fully depends on dark operative POR (DPOR) to
267 synthesize Chl *a*, likely as an adaptation to extreme shading (Iha et al. 2021). The mechanism of
268 chlorophyll biosynthesis has not been examined in the species in detail, but it is clear that the
269 endolithic lifestyle has resulted in a unique physiology. Further work will shed light on the role of
270 CAO in this and other shade-tolerant species.

271 ***CAO duplications are not ancestral within the Viridiplantae***

272 To determine whether the *CAO* gene duplication was an ancestral event, we preformed a
273 phylogenetic analysis of all high-quality CAO amino acid sequences (with the exception of
274 heterodimeric *CAO* genes). Phylogenetic patterns typically match previously shown relationship
275 among major Viridiplantae groups (Figure 2), although some relationships are poorly resolved
276 (particularly among the Rosids; Supplementary Figure 1). The presence of two *CAO* genes has been
277 previously reported in rice (*OsCAO1* and *OsCAO2*) where the two genes were positioned in tandem
278 on Chromosome 10, taken as evidence for a recent gene duplication (Lee et al. 2005) but this
279 appears to be the exception rather than the rule. In addition to rice, tandem *CAO* duplicates are
280 only found in a handful of species (Table 2). In most cases, *CAO* duplicates are located on different
281 chromosomes and are paralogous, hinting to independent non-recent duplication events (Figure 2;
282 Supplementary Figure 1). Thus, we surmise that encoding for multiple *CAO* genes is not related to
283 plant evolutionary patterns; rather, it may be that *CAO* duplication occurred independently in
284 different species.

285 There are several exceptions to this pattern, mainly in very closely related species where the
286 duplication may have originated in a recent shared ancestor. There are two possible cases in the
287 Trebouxiophyceae class within the Chlorophyte green algae, where *CAO* genes detected in closely
288 related *Chlorella* species share a close phylogenetic relationship (*Chlorella sorokiniana* UTEX1230
289 and *Chlorella variabilis*; *Chlorellaceae* sp. 4966 and *Chlorellaceae* sp. 4402) (Supplementary
290 Figure 1A). Within the bryophytes, *Spaghnum magellanicum* and *Sphagnum falax* both encode for
291 homologous *CAO* genes. These two species share a very close evolutionary relationship

(Piatkowski and Shaw 2019), and it is likely that a *CAO* duplication occurred in a recent shared ancestor (Supplementary Figure 1B).

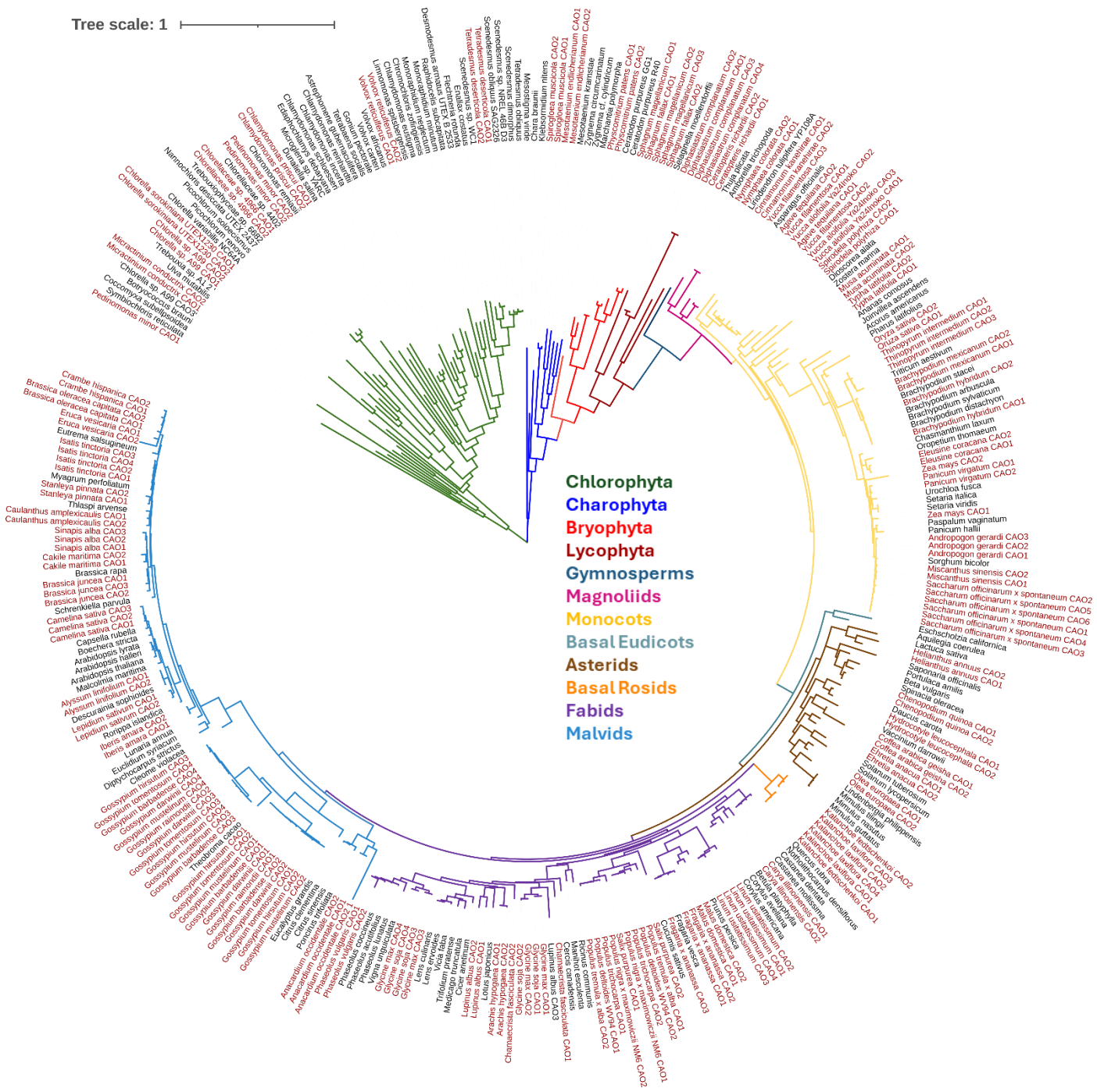

Figure 2: A phylogenetic tree of CAO genes inferred from using maximum likelihood analysis, using 326 unique CAO sequences across 246 species color-coded to represent the major Viridiplantae groups. Scale bar indicates amino acid substitutions per site. Species with more than one detected CAO gene are highlighted in red

