Chlorophyllide *a* oxygenase (CAO) gene duplication 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

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 playing a role in light harvesting, chlorophyll *b* is required at the appropriate level for the 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, regulates 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 poorly conserved in chlorophytes, but first appears as a conserved sequence in charophytes, and is very highly conserved among the embryophytes. The evolutionary history and functional role 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

Chlorophyll biosynthesis, chlorophyllide *a* oxygenase, gene duplication, degron, plants and algae

Introduction

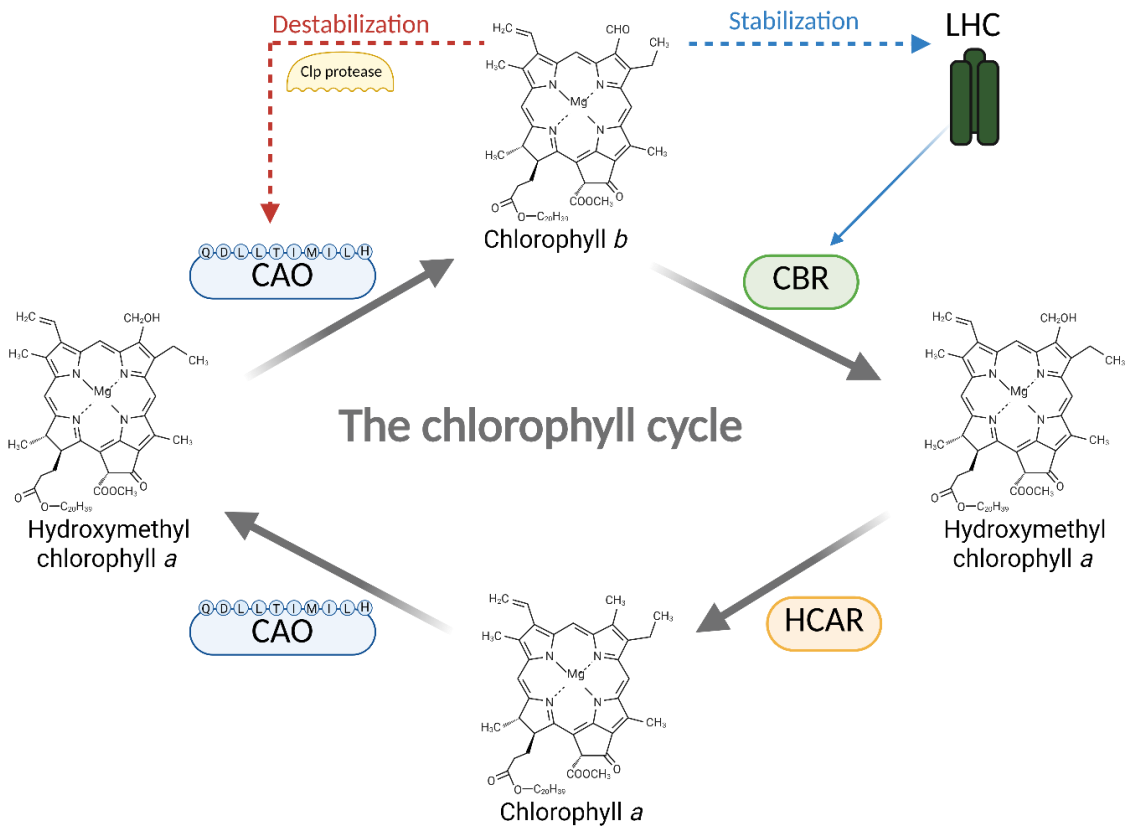
Photosynthetic organisms form the base of most food chains and provide the organic carbon compounds that support life in our biosphere. The capture of light energy by chlorophyll is the first step in the process of photosynthesis that fixes atmospheric CO_2 into stable organic compounds. Overwhelming 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) have primary chloroplasts derived from an ancient endosymbiosis, and 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 playing a role in light harvesting, Chl *b* is required at appropriate levels for the 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, which 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 complex multistep pathway (reviewed in Brzezowski et al. 2015; Willows 2020). 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* by chlorophyll *b* reductase (CBR) and 7-hydroxymethyl reductase (HCAR)

107 to complete the cycle (Figure 1; reviewed in detail in Tanaka and Tanaka, 2019). The chlorophyll
108 cycle plays an essential role in many biological processes, including biogenesis of LHCs, antenna
109 size regulation during light acclimation, and chlorophyll degradation during senescence. While
110 many of the intricacies of how this pathway is regulated are still not fully understood (Tanaka and
111 Tanaka 2019), it has been suggested that CAO levels and activity are key determinants of Chl *b*
112 accumulation.



113 **Figure 1**. The chlorophyll cycle in land plants. Chlorophyll *a* is converted to chlorophyll *b* via the
114 intermediate 7-hydromethyl chlorophyll *a*, by the action of the enzyme chlorophyllide *a* oxygenase
115 (CAO). In a negative feedback loop, high levels of chlorophyll *b* trigger the binding of a stromal Clp
116 protease to the degron motif (QDLLTIMILH) and the subsequent destabilization of CAO (dashed red
117 line). Chlorophyll *b* is necessary for the stabilization of Light Harvesting Complexes (LHC) in plants
118 (dashed blue line). Excessive amounts of LHC or accumulation of energetically uncoupled non-
119 functional LHCs (e.g., during senescence), triggers the accumulation of Chlorophyll *b* reductase
120 (CBR; blue line) and the conversion of chlorophyll *b* to chlorophyll *a* via the action of CBR and 7-
121 hydroxymethyl reductase (HCAR). This regulatory mechanism of CAO and CBR enables the fine-
122 tuning and optimization of chlorophyll *b* and LHC levels (adapted from Tanaka and Tanaka 2019).

123 CAO is part of the Rieske-type mononuclear non-heme iron oxygenase group of proteins (Gray et al.
124 2004). It catalyzes the conversion of the methyl group of Chl *a* into a formyl group to form Chl *b*
125 through two oxygenation reactions, likely using ferredoxin as a reductant (Oster et al. 2000). CAO
126 has three main regions: an N-terminal regulatory domain and a C-terminal catalytic domain, which
127 are connected by a more variable linker region (Nagata et al. 2004). The catalytic domain contains a

128 Rieske cluster and a mononuclear iron-binding domain, both of which are necessary for the
129 enzymatic activity of CAO (Yamasato et al. 2005; Kunugi et al. 2013). The regulatory domain has
130 been characterized in detail only in land plants. In *Arabidopsis*, a conserved sequence
131 (QDLLTIMILH) termed a degron was demonstrated to regulate Chl *b* accumulation by affecting CAO
132 protein turnover. In a negative feedback loop, high levels of Chl *b* trigger the binding of a stromal
133 Clp protease to the degron motif leading to destabilization of the CAO protein; this, in turn, halts
134 Chl *b* biosynthesis (Sakuraba et al., 2009; Nakagawara et al., 2007).

135 The evolutionary origins of CAO can be traced back to prochlorophytes, which encode a CAO gene
136 homolog (Tomitani et al. 1999; Nagata et al. 2004). Notably, the prokaryotic CAO-like enzymes lack
137 the regulatory N-terminal domain, suggesting that the Chl *b*-dependent mechanism for controlling
138 CAO stability was acquired in eukaryotes. Recent work on 9 chlorophyte, 2 charophyte, and 3
139 embryophyte species suggested that the degron recognition sequence first appeared in
140 charophytes as an adaptation to high-light environments in shallow water (Kunugi et al. 2016).
141 These conclusions, however, were based on a very limited sample of representative species and
142 the presence of a degron in the CAO sequence in charophytes was inferred based on a single
143 species (*Klebsormidinium flaccidum*). This hypothesis requires further validation.

144 The CAO protein is encoded by a single gene in most plants and algae (Kunugi et al. 2016;
145 Schumacher et al. 2022), with only a very few reported exceptions. Two CAO genes were detected
146 in rice (Lee et al. 2005) although only one of them is necessary for Chl *b* synthesis (Jung et al. 2021).
147 Recent work revealed CAO duplication in the bryophyte *Physcomitrium patens,* where both genes
148 contribute to Chl *b* accumulation (Zhang et al. 2023). CAO duplication was also reported in two
149 Antarctic *Chlamydomonas* species , where the CAO paralogs in each species appear to be derived
150 from independent duplications events rather than an ancestral one (Cvetkovska et al. 2019; Poirier
151 et al. 2025).

152 These insights prompted us to further examine the *CAO* gene content and the presence of the
153 regulatory domain across the diverse plant and algal groups. A systematic examination for the
154 presence of *CAO* across the Viridiplantae has not yet been preformed. In this work we took
155 advantage of the recent increase in publicly available plant and algal genomes to examine the
156 occurrence of CAO genes in diverse species across Viridiplantae and demonstrate a widespread
157 occurrence of CAO duplicates in this group. Furthermore, building on the work by Kunugi et al.
158 (2016), we have examined the occurrence of a degron sequence in 326 unique CAO sequences in
159 algae and plants. We demonstrate that this regulatory region is poorly conserved in chlorophytes,
160 compared to charophyte and embryophyte CAO sequences, suggesting a different mechanism for
161 Chl *b* regulation in early aquatic lineages.

162 **Materials and Methods**

163 ***CAO sequence identification and analyses***

164 To identify *CAO* sequences and duplication events with high accuracy, we used publicly available,
165 high quality, and fully annotated genomes. Chlorophyte and streptophyte genomes were retrieved
166 from Phycocosm (Grigoriev et al. 2021), while embryophyte genomes were obtained from
167 Phytozome v13 (Goodstein et al. 2012). Putative *CAO* genes were identified with a tBLASTn search

168 with the *Chlamydomonas reinhardtii* (CrCAO; Cre01.g043350) and the *Arabidopsis thaliana*
169 (AtCAO; AT1G44446) full-length CAO peptides (Merchant et al. 2007; Lamesch et al. 2012). A strict
170 e-value cutoff of ≤e-30 was taken as a threshold, to avoid many unspecific hits due to the
171 conserved nature of the Fe-binding and Rieske cluster domains present among diverse plant and
172 algal proteins (Schmidt and Shaw 2001; Ferraro et al. 2005; Przybyla-Toscano et al. 2021). Putative
173 *CAO* sequences were manually curated and truncated or low-quality sequences were excluded
174 from downstream analyses. Curated *CAO* sequences were aligned using MUSCLE (Edgar 2004).
175 The degron sequence (QDLLTUMILH) was identified by multiple sequence alignment of the full-
176 length CAO amino acid sequences and the regulatory N-terminal domain. In accordance with
177 Sakuraba et al. 2009, we identify the degron as "highly conserved" if 8 out of the 10 amino acids are
178 perfectly conserved, "moderately conserved" if 4 to 8 amino acids are conserved, and "poorly
179 conserved" if less than 4 amino acids are conserved. Sequence logos for each major group
180 (chlorophyte algae, charophyte algae, non-vascular plants and vascular plants) were generated by
181 Geneious Prime 2024.11 (Dotmatics), where the size of the letter reflects its frequency. To examine
182 the expression of CAO duplicate genes, we screened previously published transcriptomes from 16
183 species with representatives from the Chlorophytes (Arriola et al. 2018), Charophytes (Cheng et al.
184 2019), non-vascular (Perroud et al. 2018; Healey et al. 2023) and vascular plants (Zhu et al. 2019;
185 Wang et al. 2021; Liu et al. 2022b; Shang et al. 2023; Guo et al. 2025; He et al. 2025; Mascuñano et
186 al. 2025; Roy et al. 2025). Only transcriptomes that report FPKM, RPKM and TPM values obtained
187 from organisms cultivated at optimal conditions were considered (Supplementary Table S4).

188 ***Phylogenetic Inference***

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

197 These manually curated CAO genes were translated into amino acid sequences, aligned with
198 MUSCLE and trimmed to remove gaps and ambiguously aligned regions with Gblock (Castresana
199 2000; Talavera and Castresana 2007). Maximum likelihood trees were inferred in the CIPRES
200 Science Gateway (Miller et al. 2015) using RAxML v8.0 (Stamatakis 2014) with 1000 bootstraps with
201 the Whelan Goldman matrix for globular proteins (WAG), a gamma shape parameter, and empirical
202 estimation of invariable sites. The trees were rooted using prokaryotic CAO from *Acaryochloris*
203 *thomasi* (WP_110987895.1) and *Prochlorotrix hollandica* (WP_017713323.1) as an outgroup.
204 Tandem duplications were defined as those located on the same chromosome or contig within <10
205 genes of each other, based on the genomic coordinates retrieved from Phytozome v13 or
206 Phycocosm. To identify ancestral and lineage-specific duplications, we used Notung v2.9 (Stolzer
207 et al. 2012) and reconciled the CAO trees with a reference species tree derived from NCBI
208 taxonomy (Schoch et al. 2020) via the PhyloT v2 tool (https://phylot.biobyte.de). Putative ancestral
209 duplications were only reported when occurring on branches with bootstrap support ≥85%. All
210 trees were visualized in iTOL v6.8.2 (Letunic and Bork 2021). Previously inferred taxonomic

211 information for chlorophytes and streptophytes was confirmed via AlgaeBase, University of Galway
212 (www.algaebase.org), and for land plants via the Plants Of the World Online, Royal Botanic
213 Gardens, Kew (https://powo.science.kew.org).

214 **Results and Discussion**

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

216 Our analysis demonstrates a widespread occurrence of *CAO* gene duplication across the
217 Viridiplantae. Multiple gene copies are more prevalent among the angiosperms, where ~41% of
218 examined species encode for more than one *CAO* gene. In comparison, we detected two *CAO* gene
219 copies in only ~11% of chlorophyte and ~20% of charophyte species (Table 1; Table 2,
220 Supplementary Table S1; Supplementary Table S3). All 23 species where we detect more than two
221 *CAO* genes are members of the Embryophyta. The modern-day cultivar of sugarcane (*Saccharum*
222 sp. R570) encodes for 6 *CAO* copies, the largest number we report. This species is a result of inter-
223 specific hybridization (*Saccharum officinarum* x *spontaneum;* Dumont et al. 2022), which has
224 resulted in a very large and highly redundant genome (~0.3 Gb, ~48,000 genes) (Healey et al. 2024).
225 It must be noted that these results are likely affected by the availability of many more angiosperm
226 genomes in public databases, while other plant and algal groups are not as well represented. As
227 more genomes are sequenced, these analyses will have to be revisited.

228 We also detected multiple occurrences of heterodimeric *CAO* genes, where the functional CAO
229 protein is encoded by two genes (Table 1, Supplementary Table S1). Previous work in the
230 prasinophyte *Micromonas pusila* demonstrated that in this species CAO was encoded by two
231 separate genes, one that possesses the mononuclear iron-binding motif (*MpCAO1*) and a second
232 one that encodes for the Rieske cluster (*MpCAO2*). It was shown that a combination of both gene
233 products was necessary to form an active heterodimeric CAO protein (Kunugi et al. 2013; Dey et al.
234 2023). We expand on this work by showing the presence of a heterodimeric CAO encoded by two
235 separate genes across the Mammiellophycea. This group within the prasinophytes contains *M.*
236 *pusila* and several of its relatives (Worden et al. 2009), the ecologically important picoalgae
237 *Ostreococcus tauri*, *Ostreococcus lucimarinus* and *Bathycoccus prasinos* (Palenik et al. 2007;
238 Moreau et al. 2012; Blanc-Mathieu et al. 2014), as well as the related *Chloropicon primus*
239 (Chloropicophyceae) (Lemieux et al. 2019).