Species that encode for multiple *CAO* genes within their genomes are more prevalent among the Angiosperms. Within the monocots, *CAO* duplication arising from a shared ancestor is likely present in the Asparagaceae, specifically among the closely related *Agave tequiliana*, *Yucca filamentosa* and *Yucca aloifolia* (Ji et al. 2023), although we were only able to detect a single *CAO* copy in *Asparagus officinalis* (Supplementary Figure 1C). We detected two *CAO* gene copies in two Brachypodium species (*B. mexicanum* and *B. hybridum*) but not in their relatives from the same genus (*B. arbuscula*, *B. stacei*, *B. sylvaticum* and *B. distachyon*). The two *CAO* genes detected in *B. hybridum* appear to me more closely related to their counterparts from other species within their genome (Supplementary Figure 1C), raising the possibility of an ancestral duplication within this genus combined with gene loss in some species. All phylogenetic studies in *Brachypodium* support a rapid and recent divergence from a common ancestor, but their precise relationship is often unresolved or weakly supported (Catalan et al. 2016) making it difficult to accurately determine evolutionary relationships between them.

CAO duplication may be ancestral in the two closely related Saxifragales *Kalanchoe laxiflora* and *Kalanchoe fedtschenkoi* (Han et al. 2024). We detected two *CAO* genes in *K. fedtschenkoi* and four in *K. laxiflora*, suggesting a duplication in a recent ancestor, followed by a second duplication within *K. laxiflora* (Supplementary Figure 1E). Within the Fabids, four *Populus* species and their relative *Salix purpurea*, each encode for two *CAO* gene copies that may be a result of an ancestral duplication. The cultivated soybean *Glycine max* and its wild relative *Glycine soja* each encode for four *CAO* gene copies (Supplementary Figure 1E) that more closely related to their homologs between the two species. The phylogenetic separation of *CAO1/2* and *CAO3/4* is puzzling and could simply be an artifact of phylogenetic uncertainty. Finally, all *Gossypium* species examined here encode for four *CAO* genes, except *G. raimondii* which encodes for two copies (Supplementary Figure 1E). The only other member of the Malvaceae examined, *Theobroma cacao,* encodes for a single *CAO* gene, and although it clusters with *CAO* from *Gossypium* sp., the relationship is poorly supported.

We show that *CAO* gene duplication is not a rarity among the Viridiplantae, but do all *CAO* genes encode for an active protein? It is tempting to speculate that species that encode for multiple *CAO* genes have a more robust capacity for Chl *b* biosynthesis through the gene dosage mechanism or a finer control over Chl *b* accumulation though gene subfunctionalization (Kondrashov 2012; Panchy et al. 2016) , but experimental evidence supporting these hypotheses is largely lacking.

Current work on species that encode for multiple CAO genes suggest that the mechanisms behind the control of Chl *b* synthesis may be species specific. Mutational analysis in rice revealed that only *OsCAO1* was necessary for Chl *b* biosynthesis. While *OsCAO1* knockout mutants exhibited low Chl *b* levels, disordered thylakoid membrane arrangements, membrane damage due to excessive ROS accumulation and decreased chilling tolerance (Lee et al. 2005; Jung et al. 2021; Xiong et al. 2024), these effects were absent in *OsCAO2* knockouts (Lee et al. 2005). In contrast, knocking out either of the two *CAO* homologs in *P. patens* lead to significantly reduced Chl *b* levels suggesting a dependency on two active CAO enzymes for the appropriate control of Chl *b* synthesis in this species (Zhang et al. 2023).

356 **Table 2:** Distribution of *CAO* duplicate genes across Viridiplantae. Only genes with confirmed Riske
357 conserved cluster, Fe-binding domain (all species), and degron sequence (Streptophytes only) are
358 listed. Accession numbers can be found in Supplementary Table 1 (continued on next page).

Taxon	Species	#CAO genes	Tandem
Chlorophyta			
Chlorophyceae	*Chlamydomonas priscui*	2	x
Chlorophyceae	*Volvox reticulferus*	2	✓
Chlorophyceae	*Tetradesmus deserticola*	2	x
Trebouxiophyceae	*Chlorella sp. A99*	2	x
Trebouxiophyceae	*Chlorella sorokiniana*	2	x
Trebouxiophyceae	*Chlorellaceae sp.*	2	x
Trebouxiophyceae	*Micractinium conductrix*	2	x
Pedinophyceae	*Pedinomonas minor*	2	x
Charophyta			
Zygnematophyceae	*Mesotaenium endlicherianum*	2	x
Zygnematophyceae	*Spirogloea muscicola*	2	x
Bryophyta			
Bryopsida	*Physcomitrium patens*	2	x
Sphagnopsida	*Sphagnum fallax*	2	x
Sphagnopsida	*Sphagnum magellanicum*	3	x
Lycophyta			
Equisetopsida	*Diphasiastrum complanatum*	4	✓✓
Equisetopsida	*Ceratopteris richardii*	2	x
Angiosperms			
Laurales	*Cinnamomum kanehirae*	2	✓
Nymphaeales	*Nymphaea colorata*	2	x
Asparagales	*Agave tequilana*	2	x
Asparagales	*Yucca aloifolia*	3	x
Asparagales	*Yucca filamentosa*	3	x
Zingiberales	*Musa acuminata v1*	2	x
Alismatales	*Spirodela polyrhiza v2*	2	x
Poales	*Typha latifolia*	2	x
Poales	*Eleusine coracana*	2	x
Poales	*Oryza sativa*	2	✓
Poales	*Thinopyrum intermedium*	3	x
Poales	*Brachypodium hybridum*	2	x
Poales	*Brachypodium mexicanum*	2	x
Poales	*Zea mays*	2	x
Poales	*Andropogon gerardi*	3	x
Poales	*Miscanthus sinensis*	2	x
Poales	*Panicum virgatum*	3	x
Poales	*Saccharum officinarum*	6	x
Caryophyllales	*Chenopodium quinoa*	2	x
Gentianales	*Coffea arabica geisha*	2	x

Table 2:(continued)

Taxon	Species	#CAO genes	Tandem
Apiales	*Hydrocotyle leucocephala*	2	x
Boraginales	*Ehretia anacua*	2	x
Asterales	*Helianthus annuus*	2	x
Lamiales	*Olea europaea*	2	x
Saxifragales	*Kalanchoe fedtschenkoi*	2	x
Saxifragales	*Kalanchoe laxiflora*	4	x
Fabales	*Arachis hypogaea*	2	x
Fabales	*Chamaecrista fasciculata*	2	x
Fabales	*Glycine max*	4	x
Fabales	*Glycine soja*	4	x
Fabales	*Lupinus albus*	3	x
Fabales	*Phaseolus vulgaris*	2	x
Fagales	*Carya illinoinensis*	2	x
Rosales	*Fragaria x ananassa*	3	x
Rosales	*Malus domestica*	2	x
Malpighiales	*Linum usitatissimum*	4	x
Malpighiales	*Populus deltoides*	2	x
Malpighiales	*Populus nigra x maximowiczii*	2	x
Malpighiales	*Populus tremula x alba*	2	x
Malpighiales	*Populus trichocarpa*	2	x
Malpighiales	*Salix purpurea*	2	x
Sapindales	*Anacardium occidentale*	3	x
Malvales	*Gossypium barbadense*	4	x
Malvales	*Gossypium darwinii*	4	x
Malvales	*Gossypium hirsutum*	4	x
Malvales	*Gossypium mustelinum*	4	x
Malvales	*Gossypium raimondii*	2	x
Malvales	*Gossypium tomentosum*	4	x
Brassicales	*Alyssum linifolium*	2	x
Brassicales	*Caulanthus amplexicaulis*	2	x
Brassicales	*Iberis amara*	2	x
Brassicales	*Lepidium sativum*	2	x
Brassicales	*Stanleya pinnata*	2	x
Brassicales	*Brassica juncea*	3	x
Brassicales	*Brassica oleracea*	2	x
Brassicales	*Cakile maritima*	2	x
Brassicales	*Crambe hispanica*	2	x
Brassicales	*Eruca vesicaria*	2	x
Brassicales	*Sinapis alba*	4	x
Brassicales	*Camelina sativa*	3	x
Brassicales	*Isatis tinctoria*	4	x

Significantly, all experimental studies to date suggest that *CAO* duplicates may be differently regulated at the level of gene expression. In rice, the expression of *OsCAO1* is light-inducible, while *OsCAO2* accumulates during periods of darkness and is expressed mainly in non-photosynthetic tissues (Lee et al. 2005). In *P. patens*, both *PpCAO1* and *PpCAO2* are expressed in both light and dark conditions, but only *PpCAO1* is upregulated with prolonged light exposure. Similarly, in two *Chlamydomonas* species that harbour *CAO* duplications (*Chlamydomonas priscui* and *Chlamydomonas* sp. ICE-MDV), only one *CAO* homolog is upregulated in response to increases in light intensity, while the second one is constitutively expressed (Poirier et al. 2025). The role of light in regulating *CAO* expression has been demonstrated previously in *Arabidopsis* and tobacco (Pattanayak et al. 2005; Tanaka and Tanaka 2005; Biswal et al. 2012, 2024), species with a single *CAO* gene copy, but the implication of light regulation is not known in species that harbour multiple *CAO* genes. Two gene copies with different expression patterns could provide a fine-tuned control of Chl *b* levels and photosynthetic performance under varying light conditions, but careful experimental work is needed to substantiate this hypothesis.

Evolutionary implications of the CAO degron sequence

In addition to light-dependant regulation at the level of transcription, the main regulation of CAO occurs post-translationally. Early work in *A. thaliana* demonstrated that this regulation is governed by the presence of a degron sequence (QDLLTIMILH) in the protein's N-terminal A domain that acts as a docking site for a chloroplast Clp protease (Yamasato et al. 2005; Nakagawara et al. 2007; Sakuraba et al. 2009). It has been hypothesized that the degron sequence is "hidden" within the A domain when there are insufficient amounts of Chl *b*, but accumulation of this product causes a conformational change in the CAO protein, exposing the degron, and giving access to the protease (Tanaka and Tanaka 2011). The evolution of this negative-feedback regulatory loop has not been examined in detail, but it has been suggested that the degron was acquired early in the Viridiplantae evolution as a mechanism for high light adaptation (Kunugi et al. 2016).

Here we show that the degron sequence is highly conserved among all land plants (86% identity), less conserved among charophytes and non-vascular plants (68% and 66% identity respectively) and very poorly conserved or absent from chlorophytes (49% identity) (Figure 3; Supplementary Figure 2). Land plants (both vascular and non-vascular) have several highly conserved sites across all species (Asparagine at position 2, Isoleucine at position 6, Histidine at position 10), whereas these positions are only moderately conserved in charophytes, and poorly conserved or absent in chlorophytes. In contrast, the Rieske cluster [CXH(X)15-17CXXH] exhibits an 84% pairwise identity among all species, where the two histidine (positions 3 and 23) and cysteines (positions 1 and 20) that coordinate the metallocentre [2Fe-2S] (Przybyla-Toscano et al. 2021) are present in all examined sequences. Nearly all residues are conserved among the Viridiplantae in the C-terminal Fe-binding domain (96% pairwise identity). Overall, these results suggest that the mechanism of Chl *b* biosynthesis involving the Rieske and Fe-binding domains is ancestral and conserved among all photosynthetic species, but this is not the case for the regulatory degron domain.

The absence of a highly conserved degron sequence may have several implications in the evolution of efficient light harvesting. For instance, the N-regulatory domain (including the degron) is completely absent in marine eukaryotic Mammiellophycea that encode for a heterdimeric *CAO* gene, as well as in their deep-water relatives from the order Palmophyllales. These species typically have low Chl *a/b* ratios and, unlike land plants, can incorporate Chl *b* into their core

antennae. This was interpreted as a necessary adaptation for efficient light harvesting in deep marine environments characterized by low levels of blue-green light (Kunugi et al. 2016).