240 We also show that heterodimeric CAO is not restricted only within the Prasinophytes. We detect a
241 similar *CAO* gene arrangement in the genome of the multinucleate green alga *Caulerpa lentillifera*
242 (Bryopsidales) (Arimoto et al. 2019) , as well as in in the genome of *Bigelowiella natans*
243 (Chlorarachniophyceae) (Curtis et al. 2012). Phylogenetic analysis based on plastid-encoded
244 protein sequences revealed that chlorarachniophytes acquired their secondary plastids by the
245 uptake of a filamentous green alga in the Bryopsidales (Suzuki et al. 2016). Thus, *B. natans* likely
246 gained a heterodimeric CAO from the Bryopsidales endophyte. Considering that CAO in all other
247 green lineages (Supplementary Table 1) and within the prochlorophytes (Nagata et al. 2004) is also
248 encoded by a single gene that contains both the Rieske and iron-binding domain it is likely that the
249 occurrence of two separate *CAO* genes is not an ancestral condition. Instead, it has been
250 postulated that CAO was encoded by a single gene in the common ancestor of green plants
251 (Kunugi et al. 2016). Indeed, it has been suggested that acquisition of two separate genes encoding
252 for key functional domains may not be an evolutionarily difficult process. CAO likely functions as a

253 trimer (Kunugi et al. 2013; Liu et al. 2022a), and the electron transfer from the Rieske to the
254 mononuclear iron occurs between two neighbouring CAO proteins. A functionally similar process
255 likely occurs in the CAO enzyme encoded by two separate genes.

256 **Table 1**: Number of species across the Viridiplantae with fully sequenced and annotated genomes
257 available through Phytozome (Total), compared to the number of species that encode for multiple
258 *CAO* genes (Multiple CAO) and number of species that encode for two *CAO* genes that form
259 heterodimeric complex (Heterodimeric CAO).
260

Group	Total	Heterodimeric CAO	Multiple CAO
Chlorophytes*	70	11	8
Charophytes	10	0	2
Non-vascular plants	6	0	3
Lycophytes	3	0	2
Gymnosperms	1	0	0
Angiosperms	151	0	62

261 *The Chlorophyte group also includes *Bigelowiella natans* (Chlorarachniophyceae) that has gained
262 secondary plastids by the uptake of a green alga.

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

270 We were also not able to conclusively identify a *CAO* gene in the genomes of the endolithic coral
271 holobiont *Ostreobium quekettii* (Bryopsidales) (Iha et al. 2021), the aquaculture species
272 *Tetraselmis striata* (Chlorodendrophyceae) (Steadman Tyler et al. 2019) and the halotolerant
273 *Picocystis* sp. ML (Picocystophyceae) (Junkins et al. 2019). All identified sequences within the
274 genomes of these species had e-values higher than the cut-off (≥e-15) used in this work and high
275 similarity to pheophorbide *a* oxygenase (PAO), a Rieske-type protein involved in chlorophyll
276 degradation (Pružinská et al. 2003). The lack of *CAO* genes in these algae is surprising and could be
277 an artifact of genome assembly, as Chl *b* has been detected in all three species (Roesler et al.
278 2002; Massé et al. 2020; Conlon et al. 2024). The genome of *O. queketti* has a low BUSCO score
279 (60.7%) (Iha et al. 2021) and the BUSCO score is not reported for the draft genomes of *T. striata* and
280 *P.* sp. ML (Junkins et al. 2019; Steadman Tyler et al. 2019). The presence of CAO should be
281 experimentally examined in these species.

282 It is tempting to speculate that these species may have a different mechanism for Chl *b*
283 biosynthesis or a CAO sequence significantly different than most algae and plants. For instance,
284 *Ostreobium* is an endolithic coral holobiont adapted to extreme shading, and has lost many genes
285 associated with photoprotection and light perception due to evolution in a light-limited
286 environment (Iha et al. 2021). This alga is also the only known eukaryote that has lost the gene
287 encoding light-dependent protochlorophyllide oxidoreductase (LPOR), a key enzyme that catalyzes

288 the conversion of protochlorophyllide to chlorophyllide *a*, a necessary step towards the synthesis
289 of Chl *a* (Reinbothe et al. 2010). Instead, this alga fully depends on dark operative POR (DPOR) to
290 synthesize Chl *a*, likely as an adaptation to extreme shading (Iha et al. 2021). The mechanism of
291 chlorophyll biosynthesis has not been examined in the species in detail, but it is clear that the
292 endolithic lifestyle has resulted in a unique physiology. Further work will shed light on the role of
293 CAO in this and other shade-tolerant species.

294 We used a very conservative method to identify putative CAO sequences with both a Rieske and
295 Fe-binding domains, but this strict approach may have prevented the identification of duplicates
296 that may have diverged or become pseudogenes. Indeed, using a less conservative cutoff (≤e-5)
297 when screening the genomes of *A. thaliana* and *C. reinhardtii* (Wang et al. 2022; Craig et al. 2022)
298 reveals that *PAO* and Translocon at the inner envelope membrane of chloroplasts 55 (*TIC55*) may
299 be paralogous to *CAO* (Supplementary Table S2). *PAO* and *TIC55* encode for enzymes with a Rieske
300 and Fe-binding domains but share a very low sequence similarity with CAO and each other
301 (Supplementary Table S2, Supplementary Figure S1). Both PAO and TIC55 are involved in
302 chlorophyll breakdown and allow for the degradation of the highly phototoxic pheophorbide *a* and
303 its export from the chloroplast (recently reviewed in Kuai et al. 2018). Evolutionary analysis of the
304 genes involved in chlorophyll degradation revealed that this pathway was already present in the
305 common ancestor of land plants, suggesting that *PAO* and *TIC55* may share their origins with *CAO*
306 (Schumacher et al. 2022). This hypothesis has not yet been experimentally supported as ancestral
307 monooxygenase have been difficult to distinguish without detailed knowledge on their biochemical
308 activity and function (Schumacher et al. 2022). This possibility raises interesting questions about
309 the ancient evolutionary origins of chlorophyll metabolism within the Viridiplantae.

310 ***Was CAO gene duplication an ancestral event?***

311 To determine whether the *CAO* gene duplication was an ancestral event, we performed a
312 phylogenetic analysis of all high-quality CAO amino acid sequences (except heterodimeric CAO).
313 Phylogenetic patterns typically match previously shown relationships among major Viridiplantae
314 groups (Figure 2), although some relationships are poorly resolved (particularly among the algal
315 groups and the Rosids; Supplementary Figure 2). The presence of two *CAO* genes has been
316 previously reported in rice (*OsCAO1* and *OsCAO2*) where the two genes were positioned in tandem
317 on Chromosome 10, taken as evidence for a recent gene duplication in this species (Lee et al.
318 2005). A tandem CAO duplication, however, appears to be the exception rather than the rule. In
319 addition to rice, we only observed *CAO* duplicates located in proximity to each other in the
320 genomes of a handful of species: the green alga *Volvox reticuliferus*, the lycophyte *Diphastrium*
321 *complanatum*, and the magnoliid *Cinnamomum kanehirae* (Table 2).

322 There are several instances, mainly in very closely related species, where the duplication may have
323 originated in a recent shared ancestor. There are several possible cases in the Trebouxiophyceae
324 and Chlorophyceae class, where *CAO* duplicates are detected in closely related *Chlorella* and
325 *Scenedesmus* species (Supplementary Figure 2A). This insight must be interpreted with caution as
326 phylogenetic resolution within the chlorophytes is still weak due to limited gene and taxon
327 sampling (Li et al. 2021). Within the bryophytes, *Spaghnum magellanicum* and *Sphagnum falax*,
328 species that share a very close evolutionary relationship (Piatkowski and Shaw 2019; Bell et al.
329 2020), both have two *CAO* genes possibly arising from an ancestral duplication. Two CAO genes

are also encoded in the genome of *P. patens* (Zhang et al. 2023). This raises the possibility of a CAO duplication in a recent shared ancestor (Supplementary Figure 2B), but the bryophytes are represented by only 5 species in our analysis. Additional genomic sequencing and a deeper phylogenetic analysis will be needed to clarify the evolution of CAO within these groups.

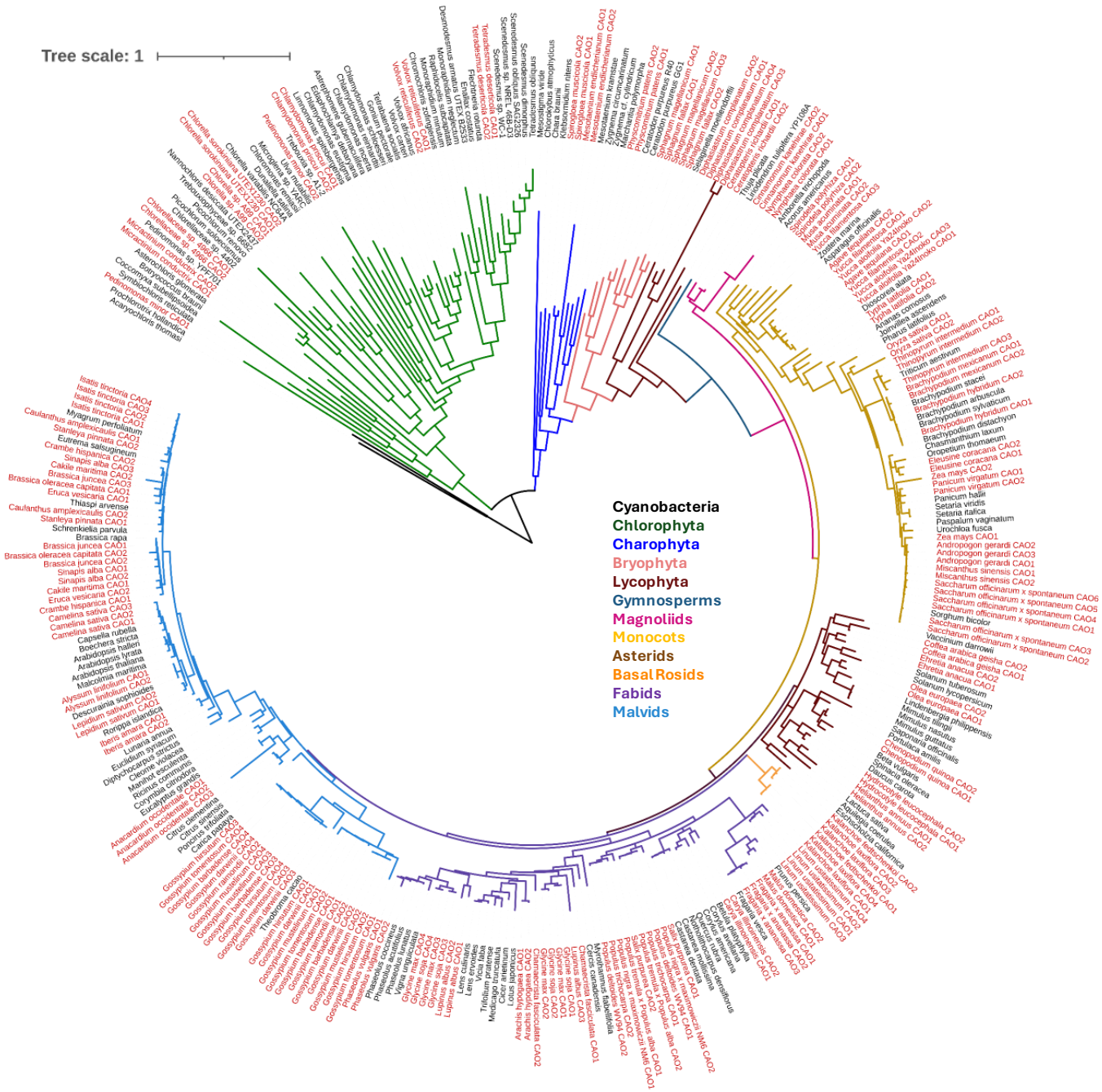


Figure 2: A phylogenetic tree based on CAO amino acid sequences inferred from using maximum likelihood analysis, using 326 unique CAO sequences across 246 species color-coded to represent the major Viridiplantae groups. Scale bar indicates amino acid substitutions per site. Species with more than one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

338 Species that encode for multiple *CAO* genes within their genomes are more prevalent among
339 angiosperms. Within the monocots, *CAO* duplicates arising from a shared ancestral gene are likely
340 present in the Asparagaceae, specifically among the closely related *Agave tequiliana*, *Yucca*
341 *filamentosa* and *Yucca aloifolia* (Ji et al. 2023) , although we were only able to detect a single *CAO*
342 copy in *Asparagus officinalis* (Supplementary Figure 2C). We detected two *CAO* gene copies in two
343 *Brachypodium* species (*B. mexicanum* and *B. hybridum*) but not in their relatives from the same
344 genus (*B. arbuscula*, *B. stacei*, *B. sylvaticum* and *B. distachyon*), raising the possibility of an
345 ancestral duplication followed by gene loss. All phylogenetic studies in *Brachypodium* support a
346 rapid and recent divergence from a common ancestor (Catalan et al. 2016), supporting this
347 possibility.

348 *CAO* duplication may be ancestral in the two closely related Saxifragales *Kalanchoe laxiflora* and
349 *Kalanchoe fedtschenkoi* (Han et al. 2024). We detected two *CAO* genes in *K. fedtschenkoi* and four
350 in *K. laxiflora*, possibly arising in a recent ancestor, followed by a second duplication within *K.*
351 *laxiflora* (Supplementary Figure 2D). Within the Fabids, four *Populus* species and their relative *Salix*
352 *purpurea*, each encode for two *CAO* gene copies that may be a result of an ancestral duplication.
353 The cultivated soybean *Glycine max* and its wild relative *Glycine soja* each encode for four *CAO*
354 gene copies (Supplementary Figure 2D) that are closely related to their homologs between the two
355 species. The phylogenetic separation of *CAO1/2* and *CAO3/4* from *G. max* and *G. soja* is puzzling,
356 as the two species are the closest relatives within the *Glycine* genus (Zhuang et al. 2022). All
357 *Gossypium* species examined here encode for four *CAO* genes, except *G. raimondii* which encodes
358 for two copies (Supplementary Figure 2E). The only other member of the Malvaceae examined,
359 *Theobroma cacao,* encodes for a single *CAO* gene, and although it clusters with *CAO* from
360 *Gossypium* sp., the relationship is poorly supported. Finally, there is a possibility of ancestral
361 duplications within the Brassicaceae (Supplementary Figure 2E), but this family has a poor
362 resolution due to factors such as incomplete sampling, rapid radiation, and hybridization (Huang et
363 al. 2016; Hendriks et al. 2023).