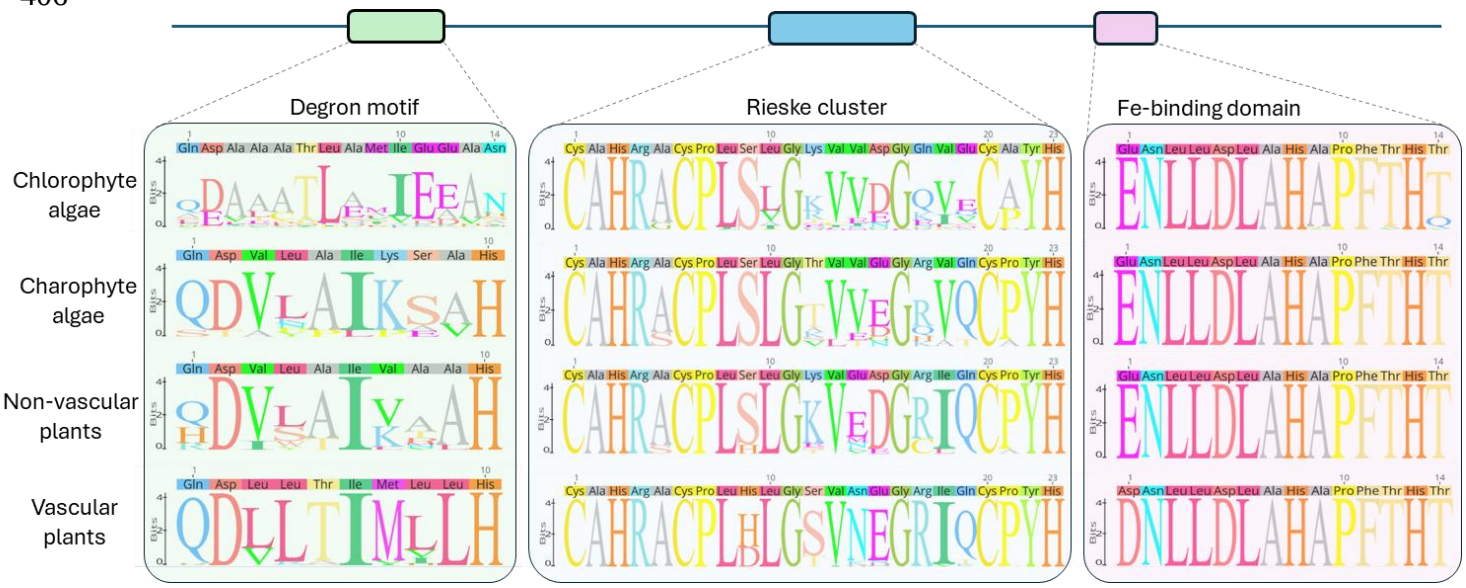

Figure 3: Sequence logos of the conserved motifs identified in the CAO protein sequence. All sequence logos are superimposed on a graphical representation of plant CAO sequences (not to scale) showing the regulatory degron motif, conserved Rieske domain cluster and Fe-binding domain (Sakuraba et al., 2009; Liu et al., 2022). The logos were generated for all chlorophyte algae, charophyte algae, non-vascular plants, and vascular plants. The sequences on top of each logo are the consensus sequence with the most common base in each sequence alignment.

It has been suggested that the degron sequence evolved in shallow-water algae and land plants as a means to regulate Chl *b* accumulation and limit Chl *b* incorporation into the reaction center in an environment where they are exposed to high light intensities (Kunugi et al. 2016). Indeedsuch species exclusively harbour Chl *a* in their core complexes, but this appears to be controlled by pigment availability rather than the chemical nature of the two chlorophyll types. The binding affinity of Chl *a* and *b* to the PSI complex is similar (Ikegami et al. 2007) and cyanobacteria that do not naturally produce Chl *b* successfully incorporated this pigment into their reaction centers when transformed with plant CAO (Satoh et al. 2001; Xu et al. 2001). Incorporation of Chl *b* into the reaction center accompanied by increased Chl *b* accumulation was also observed in *Arabidopsis* when full length CAO was overexpressed or replaced with a prokaryotic CAO that lacked a degron sequence (Hirashima et al. 2006; Sakuraba et al. 2009).

The chlorophyll cycle has not been examined in detail in the core chlorophytes (Chlorophyceae, Trebouxiophyceae, Ulvophyceae), a diverse group that inhabits marine, freshwater and terrestrial habitats (Leliaert et al. 2012). Although many chlorophytes thrive in high light, others are adapted to light limited environments, including deep waters. As a result, these species have developed adaptations allowing for efficient light harvesting in low light. For instance, *Chlamydomonas priscui* is endemic to the depths of the Antarctic Lake Bonney, an environment characterized by extreme shading and dominated by blue-green wavelengths (Neale and Priscu 1995). This alga displays a host of adaptations to extreme shading including a large LHC antennae size (reviewed in

Cvetkovska et al. 2017; Hüner et al. 2022). As suggested for marine prasinophytes (Kunugi et al. 2016), such low-light adapted chlorophytes could benefit from increased Chl *b* levels and could thrive without a Chl *b*-dependent feedback mechanisms.

Despite the lack of a conserved degron sequence to regulate CAO protein levels (Figure 2), chlorophytes generally have maintained the ability to alter their Chl *a/b* ratio in response to light (Leliaert et al. 2012). Transcriptional control of CAO levels was postulated to be a minor aspect of adjusting Chl *b* levels in plants (Tanaka and Tanaka 2019), but this process could play a bigger role in chlorophytes. Finally, it is possible that a different, yet unidentified, mechanism exists for Chl *b* turnover independent of a degron sequence, allowing these species to modify their Chl *a/b* ratios. A detailed experimental investigation on the control and mechanism of Chl *b* biosynthesis is clearly needed among the Chlorophytes.

Conclusion and Drawbacks

Our work suggests that CAO gene duplication is widespread among the Viridiplantae, likely originating from multiple independent duplication events throughout plant evolution. While we show that the presence of multiple *CAO* gene copies is highly prevalent in land plants, these results must be interpreted with caution. Many more angiosperm genomes are available in public databases, while other plant and algal groups are not as well represented. Furthermore, many genome sequencing initiatives are focused on plants of economic and agricultural importance, in which domestication and selective breeding may have affectted the genomic content (Turner-Hissong et al. 2020).While the exact impact of CAO duplication is unknown, the study of transgenic tobacco overexpressing CAO (Biswal et al. 2012) showed that an increase in Chl *b* may be a strategy to enhance CO_2 assimilation, delay senescence, and enhance productivity - traits that would be attractive in crop breeding programs (Voitsekhovskaja and Tyutereva 2015). Therefore, the high prevalence of CAO duplication seen in land plants in this study may have been selected for during plant breeding, as the traits resulting from the duplication may enhance crop productivity. Our understanding of the prevalence and evolution of CAO at the genetic level will be improved in future years, as more algal and non-vascular plant genomes become available. Furthermore, there are still numerous complex and outstanding questions regarding the control of Chl *b* accumulation in photosynthetic species. In particular, the chlorophytes are an understudied group when it comes to Chl *b* regulating mechanism. Detailed genomic and experimental studies will shed light on the role and regulation of Chl *b* in these species and will deepen our understanding on this crucial photosynthetic pigment.