364 The genomes of multiple species across the Viridiplantae encode for more than one *CAO* gene, but
365 are these gene duplications related to plant evolutionary patterns? The green lineage is
366 characterized by a dynamic and complex evolutionary history marked by multiple whole genome
367 (WGD) and segmental duplication events, gene family expansions, and lineage-specific gene
368 losses (Panchy et al. 2016; Kersey 2019; Leebens-Mack et al. 2019; Soltis and Soltis 2021). Such
369 tremendous diversity in genome content can be a challenge for elucidating *CAO* evolutionary
370 trajectories even withing a single family. For instance, the Brassicaceae have been shaped by
371 multiple well-established WGD events (Mabry et al. 2020), but we detected only a single *CAO* gene
372 in the genome of *A. thaliana* and its closest relatives. In contrast, the genomes of several other
373 Brassica species encode for two or more CAO genes (Figure 2). This highlights the importance of
374 considering not only gene duplication but also gene retention dynamics when interpreting the
375 evolution of CAO. Future work that incorporates in-depth analyses, such as relative dating of CAO
376 duplication events and genome-wide synteny, will further clarify the evolutionary trajectory of CAO
377 within the Viridiplantae.

378 We show that *CAO* gene duplication is not a rarity among the Viridiplantae, but are all *CAO*
379 duplicates expressed and do they encode for active proteins? In our work, we focused only on CAO
380 genes with highly conserved catalytic domains likely to encode for active enzymes, but the

381 functional role of CAO duplication may be complex. A survey of published transcriptomes from 16
382 representative species from all major groups (Chlorophytes, Charophytes, non-vascular and
383 vascular plants) suggests that all *CAO* duplicate genes within a species are expressed, albeit at
384 different levels. Typically, one *CAO* gene has a higher expression than its paralog within the same
385 genome (Supplementary Table 4) suggesting differential regulation of *CAO* expression.

386 Significantly, all experimental studies to date suggest that the expression of *CAO* duplicates may
387 be differently regulated by light availability. In rice, the expression of *OsCAO1* is light-inducible,
388 while *OsCAO2* accumulates during periods of darkness and is expressed mainly in non-
389 photosynthetic tissues (Lee et al. 2005). In *P. patens*, both *PpCAO1* and *PpCAO2* are expressed in
390 both light and dark conditions, but only *PpCAO1* is upregulated with prolonged light exposure.
391 Similarly, in two *Chlamydomonas* species that harbour *CAO* duplicate genes (*Chlamydomonas*
392 *priscui* and *Chlamydomonas* sp. ICE-MDV), only one *CAO* homolog is upregulated in response to
393 increases in light intensity, while the second one is constitutively expressed (Poirier et al. 2025).
394 The role of light in regulating *CAO* expression has been also demonstrated previously in
395 *Arabidopsis* and tobacco, species with a single CAO gene copy (Pattanayak et al. 2005; Tanaka and
396 Tanaka 2005; Biswal et al. 2012, 2024).

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

405 It is tempting to speculate that species that encode for multiple *CAO* genes have a more robust
406 capacity for Chl *b* biosynthesis through an increase in gene dosage or a finer control over Chl *b*
407 accumulation though gene subfunctionalization or differential expression (Kondrashov 2012;
408 Panchy et al. 2016). Two gene copies with different expression patterns or activity levels could
409 provide a fine-tuned control of Chl *b* levels and photosynthetic performance under varying
410 environmental conditions, but additional experimental work is needed to substantiate this
411 hypothesis.

412

413

414

415 **Table 2:** Distribution of *CAO* duplicate genes across Viridiplantae. Only genes with a confirmed
416 Riske conserved cluster, Fe-binding domain (all species), and degron sequence (Streptophytes
417 only) are listed. A checkmark (✓) indicates that the CAO genes are positioned in a tandem on the
418 same chromosome within the genome, while a cross (x) indicates the genes are located on
419 different chromosomes. Accession numbers can be found in Supplementary Table 1 (continued on
420 next page).

Taxon	Species	#CAO genes	Tandem
Chlorophyta			
Chlorophyceae	*Chlamydomonas priscui*	2	x
Chlorophyceae	*Volvox reticuilferus*	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*	2	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*	2	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*	3	x
Brassicales	*Camelina sativa*	3	x
Brassicales	*Isatis tinctoria*	4	x

423 ***Evolutionary implications of the CAO degron sequence***

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

434 Here we show that the degron sequence is highly conserved among all land plants (86% identity),
435 moderately conserved among charophytes and non-vascular plants (68% and 66% identity
436 respectively), and very poorly conserved in chlorophytes (39% identity) (Figure 3; Supplementary
437 Figure 3). Land plants (both vascular and non-vascular) have several highly conserved sites across
438 all species (Asparagine at position 2, Isoleucine at position 6, Histidine at position 10), whereas
439 these positions are only moderately conserved in charophytes, and poorly conserved in
440 chlorophytes. In contrast, the Rieske cluster [CXH(X)15-17CXXH] exhibits an 84% pairwise identity
441 among all species, where the two histidine (positions 3 and 23) and cysteines (positions 1 and 20)
442 that coordinate the metallocentre [2Fe-2S] (Przybyla-Toscano et al. 2021) are present in all
443 examined sequences. Nearly all residues are conserved among the Viridiplantae in the C-terminal
444 Fe-binding domain (96% pairwise identity). Overall, these results suggest that the mechanism of
445 Chl *b* biosynthesis involving the Rieske and Fe-binding domains is ancestral and conserved among
446 all photosynthetic species, but this is not the case for the regulatory degron domain.

447 The lack of a highly conserved degron sequence may have several implications in the evolution of
448 efficient light harvesting. For instance, the N-regulatory domain (including the degron) is
449 completely absent in marine eukaryotic Mammiellophycea that encode for a heterodimeric *CAO*
450 gene, as well as in their deep-water relatives from the order Palmophyllales. These species
451 typically have low Chl *a/b* ratios and, unlike land plants, can incorporate Chl *b* into their core
452 antennae. This was interpreted as a necessary adaptation for efficient light harvesting in deep
453 marine environments characterized by low levels of blue-green light (Kunugi et al. 2016).