Supplementary Information

Supplementary Table 1: A summary of genes encoding for chlorophyll *a* oxygenase (CAO) across the Viridiplantae used in this work. All sequences were obtained from Phycocosm (Chlorophytes, Euglenophytes, Cercozoa, Charophytes) or Phytozome (Bryophytes, Marchantiophytes, Lycophytes, Gymnosperms, Angiosperms). Accession numbers represent the unique IDs assigned to each gene (Phycocosm: Gene portal|Protein ID|Gene Model; Phytozome: Gene Model).

Supplementary Table 2: A summary of the occurrence of CAO duplicate genes across the major groups within Viridiplantae. The number in brackets represents the proportion of species which encode for multiple copies of the CAO gene (Mutiple CAO) or encode for two CAO genes that form heterodimeric complex (Heterodimeric CAO).

Supplementary Figure 1: A phylogenetic tree of CAO genes inferred from using maximum likelihood analysis. The node labels show the bootstrap support, with dashed lines indicating <50 bootstrap support. Scale bar indicates 0.1 amino acid substitutions per site. Species with more than one detected CAO gene are highlighted in red. Cases where CAO duplication may have arisen in common shared ancestor are highlighted with a blue arrow. The inset shows the outline of the full CAO phylogenetic tree inferred from using maximum likelihood analysis, using 326 unique CAO sequences across 246 species color coded to represent the major Viridiplantae groups. The position of the chlorophytes (green) is highlighted with a black box. (A) Chlorophytes; (B) Charophytes, bryophytes (including marachantiophytes), gymnosperms and magnolids; (C) Monocots; (D) Basal Eudicots and Asterids; (E) Basal Rosids and Fabids; (F) Malvids.

Supplementary Figure 2: Multiple sequence alignment of the predicted CAO amino acid sequences from Chlorophytes (A), Charophytes (B), non-vascular plants (C) and vascular plants (D). Global alignments were performed with Clustal Omega. The colour of the residue depends on the degree of conservation among species (blue, fully conserved; green, 80–100% similar; yellow, 60–80% similar; gray, < 60% similar). As a reference, the CAO sequence from Arabidopsis thaliana is present in all alignments (top), with the key domains annotated (green, Degron sequence; blue, conserved Rieske cluster; purple, Fe-binding domain)

References

Arimoto A, Nishitsuji K, Higa Y, et al (2019) A siphonous macroalgal genome suggests convergent functions of homeobox genes in algae and land plants. DNA Research 26:183–192. https://doi.org/10.1093/dnares/dsz002

Biswal AK, Pattanayak GK, Pandey SS, et al (2012) Light Intensity-Dependent Modulation of Chlorophyll *b* Biosynthesis and Photosynthesis by Overexpression of Chlorophyllide *a* Oxygenase in Tobacco. Plant Physiol 159:433–449. https://doi.org/10.1104/pp.112.195859

Biswal AK, Pattanayak GK, Ruhil K, et al (2024) Reduced expression of chlorophyllide *a* oxygenase (CAO) decreases the metabolic flux for chlorophyll synthesis and downregulates photosynthesis in tobacco plants. Physiol Mol Biol Plants 30:1–16. https://doi.org/10.1007/s12298-023-01395-5

Blanc-Mathieu R, Verhelst B, Derelle E, et al (2014) An improved genome of the model marine alga *Ostreococcus tauri* unfolds by assessing Illumina de novo assemblies. BMC Genomics 15:1–12. https://doi.org/10.1186/1471-2164-15-1103

Blankenship RE (2010) Early Evolution of Photosynthesis. Plant Physiol 154:434–438. https://doi.org/10.1104/pp.110.161687

Brzezowski P, Richter AS, Grimm B (2015) Regulation and function of tetrapyrrole biosynthesis in plants and algae. Biochimica et Biophysica Acta - Bioenergetics 1847:968–985. https://doi.org/10.1016/j.bbabio.2015.05.007

Bujaldon S, Kodama N, Rappaport F, et al (2017) Functional Accumulation of Antenna Proteins in Chlorophyll b-Less Mutants of *Chlamydomonas reinhardtii*. Molecular Plant 10:115–130. https://doi.org/10.1016/j.molp.2016.10.001

Cardona T (2019) Thinking twice about the evolution of photosynthesis. Open Biology 9:180246. https://doi.org/10.1098/rsob.180246

Castresana J (2000) Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Molecular Biology and Evolution 17:540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Catalan P, López-Álvarez D, Díaz-Pérez A, et al (2016) Phylogeny and Evolution of the Genus *Brachypodium*. In: Vogel JP (ed) Genetics and Genomics of Brachypodium. Springer International Publishing, Cham, pp 9–38

Conlon T, Parkes R, Fierli D, Touzet N (2024) Comparative pigment and fatty acid profiling of marine species within the chlorophyte genus *Tetraselmis*. Food Bioscience 58:103660. https://doi.org/10.1016/j.fbio.2024.103660

Curtis BA, Tanifuji G, Burki F, et al (2012) Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature 492:59–65. https://doi.org/10.1038/nature11681

Cvetkovska M, Hüner NPA, Smith DR (2017) Chilling out: the evolution and diversification of psychrophilic algae with a focus on Chlamydomonadales. Polar Biol 40:1169–1184. https://doi.org/10.1007/s00300-016-2045-4

Cvetkovska M, Orgnero S, Hüner NPA, Smith DR (2019) The enigmatic loss of light-independent chlorophyll biosynthesis from an Antarctic green alga in a light-limited environment. New Phytol 222:651–656. https://doi.org/10.1111/nph.15623

de Vries J, Archibald JM (2018) Plant evolution: landmarks on the path to terrestrial life. New Phytologist 217:1428–1434. https://doi.org/10.1111/nph.14975

Dey D, Tanaka R, Ito H (2023) Structural Characterization of the Chlorophyllide *a* Oxygenase (CAO) Enzyme Through an In Silico Approach. J Mol Evol 91:225–235. https://doi.org/10.1007/s00239-023-10100-9

Dumont T, Barau L, Thong-Chane A, et al (2022) Sugarcane Breeding in Reunion: Challenges, Achievements and Future Prospects. Sugar Tech 24:181–192. https://doi.org/10.1007/s12355-021-00998-0

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113. https://doi.org/10.1186/1471-2105-5-113

Espineda CE, Linford AS, Devine D, Brusslan JA (1999) The AtCAO gene, encoding chlorophyll *a* oxygenase, is required for chlorophyll *b* synthesis in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America 96:10507–10511. https://doi.org/10.1073/pnas.96.18.10507