454

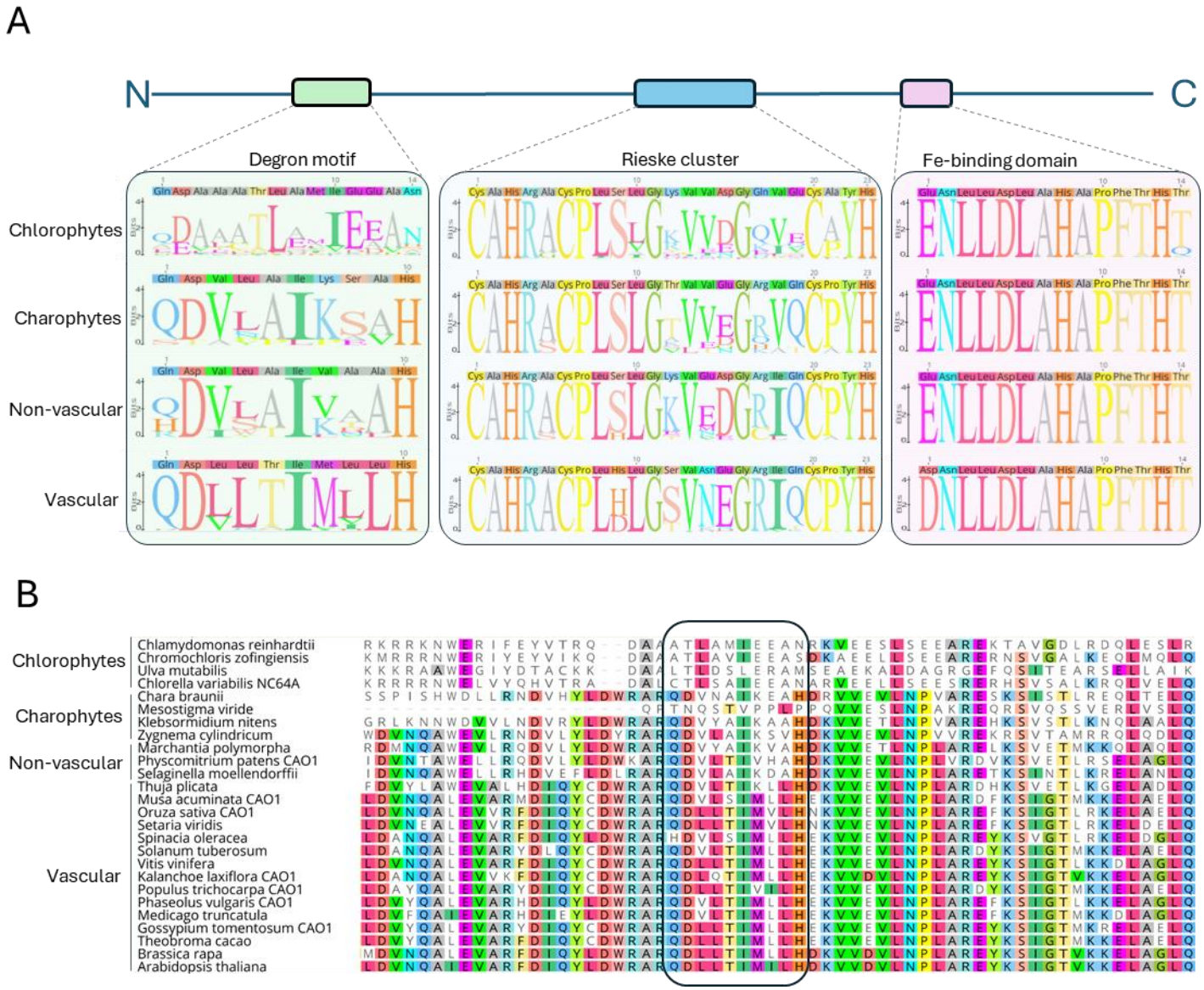


Figure 3: Key domains in the CAO protein across the Viridiplantae. (A) Sequence logos of the conserved motifs identified in the CAO protein sequences. N and C represent the N-terminus and C-terminus, respectively. All sequence logos are superimposed on a graphical representation of plant CAO sequences (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 CAO proteins from 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. (B) Comparison of the CAO amino acid sequences corresponding to the degron and its surrounding region in the N-terminal regulatory domain in representative species from Chlorophyte, Charophyte, non-vascular and vascular land plants. The region that corresponds to the degron sequence as described in *A. thaliana* is emphasized with a box. Amino acids that are identical with the *A. thalina* sequence are colored, and those that are different are shown in white.

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

479 The chlorophyll cycle has not been examined in detail in the core chlorophytes (Chlorophyceae,
480 Trebouxiophyceae, Ulvophyceae), a diverse group that inhabits marine, freshwater and terrestrial
481 habitats (Leliaert et al. 2012). Although many chlorophytes thrive in high light, others are adapted to
482 light limited environments, including deep waters. As a result, these species have developed
483 adaptations allowing for efficient light harvesting in low light. For instance, *Chlamydomonas priscui*
484 is endemic to the depths of the Antarctic Lake Bonney, an environment characterized by extreme
485 shading and dominated by blue-green wavelengths (Neale and Priscu 1995). This alga displays a
486 host of adaptations to extreme shading including a large LHC antennae size (reviewed in
487 Cvetkovska et al. 2017; Hüner et al. 2022). As suggested for marine prasinophytes (Kunugi et al.
488 2016), such low-light adapted chlorophytes could benefit from increased Chl *b* levels and could
489 thrive without a Chl *b*-dependent feedback mechanism.

490 Despite the lack of a conserved degron sequence to regulate CAO protein levels (Figure 3),
491 chlorophytes have maintained the ability to alter their Chl *a/b* ratio in response to light (Leliaert et
492 al. 2012). Transcriptional control of CAO levels was postulated to be a minor aspect of adjusting
493 Chl *b* levels in plants (Tanaka and Tanaka 2019), but this process could play a bigger role in
494 chlorophytes. It is also possible that a different, yet unidentified, mechanism exists for Chl *b*
495 turnover independent of a degron sequence, allowing these species to modify their Chl *a/b* ratios.
496 Finally, we can not rule out that the region corresponding to a degron in Chlorophyte CAO may still
497 have a similar function despite poor sequence conservation with land plants. The mechanism of
498 regulating protein degradation via degron motifs is well understood in plants (Isono et al. 2024), but
499 insights on algal degrons is lacking. A detailed experimental investigation on the control and
500 mechanism of Chl *b* biosynthesis is clearly needed among the Chlorophytes. For instance, targeted
501 mutagenesis of the N-terminus of algal CAO will reveal the importance and role of this regulatory
502 region in chlorophyll turnover.

503 ***Conclusion and Drawbacks***

504 Our work suggests that *CAO* gene duplication is widespread among the Viridiplantae, likely
505 originating from multiple independent duplication events throughout plant evolution. While we
506 show that the presence of multiple *CAO* gene copies is highly prevalent in land plants, these
507 results must be interpreted with caution. Many more angiosperm genomes are available in public
508 databases, while other plant and algal groups are not as well represented. Thus, the conclusions of
509 our study must be re-visited once more genomes are sequenced. Furthermore, many genome

510 sequencing initiatives are focused on plants of economic and agricultural importance, in which
511 domestication and selective breeding may have affected the genomic content (Turner-Hissong et
512 al. 2020). While the exact physiological impact of CAO duplication is unknown, a study in
513 transgenic tobacco overexpressing CAO (Biswal et al. 2012) showed that an increase in Chl *b* may
514 be a strategy to enhance CO_2 assimilation, delay senescence, and enhance productivity. Such
515 traits would be attractive in crop breeding programs (Voitsekhovskaja and Tyutereva 2015), and the
516 high prevalence of CAO duplication seen in crop plants in this study may have been selected for
517 during plant breeding. Indeed, several crop plants (including corn, soybean, and olives, among
518 others) appear to express all CAO genes encoded in their genome (Supplementary Table 4). The
519 functional role of these duplicates in plant physiology, photosynthetic efficiency, and productivity
520 warrants further investigation.

521 Our understanding of the prevalence and evolution of CAO at the genetic level will be improved in
522 future years, as more algal and non-vascular plant genomes become available. Furthermore, there
523 are still numerous complex and outstanding questions regarding the control of Chl *b* accumulation
524 in photosynthetic species. In particular, the chlorophytes are an understudied group when it
525 comes to Chl *b* regulating mechanism. Detailed genomic and experimental studies will shed light
526 on the role and regulation of Chl *b* in these species and will deepen our understanding on this
527 crucial photosynthetic pigment.

528 **Supplementary Information**

529 **Supplementary Table S1**: A summary of genes encoding for chlorophyllide *a* oxygenase (*CAO*)
530 across the Viridiplantae used in this work. All sequences were obtained from Phycocosm
531 (Chlorophytes, Euglenophytes, Cercozoa, Charophytes) or Phytozome v13 (Bryophytes,
532 Marchantiophytes, Lycophytes, Gymnosperms, Angiosperms). Accession numbers represent the
533 unique IDs assigned to each gene (Phycocosm: Gene portal|Protein ID|Gene Model; Phytozome:
534 Gene Model). E-values are based on a tBLASTn search with the *Chlamydomonas reinhardtii* CAO
535 (Cre01.g043350) for species within the Chlorophyta and Charophyta, and the *Arabidopsis thaliana*
536 CAO (AT1G44446) for all other species

537 **Supplementary Table S2:** BLAST results obtained by screening the genome of *Arabidopsis thaliana*
538 and *Chlamydomonas reinhardtii* for CAO homologs. Accession numbers were obtained from
539 Phytozome v13 (Goodstein et al. 2012). The functional domains according to InterPro (Blum et al.
540 2025) that are shared among all proteins are highlighted with an asterisk (Rieske Domain; InterPro:
541 IPR017941; Pheophorbide a Oxygenase Domain; InterPro: IPR013626).

542 **Supplementary Table S3**: A summary of the occurrence of *CAO* duplicate genes across the major
543 groups within Viridiplantae. The number in brackets represents the proportion of species which
544 encode for multiple copies of the *CAO* gene (Mutiple CAO) or encode for two *CAO* genes that form
545 heterodimeric complex (Heterodimeric CAO).

546 **Supplementary Table S4**: *CAO* gene expression in species whose genome encodes for more than
547 one *CAO* gene. All data was obtained from previously published work. Gene expression is
548 presented as FPKM, RPKM or TPM averages with standard deviation. In all cases, gene expression
549 ratios were determined by dividing all expression values by the highest CAO expression level within
550 the species.

551 **Supplementary Figure 1**: An alignment of the CAO protein from (A) *Arabidopsis thaliana* and (B)
552 *Chlamydomonas reinhardtii* with all other proteins identified using the CAO amino acid sequence
553 as a query. The color of the residue depends on the degree of conservation among proteins (blue
554 fully conserved; green 80–100% similar; yellow 60–80% similar; white < 60% similar). The important
555 domains that are shared among all proteins are highlighted in green (Rieske Domain; InterPro:
556 IPR017941), yellow (Fe-binding domain) and blue (Pheophorbide a Oxygenase Domain; InterPro:
557 IPR013626).

558 **Supplementary Figure 2**: A phylogenetic tree of CAO genes inferred from using maximum
559 likelihood analysis. The node labels show the bootstrap support, with dashed lines indicating <50
560 bootstrap support. Scale bar indicates amino acid substitutions per site. Species with more than
561 one detected CAO gene are highlighted in red. Cases where CAO duplication may have arisen in a
562 common shared ancestor are marked with a blue arrow, and only those located on branches with a
563 strong bootstrap support (>85%) are shown. The inset shows the outline of the full CAO
564 phylogenetic tree inferred from using maximum likelihood analysis, using 326 unique CAO
565 sequences across 246 species color coded to represent the major Viridiplantae groups. The
566 position of the chlorophytes (green) is highlighted with a black box. (**A)** Chlorophytes; (**B**)
567 Charophytes, bryophytes (including marachantiophytes), and lycophytes; (**C**) Gymnosperms,
568 magnoliids, and monocots; (**D**) Basal eudicots, asterids and fabids; (**F**) Malvids.

569 **Supplementary Figure 3**: Multiple sequence alignment of the predicted CAO amino acid
570 sequences from Chlorophytes (**A**), Charophytes (**B**), non-vascular plants (**C**) and vascular plants
571 (**D**). Global alignments were performed with MUSCLE. The colour of the residue depends on the
572 degree of conservation among species (blue, fully conserved; green, 80–100% similar; yellow, 60–
573 80% similar; gray, < 60% similar). As a reference, the CAO sequence from *Arabidopsis thaliana* is
574 present in all alignments (top), with the key domains annotated (green, Degron sequence; blue,
575 conserved Rieske cluster; purple, Fe-binding domain)

576 **References**

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

580 Arriola MB, Velmurugan N, Zhang Y, et al (2018) Genome sequences of Chlorella sorokiniana UTEX
581 1602 and Micractinium conductrix SAG 241.80: implications to maltose excretion by a
582 green alga. The Plant Journal 93:566–586. https://doi.org/10.1111/tpj.13789

583 Bell D, Lin Q, Gerelle WK, et al (2020) Organellomic data sets confirm a cryptic consensus on
584 (unrooted) land-plant relationships and provide new insights into bryophyte molecular
585 evolution. Am J Bot 107:91–115. https://doi.org/10.1002/ajb2.1397

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

589 Biswal AK, Pattanayak GK, Ruhil K, et al (2024) Reduced expression of chlorophyllide a oxygenase
590 (CAO) decreases the metabolic flux for chlorophyll synthesis and downregulates

591 photosynthesis in tobacco plants. Physiol Mol Biol Plants 30:1–16.
592 https://doi.org/10.1007/s12298-023-01395-5

593 Blanc-Mathieu R, Verhelst B, Derelle E, et al (2014) An improved genome of the model marine alga
594 Ostreococcus tauri unfolds by assessing Illumina de novo assemblies. BMC Genomics
595 15:1–12. https://doi.org/10.1186/1471-2164-15-1103

596 Blankenship RE (2010) Early Evolution of Photosynthesis1. Plant Physiol 154:434–438.
597 https://doi.org/10.1104/pp.110.161687

598 Blum M, Andreeva A, Florentino LC, et al (2025) InterPro: the protein sequence classification
599 resource in 2025. Nucleic Acids Research 53:D444–D456.
600 https://doi.org/10.1093/nar/gkae1082

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

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

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

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

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

615 Cheng S, Xian W, Fu Y, et al (2019) Genomes of Subaerial Zygnematophyceae Provide Insights into
616 Land Plant Evolution. Cell 179:1057–1067. https://doi.org/10.1016/j.cell.2019.10.019

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

620 Craig RJ, Gallaher SD, Shu S, et al (2022) The Chlamydomonas Genome Project, version 6:
621 reference assemblies for mating type plus and minus strains reveal extensive structural
622 mutation in the laboratory. The Plant Cell koac347. https://doi.org/10.1093/plcell/koac347

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

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

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

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

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

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

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

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

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

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

650 Gray J, Wardzala E, Yang M, et al (2004) A small family of LLS1-related non-heme oxygenases in
651 plants with an origin amongst oxygenic photosynthesizers. Plant Molecular Biology 54:39–
652 54

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

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

658 Guo Q, Zhu S, Lai T, et al (2025) A Phosphate-Starvation Enhanced Purple Acid Phosphatase,
659 GmPAP23 Mediates Intracellular Phosphorus Recycling and Yield in Soybean. Plant, Cell
660 and Environment 1–18. https://doi.org/10.1111/pce.15400

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

664 He Y, Deng Z, He S, et al (2025) Transcriptome and co-expression network analysis reveal the
665 genetic basis of cell wall components in maize stalks. BMC Genomics 26:.
666 https://doi.org/10.1186/s12864-025-11816-2

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

669 Healey AL, Piatkowski B, Lovell JT, et al (2023) Newly identified sex chromosomes in the Sphagnum
670 (peat moss) genome alter carbon sequestration and ecosystem dynamics. Nature Plants
671 9:238–254. https://doi.org/10.1038/s41477-022-01333-5

672 Hendriks KP, Kiefer C, Al-Shehbaz IA, et al (2023) Global Brassicaceae phylogeny based on filtering
673 of 1,000-gene dataset. Current Biology 33:4052-4068.e6.
674 https://doi.org/10.1016/j.cub.2023.08.026