Ferraro DJ, Gakhar L, Ramaswamy S (2005) Rieske business: Structure–function of Rieske non-heme oxygenases. Biochemical and Biophysical Research Communications 338:175–190. https://doi.org/10.1016/j.bbrc.2005.08.222

549 Goodstein DM, Shu S, Howson R, et al (2012) Phytozome: a comparative platform for green plant
550 genomics. Nucleic Acids Research 40:D1178–D1186. https://doi.org/10.1093/nar/gkr944

551 Gray J, Wardzala E, Yang M, et al (2004) A small family of LLS1-related non-heme oxygenases in
552 plants with an origin amongst oxygenic photosynthesizers. Plant Molecular Biology 54:39–
553 54

554 Grigoriev IV, Hayes RD, Calhoun S, et al (2021) PhycoCosm, a comparative algal genomics
555 resource. Nucleic Acids Research 49:D1004–D1011. https://doi.org/10.1093/nar/gkaa898

556 Guiry MD (2024) How many species of algae are there? A reprise. Four kingdoms, 14 phyla, 63
557 classes and still growing. Journal of Phycology 60:214–228.
558 https://doi.org/10.1111/jpy.13431

559 Han S, Zhang S, Yi R, et al (2024) Phylogenomics and plastomics offer new evolutionary
560 perspectives on Kalanchoideae (Crassulaceae). Annals of Botany 133:585–604.
561 https://doi.org/10.1093/aob/mcae017

562 Healey AL, Garsmeur O, Lovell JT, et al (2024) The complex polyploid genome architecture of
563 sugarcane. Nature 628:804–810. https://doi.org/10.1038/s41586-024-07231-4

564 Hirashima M, Satoh S, Tanaka R, Tanaka A (2006) Pigment Shuffling in Antenna Systems Achieved
565 by Expressing Prokaryotic Chlorophyllide a Oxygenase in *Arabidopsis*. The Journal of
566 Biological Chemistry 281:15385–15393. https://doi.org/10.1074/jbc.M602903200

567 Hüner NPA, Smith DR, Cvetkovska M, et al (2022) Photosynthetic adaptation to polar life: Energy
568 balance, photoprotection and genetic redundancy. Journal of Plant Physiology 268:153557.
569 https://doi.org/10.1016/j.jplph.2021.153557

570 Iha C, Dougan KE, Varela JA, et al (2021) Genomic adaptations to an endolithic lifestyle in the coral-
571 associated alga *Ostreobium*. Current Biology 31:1393-1402.e5.
572 https://doi.org/10.1016/j.cub.2021.01.018

573 Ikegami I, Satoh S, Aoki M (2007) Binding Affinity of Chl *b* for the Chl *a* -Binding Sites in PSI Core
574 Complexes. Plant Cell Physiology 48:1092–1097. https://doi.org/10.1093/pcp/pcm078

575 Ji Y, Landis JB, Yang J, et al (2023) Phylogeny and evolution of Asparagaceae subfamily Nolinoideae:
576 new insights from plastid phylogenomics. Annals of Botany 131:301–312.
577 https://doi.org/10.1093/aob/mcac144

578 Jung YJ, Lee HJ, Yu J, et al (2021) Transcriptomic and physiological analysis of OsCAO1 knockout
579 lines using the CRISPR/Cas9 system in rice. Plant Cell Rep 40:1013–1024.
580 https://doi.org/10.1007/s00299-020-02607-y

581 Junkins EN, Stamps BW, Corsetti FA, et al (2019) Draft Genome Sequence of *Picocystis* sp. Strain
582 ML, Cultivated from Mono Lake, California. Microbiol Resour Announc 8:e01353-18.
583 https://doi.org/10.1128/MRA.01353-18

584 Kim EH, Li XP, Razeghifard R, et al (2009) The multiple roles of light-harvesting chlorophyll *a/b*-
585 protein complexes define structure and optimize function of *Arabidopsis* chloroplasts: A

study using two chlorophyll b-less mutants. Biochimica et Biophysica Acta - Bioenergetics 1787:973–984. https://doi.org/10.1016/j.bbabio.2009.04.009

Kondrashov FA (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. Proceedings of the Royal Society B: Biological Sciences 279:5048–5057. https://doi.org/10.1098/rspb.2012.1108

Król M, Spangfort MD, Huner NPA, et al (1995) Chlorophyll *a/b*-binding proteins, pigment conversions, and early light-induced proteins in a chlorophyll *b*-less Barley mutant. Plant Physiology 107:873–883. https://doi.org/10.1104/pp.107.3.873

Kume A, Akitsu T, Nasahara KN (2018) Why is chlorophyll *b* only used in light-harvesting systems? Journal of Plant Research 131:961–972. https://doi.org/10.1007/s10265-018-1052-7

Kunugi M, Satoh S, Ihara K, et al (2016) Evolution of Green Plants Accompanied Changes in Light-Harvesting Systems. Plant and Cell Physiology 57:1231–1243. https://doi.org/10.1093/pcp/pcw071

Kunugi M, Takabayashi A, Tanaka A (2013) Evolutionary changes in chlorophyllide *a* oxygenase (CAO) structure contribute to the acquisition of a new light-harvesting complex in micromonas. Journal of Biological Chemistry 288:19330–19341. https://doi.org/10.1074/jbc.M113.462663

Lamesch P, Berardini TZ, Li D, et al (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 40:D1202-1210. https://doi.org/10.1093/nar/gkr1090

Lee S, Kim JH, Eun SY, et al (2005) Differential regulation of chlorophyll *a* oxygenase genes in rice. Plant Molecular Biology 57:805–818. https://doi.org/10.1007/s11103-005-2066-9

Leliaert F, Smith DR, Moreau H, et al (2012) Phylogeny and Molecular Evolution of the Green Algae. Critical Reviews in Plant Sciences 31:1–46. https://doi.org/10.1080/07352689.2011.615705

Lemieux C, Turmel M, Otis C, Pombert J-F (2019) A streamlined and predominantly diploid genome in the tiny marine green alga *Chloropicon primus*. Nat Commun 10:4061. https://doi.org/10.1038/s41467-019-12014-x

Letunic I, 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

Liu J, Knapp M, Jo M, et al (2022) Rieske Oxygenase Catalyzed C–H Bond Functionalization Reactions in Chlorophyll *b* Biosynthesis. ACS Cent Sci 8:1393–1403. https://doi.org/10.1021/acscentsci.2c00058