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

678 Huang C-H, Sun R, Hu Y, et al (2016) Resolution of Brassicaceae Phylogeny Using Nuclear Genes
679 Uncovers Nested Radiations and Supports Convergent Morphological Evolution. Molecular
680 Biology and Evolution 33:394–412. https://doi.org/10.1093/molbev/msv226

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

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

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

689 Isono E, Li J, Pulido P, et al (2024) Protein degrons and degradation: Exploring substrate recognition
690 and pathway selection in plants. The Plant Cell 36:3074–3098.
691 https://doi.org/10.1093/plcell/koae141

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

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

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

701 Kersey PJ (2019) Plant genome sequences: past, present, future. Current Opinion in Plant Biology
702 48:1–8. https://doi.org/10.1016/j.pbi.2018.11.001

703 Kim EH, Li XP, Razeghifard R, et al (2009) The multiple roles of light-harvesting chlorophyll a/b-
704 protein complexes define structure and optimize function of Arabidopsis chloroplasts: A
705 study using two chlorophyll b-less mutants. Biochimica et Biophysica Acta - Bioenergetics
706 1787:973–984. https://doi.org/10.1016/j.bbabio.2009.04.009

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

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

713 Kuai B, Chen J, Hörtensteiner S (2018) The biochemistry and molecular biology of chlorophyll
714 breakdown. Journal of Experimental Botany 69:751–767. https://doi.org/10.1093/jxb/erx322

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

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

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

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

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

729 Lee S, Kim J-H, Yoo ES, et al (2005b) Differential regulation of chlorophyll a oxygenase genes in rice.
730 Plant Mol Biol 57:805–818. https://doi.org/10.1007/s11103-005-2066-9

731 Leebens-Mack JH, Barker MS, Carpenter EJ, et al (2019) One thousand plant transcriptomes and
732 the phylogenomics of green plants. Nature 574:679–685. https://doi.org/10.1038/s41586-
733 019-1693-2

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

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

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

741 Li X, Hou Z, Xu C, et al (2021) Large Phylogenomic Data sets Reveal Deep Relationships and Trait
742 Evolution in Chlorophyte Green Algae. Genome Biol Evol 13:evab101.
743 https://doi.org/10.1093/gbe/evab101

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

747 Liu Y, Liu J, Kong Z, et al (2022b) Transcriptomics and metabolomics analyses of the mechanism of
748 flavonoid synthesis in seeds of differently colored quinoa strains. Genomics 114:138–148.
749 https://doi.org/10.1016/j.ygeno.2021.11.030

750 Mabry ME, Brose JM, Blischak PD, et al (2020) Phylogeny and multiple independent whole-genome
751 duplication events in the Brassicales. Am J Bot 107:1148–1164.
752 https://doi.org/10.1002/ajb2.1514

753 Mascuñano B, Coto-Elena J, Guerrero-Sánchez VM, et al (2025) Transcriptome analysis of wild olive
754 (Olea Europaea L. subsp. europaea var. sylvestris) clone AC18 provides insight into the role
755 of lignin as a constitutive defense mechanism underlying resistance to Verticillium wilt.
756 BMC Plant Biology 25:. https://doi.org/10.1186/s12870-025-06301-7

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

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

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

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

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

772 Nagata N, Satoh S, Tanaka R, Tanaka A (2004) Domain structures of chlorophyllide a oxygenase of
773 green plants and Prochlorothrix hollandica in relation to catalytic functions. Planta
774 218:1019–1025. https://doi.org/10.1007/s00425-003-1181-6

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

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

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

784 Nick S, Meurer J, Soll J, Ankele E (2013) Nucleus-encoded light-harvesting chlorophyll a/b proteins
785 are imported normally into chlorophyll b-free chloroplasts of arabidopsis. Molecular Plant
786 6:860–871. https://doi.org/10.1093/mp/sss113

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

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

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

795 Panchy N, Lehti-Shiu M, Shiu S-H (2016) Evolution of Gene Duplication in Plants. Plant Physiol
796 171:2294–2316. https://doi.org/10.1104/pp.16.00523

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

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

804 Perroud PF, Haas FB, Hiss M, et al (2018) The Physcomitrella patens gene atlas project: large-scale
805 RNA-seq based expression data. Plant Journal 95:168–182.
806 https://doi.org/10.1111/tpj.13940

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

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

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

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

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

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

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

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

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

836 Roy P, Sethi S, New J, et al (2025) Transcriptomic Response of White Lupin Roots to Short-Term
837 Sucrose Treatment. Plants 14:14–17. https://doi.org/10.3390/plants14030381

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

840 Sakuraba Y, Tanaka R, Yamasato A, Tanaka A (2009) Determination of a chloroplast degron in the
841 regulatory domain of chlorophyllide a oxygenase. Journal of Biological Chemistry
842 284:36689–36699. https://doi.org/10.1074/jbc.M109.008144

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

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

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

852 Schoch CL, Ciufo S, Domrachev M, et al (2020) NCBI Taxonomy: a comprehensive update on
853 curation, resources and tools. Database (Oxford) 2020:baaa062.
854 https://doi.org/10.1093/database/baaa062

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

858 Shang S, Zhang Z, Li L, et al (2023) Transcriptome analysis reveals genes expression pattern of
859 Spirodela polyrhiza response to heat stress. International Journal of Biological
860 Macromolecules 225:767–775. https://doi.org/10.1016/j.ijbiomac.2022.11.139

861 Soltis PS, Soltis DE (2021) Plant genomes: Markers of evolutionary history and drivers of
862 evolutionary change. PLANTS, PEOPLE, PLANET 3:74–82.
863 https://doi.org/10.1002/ppp3.10159

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

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

870 Stolzer M, Lai H, Xu M, et al (2012) Inferring duplications, losses, transfers and incomplete lineage
871 sorting with nonbinary species trees. Bioinformatics 28:i409–i415.
872 https://doi.org/10.1093/bioinformatics/bts386

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

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

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

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

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

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

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

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

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

900 Tomitani A, Okada K, Miyashita H, et al (1999b) Chlorophyll b and phycobilins in the common
901 ancestor of cyanobacteria and chloroplasts. Letters to nature 400:159–162

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

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

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

910 Wang B, Yang X, Jia Y, et al (2022) High-quality Arabidopsis thaliana Genome Assembly with
911 Nanopore and HiFi Long Reads. Genomics Proteomics Bioinformatics 20:4–13.
912 https://doi.org/10.1016/j.gpb.2021.08.003

913 Wang W, Wang L, Wang L, et al (2021) Transcriptome analysis and molecular mechanism of linseed
914 (Linum usitatissimum L.) drought tolerance under repeated drought using single-molecule
915 long-read sequencing. BMC Genomics 22:1–23. https://doi.org/10.1186/s12864-021-
916 07416-5

917 Willows RD (2020) Biosynthesis of Chlorophyll and Bilins in Algae. In: Larkum AWD, Grossman AR,
918 Raven JA (eds) Advances in Photosynthesis and Respiration, 45th edn. Springer Nature, pp
919 83–103

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

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

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

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

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

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

938 Zhu X, Tai X, Ren Y, et al (2019) Genome-wide analysis of coding and long non-coding RNAs
939 involved in cuticular wax biosynthesis in cabbage (Brassica oleracea L. var. capitata).
940 International Journal of Molecular Sciences 20:. https://doi.org/10.3390/ijms20112820

941 Zhuang Y, Wang X, Li X, et al (2022) Phylogenomics of the genus Glycine sheds light on polyploid
942 evolution and life-strategy transition. Nat Plants 8:233–244.
943 https://doi.org/10.1038/s41477-022-01102-4

944

945