Massé A, Tribollet A, Meziane T, et al (2020) Functional diversity of microboring Ostreobium algae isolated from corals. Environmental Microbiology 22:4825–4846. https://doi.org/10.1111/1462-2920.15256

Merchant SS, Prochnik SE, Vallon O, et al (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318:245–250. https://doi.org/10.1126/science.1143609

Miller MA, Schwartz T, Pickett BE, et al (2015) A RESTful API for Access to Phylogenetic Tools via the CIPRES Science Gateway. Evol Bioinform Online 11:43–48. https://doi.org/10.4137/EBO.S21501

Moreau H, Verhelst B, Couloux A, et al (2012) Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. Genome Biol 13:1–16. https://doi.org/10.1186/gb-2012-13-8-r74

Mueller AH, Dockter C, Gough SP, Lundqvist U (2012) Characterization of Mutations in Barley fch2 Encoding Chlorophyllide *a* Oxygenase. Plant and Cell Physiology 53:1232–1246. https://doi.org/10.1093/pcp/pcs062

Nagata N, Satoh S, Tanaka R, Tanaka A (2004) Domain structures of chlorophyllide *a* oxygenase of green plants and Prochlorothrix hollandica in relation to catalytic functions. Planta 218:1019–1025. https://doi.org/10.1007/s00425-003-1181-6

Nakagawara E, Sakuraba Y, Yamasato A, et al (2007) Clp protease controls chlorophyll *b* synthesis by regulating the level of chlorophyllide a oxygenase. The Plant Journal 49:800–809. https://doi.org/10.1111/j.1365-313X.2006.02996.x

Neale PJ, Priscu JC (1995) The Photosynthetic Apparatus of Phytoplankton from a Perennially Ice-Covered Antarctic Lake: Acclimation to an Extreme Shade Environment. Plant and Cell Physiology 36:253–263. https://doi.org/10.1093/oxfordjournals.pcp.a078757

Neilson JAD, Durnford DG (2010) Structural and functional diversification of the light-harvesting complexes in photosynthetic eukaryotes. Photosynthesis Research 106:57–71. https://doi.org/10.1007/s11120-010-9576-2

Nick S, Meurer J, Soll J, Ankele E (2013) Nucleus-encoded light-harvesting chlorophyll a/b proteins are imported normally into chlorophyll b-free chloroplasts of Arabidopsis. Molecular Plant 6:860–871. https://doi.org/10.1093/mp/sss113

Oster U, Tanaka R, Tanaka A, Rüdiger W (2000) Cloning and functional expression of the gene encoding the key enzyme for chlorophyll *b* biosynthesis (CAO) from *Arabidopsis thaliana*. Plant Journal 21:305–310. https://doi.org/10.1046/j.1365-313X.2000.00672.x

Palenik B, Grimwood J, Aerts A, et al (2007) The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. Proceedings of the National Academy of Sciences 104:7705–7710. https://doi.org/10.1073/pnas.0611046104

Palenik B, Haselkorn R (1992) Multiple evolutionary origins of prochlorophytes, the chlorophyll *b*-containing prokaryotes. Nature 355:265–267. https://doi.org/10.1038/355265a0

Panchy N, Lehti-Shiu M, Shiu S-H (2016) Evolution of Gene Duplication in Plants. Plant Physiol 171:2294–2316. https://doi.org/10.1104/pp.16.00523

Partensky F, Six C, Ratin M, et al (2018) A novel species of the marine cyanobacterium Acaryochloris with a unique pigment content and lifestyle. Sci Rep 8:9142. https://doi.org/10.1038/s41598-018-27542-7

Pattanayak GK, Biswal AK, Reddy VS, Tripathy BC (2005) Light-dependent regulation of chlorophyll *b* biosynthesis in chlorophyllide *a* oxygenase overexpressing tobacco plants. Biochemical and Biophysical Research Communications 326:466–471. https://doi.org/10.1016/j.bbrc.2004.11.049

Piatkowski BT, Shaw AJ (2019) Functional trait evolution in Sphagnum peat mosses and its relationship to niche construction. New Phytologist 223:939–949. https://doi.org/10.1111/nph.15825

Poirier MC, Fugard K, Cvetkovska M (2025) Light quality affects chlorophyll biosynthesis and photosynthetic performance in Antarctic Chlamydomonas. Photosynth Res 163:9. https://doi.org/10.1007/s11120-024-01127-0

Pombert J-F, Blouin NA, Lane C, et al (2014) A Lack of Parasitic Reduction in the Obligate Parasitic Green Alga Helicosporidium. PLoS Genet 10:e1004355. https://doi.org/10.1371/journal.pgen.1004355

Pružinská A, Tanner G, Anders I, et al (2003) Chlorophyll breakdown: Pheophorbide *a* oxygenase is a Rieske-type iron–sulfur protein, encoded by the accelerated cell death 1 gene. Proceedings of the National Academy of Sciences 100:15259–15264. https://doi.org/10.1073/pnas.2036571100

Przybyla-Toscano J, Boussardon C, Law SR, et al (2021) Gene atlas of iron-containing proteins in *Arabidopsis thaliana*. The Plant Journal 106:258–274. https://doi.org/10.1111/tpj.15154

Reinbothe C, Bakkouri ME, Buhr F, et al (2010) Chlorophyll biosynthesis: spotlight on protochlorophyllide reduction. Trends in Plant Science 15:614–624. https://doi.org/10.1016/j.tplants.2010.07.002

Reinbothe C, Bartsch S, Eggink LL, et al (2006) A role for chlorophyllide *a* oxygenase in the regulated import and stabilization of light-harvesting chlorophyll *a/b* proteins. Proceedings of the National Academy of Sciences of the United States of America 103:4777–4782. https://doi.org/10.1073/pnas.0511066103

Rensing SA (2014) Gene duplication as a driver of plant morphogenetic evolution. Current Opinion in Plant Biology 17:43–48. https://doi.org/10.1016/j.pbi.2013.11.002

Roche JL, van der Staay GWM, Partensky F, et al (1996) Independent evolution of the prochlorophyte and green plant chlorophyll *a/b* light-harvesting proteins. Proceedings of the National Academy of Sciences 93:15244–15248. https://doi.org/10.1073/pnas.93.26.15244

Roesler CS, Culbertson CW, Etheridge SM, et al (2002) Distribution, production, and ecophysiology of *Picocystis* strain ML in Mono Lake, California. Limnology and Oceanography 47:440–452. https://doi.org/10.4319/lo.2002.47.2.0440

Ruggiero MA, Gordon DP, Orrell TM, et al (2015) A Higher Level Classification of All Living Organisms. PLoS One 10:e0119248. https://doi.org/10.1371/journal.pone.0119248

Sakuraba Y, Tanaka R, Yamasato A, Tanaka A (2009) Determination of a chloroplast degron in the regulatory domain of chlorophyllide *a* oxygenase. Journal of Biological Chemistry 284:36689–36699. https://doi.org/10.1074/jbc.M109.008144

Sánchez-Baracaldo P, Cardona T (2020) On the origin of oxygenic photosynthesis and Cyanobacteria. New Phytologist 225:1440–1446. https://doi.org/10.1111/nph.16249

Satoh S, Ikeuchi M, Mimuro M, Tanaka A (2001) Chlorophyll *b* Expressed in Cyanobacteria Functions as a Light-harvesting Antenna in Photosystem I through Flexibility of the Proteins. The Journal of Biological Chemistry 276:4293–4297. https://doi.org/10.1074/jbc.M008238200

Schmidt CL, Shaw L (2001) A Comprehensive Phylogenetic Analysis of Rieske and Rieske-Type Iron-Sulfur Proteins. J Bioenerg Biomembr 33:9–26. https://doi.org/10.1023/A:1005616505962

Schumacher I, Menghini D, Ovinnikov S, et al (2022) Evolution of chlorophyll degradation is associated with plant transition to land. The Plant Journal 109:1473–1488. https://doi.org/10.1111/tpj.15645

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Steadman Tyler CR, Hovde BT, Daligault HE, et al (2019) High-Quality Draft Genome Sequence of the Green Alga *Tetraselmis striata* (Chlorophyta) Generated from PacBio Sequencing. Microbiology Resource Announcements 8:10.1128/mra.00780-19. https://doi.org/10.1128/mra.00780-19

Suzuki S, Hirakawa Y, Kofuji R, et al (2016) Plastid genome sequences of *Gymnochlora stellata*, *Lotharella vacuolata*, and *Partenskyella glossopodia* reveal remarkable structural conservation among chlorarachniophyte species. J Plant Res 129:581–590. https://doi.org/10.1007/s10265-016-0804-5

Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56:564–577. https://doi.org/10.1080/10635150701472164

Tanaka A, Ito H, Tanaka R, et al (1998) Chlorophyll *a* oxygenase (CAO) is involved in chlorophyll *b* formation from chlorophyll *a*. Proceedings of the National Academy of Sciences of the United States of America 95:12719–12723. https://doi.org/10.1073/pnas.95.21.12719

Tanaka A, Tanaka R (2019) The biochemistry, physiology, and evolution of the chlorophyll cycle, 1st edn. Elsevier Ltd.

Tanaka R, Tanaka A (2007) Tetrapyrrole biosynthesis in higher plants. Annual Review of Plant Biology 58:321–346. https://doi.org/10.1146/annurev.arplant.57.032905.105448

Tanaka R, Tanaka A (2005) Effects of chlorophyllide *a* oxygenase overexpression on light acclimation in *Arabidopsis thaliana*. Photosynth Res 85:327–340. https://doi.org/10.1007/s11120-005-6807-z

Tanaka R, Tanaka A (2011) Chlorophyll cycle regulates the construction and destruction of the light-harvesting complexes. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1807:968–976. https://doi.org/10.1016/j.bbabio.2011.01.002

Tartar A, Boucias DG, Becnel JJ, Adams BJ (2003) Comparison of plastid 16S rRNA (rrn16) genes from *Helicosporidium* spp.: evidence supporting the reclassification of Helicosporidia as green algae (Chlorophyta). Int J Syst Evol Microbiol 53:1719–1723. https://doi.org/10.1099/ijs.0.02559-0

Tomitani A, Okada K, Miyashita H, et al (1999) Chlorophyll *b* and phycobilins in the common ancestor of cyanobacteria and chloroplasts. Nature 400:159–162. https://doi.org/10.1038/22101

Turner-Hissong SD, Mabry ME, Beissinger TM, et al (2020) Evolutionary insights into plant breeding. Current Opinion in Plant Biology 54:93–100. https://doi.org/10.1016/j.pbi.2020.03.003

Ueno Y, Aikawa S, Kondo A, Akimoto S (2019) Adaptation of light-harvesting functions of unicellular green algae to different light qualities. Photosynthesis Research 139:145–154. https://doi.org/10.1007/s11120-018-0523-y

Voitsekhovskaja OV, Tyutereva EV (2015) Chlorophyll *b* in angiosperms: Functions in photosynthesis, signaling and ontogenetic regulation. Journal of Plant Physiology 189:51–64. https://doi.org/10.1016/j.jplph.2015.09.013

Willows RD (2020) Biosynthesis of Chlorophyll and Bilins in Algae. In: Larkum AWD, Grossman AR, Raven JA (eds) Advances in Photosynthesis and Respiration, 45th edn. Springer Nature, pp 83–103

Worden AZ, Lee J-H, Mock T, et al (2009) Green Evolution and Dynamic Adaptations Revealed by Genomes of the Marine Picoeukaryotes Micromonas. Science 324:268–272. https://doi.org/10.1126/science.1167222

Xiong J, Bauer CE (2002) COMPLEX EVOLUTION OF PHOTOSYNTHESIS. Annual Review of Plant Biology 53:503–521. https://doi.org/10.1146/annurev.arplant.53.100301.135212

Xiong J, Wen G, Song J, et al (2024) Knockout of the Chlorophyll *a* Oxygenase Gene OsCAO1 Reduces Chilling Tolerance in Rice Seedlings. Genes 15:721. https://doi.org/10.3390/genes15060721

Xu H, Vavilin D, Vermaas W (2001) Chlorophyll *b* can serve as the major pigment in functional photosystem II complexes of cyanobacteria. Proceedings of the National Academy of Sciences of the United States of America 98:14168–14173. https://doi.org/10.1073/pnas.251530298

Yamasato A, Nagata N, Tanaka R, Tanaka A (2005) The N-Terminal Domain of Chlorophyllide a Oxygenase Confers Protein Instability in Response to Chlorophyll *b* Accumulation in Arabidopsis. Plant Cell 17:1585–1597. https://doi.org/10.1105/tpc.105.031518

Zhang L, Yang C, Liu C (2023) Revealing the significance of chlorophyll *b* in the moss *Physcomitrium patens* by knocking out two functional chlorophyllide a oxygenase. Photosynthesis Research 158:171–180. https://doi.org/10.1007/s11120-023-01044-8