Streamlined regulation of chloroplast development in the liverwort *Marchantia polymorpha*

Nataliya E. Yelina[1,2,*], Eftychios Frangedakis[1,*], Tina B. Schreier[1], Jenna Rever[1], Marta Tomaselli[1], Jim Haseloff[1] and Julian M. Hibberd[1,$]

[1]Department of Plant Sciences, University of Cambridge, Cambridge, CB3 EA, UK

[2]Present address: Crop Science Centre, University of Cambridge, 93 Lawrence Weaver Road, Cambridge CB3 0LE, UK

* equal contributions

$ correspondence: Julian M. Hibberd jmh65@cam.ac.uk

Keywords: Chloroplast biogenesis, chlorophyll, Marchantia, transcription factors

Abstract

Photosynthesis in eukaryotic cells takes place in specialised plastids. The regulation of plastid development is crucial for multicellular systems such as plants. Two families of transcription factors known as Golden2-like (GLK) and GATA regulate plant chloroplast development, and the miR171-targeted SCARECROW-LIKE (SCL) GRAS transcription factors regulate chlorophyll biosynthesis. The extent to which these proteins carry out conserved roles in non-seed plants such as the liverworts is not known. Here we determine the degree of functional conservation of the GLK, GATA and SCL proteins in controlling chloroplast development in the model liverwort *Marchantia polymorpha*. Our results indicate that GATA and SCL do not play a detectable role in chloroplast biogenesis but loss of GLK function leads to reduced chloroplast size, underdeveloped thylakoid membranes and lower chlorophyll accumulation. These findings suggest that the functioning of GATA and SCL in chloroplast development either evolved after the divergence of vascular plants from bryophytes, that both roles were secondarily lost in *M. polymorpha*, or that functional redundancy is masking their roles. In contrast, and consistent with its presence in algae, GLK plays a conserved role in chloroplast biogenesis of liverworts and vascular plants.

Introduction

Photosynthesis sustains the majority of life on Earth. In eukaryotes it is performed in subcellular organelles known as chloroplasts that are thought to have originated from an endosymbiotic event between a photosynthetic prokaryote similar to cyanobacteria and a eukaryotic cell (Archibald 2009; Gould et al. 2008). Chloroplasts develop in response to light from undifferentiated proplastids in embryos, the shoot apical meristem or gametophytic cells. Key processes during chloroplast biogenesis include synthesis of chlorophyll, assembly of the thylakoid membranes and the photosynthetic apparatus and accumulation of enzymes of the Calvin Benson Bassham cycle in the chloroplast stroma (Jarvis and López-Juez 2013). Underpinning all of these events is the targeting and accumulation of transporter proteins in the chloroplast envelope because, since the endosymbiotic event, the majority of the genes controlling chloroplast biogenesis have transferred from the plastid to the nucleus and therefore need to be imported from the cytosol into the chloroplast (de Vries and Gould 2018; Gould et al. 2008). Chloroplast biogenesis therefore needs to be responsive to the external environment as well as the cell and so nuclear-encoded photosynthesis genes are regulated by light and hormones. Key intermediaries allowing responsiveness are transcription factors.

Our understanding of transcription factors acting on photosynthesis and chloroplast biogenesis is based primarily on analysis of the model flowering plant *Arabidopsis thaliana* (Figure 1A). For example, regulators that belong to the GARP and GATA families of transcription factors play key roles in chloroplast biogenesis (Cackett et al. 2022) and members of the SCARECROW-LIKE (SCL) GRAS family impact on chlorophyll accumulation (Ma et al. 2014). Within the GARP superfamily, Golden2-like (GLK) are positive regulators of nuclear-encoded chloroplast and photosynthesis related genes (Waters et al. 2008; Waters et al. 2009; Fitter et al. 2002; Bravo-Garcia et al. 2009). In the GATA family, GATA Nitrate-inducible Carbon metabolism-involved (GNC) and Cytokinin-Responsive GATA Factor 1 (CGA1) induce genes involved in chlorophyll biosynthesis and suppress phytochrome interacting factors as well as brassinosteroid (BR) related genes to promote chloroplast biogenesis and division (Chiang et al. 2012; Hudson et al. 2011; Naito et al. 2007; Bi et al. 2005). Moreover, GATA2 promotes photomorphogenesis by directly binding to light-responsive promoters (Luo et al. 2010) and in the absence of light the BR activated transcription factor BRASSINAZOLE RESISTANT1 (BZR1) represses GATA2 expression to inhibit photomorphogenesis. Lastly, three *A. thaliana* targeted SCL transcription factors redundantly regulate chlorophyll biosynthesis by suppressing the expression of *Protochlorophyllide Oxidoreductase C,* a key enzyme for chlorophyll production (Ma et al. 2014).

69 Land plants evolved from aquatic green algae, and approximately 500 MYA diverged into two
70 major monophyletic clades, the vascular plants (which include angiosperms, gymnosperms, ferns
71 and lycophytes) and the bryophytes (which include hornworts, liverworts and mosses) (Figure 1B)
72 (Li et al. 2020; de Vries and Archibald 2018). Despite our detailed knowledge of chloroplast
73 biogenesis in angiosperms, our understanding of this process in other land lineages including the
74 bryophytes is poor. With the exception of a detailed study of GLK function in the moss
75 *Physcomitrium patens* (Yasumura et al. 2005) little is known about chloroplast biogenesis in
76 bryophytes. Consequently, how this fundamental biological process may have evolved is unclear.
77 Bryophytes are of particular importance to infer more accurately the ancestral state of a trait.
78 Whilst earlier evolutionary hypotheses proposed that bryophytes represent a collection of
79 paraphyletic lineages (Qiu et al. 2006), current phylogenetic analyses indicate that the liverworts
80 and mosses form a monophyletic group that split from hornworts approximately 400 MYA (Li et
81 al. 2020; One Thousand Plant Transcriptomes Initiative 2019). This revised land plant phylogeny
82 supports the notion that traits of the common ancestor of land plants could have diversified not
83 only within the vascular plant clade but also in the three deeply divergent groups of bryophytes.
84 Thus, studying representative species from more than one bryophyte clade is useful to gain
85 insight into the likely ancestral state of any trait of interest. We therefore selected the model
86 liverwort *Marchantia polymorpha* as a system to further investigate processes underpinning the
87 evolution of chloroplast biogenesis. *M. polymorpha* has a small and well-annotated genome, key
88 steps in its development are easily accessible for observation, and an extensive set of genetic
89 manipulation tools are available (Bowman et al. 2022; Sauret-Güeto et al. 2020).
90 Phylogenetic analysis identified putative homologs for GATAs (GNC/CGA1 and GATA2), the
91 SCL-miR171 module as well as GLK and suggested the regulatory network controlling chloroplast
92 biogenesis in *A. thaliana* may be conserved in *M. polymorpha*. To test this we generated knock-
93 out mutants of each protein. Although thallus growth was perturbed by editing the *GATA* and *SCL*
94 genes, no differences in chlorophyll content were detectable. However, when *GLK* was mutated
95 chlorophyll accumulation was reduced and chloroplast development perturbed. We conclude that
96 in *M. polymorpha* the regulatory network controlling chloroplast biogenesis may be simpler than
97 in angiosperms and that of the genes tested in this study *GLK* is the only conserved component.
98 To better understand the extent to which *GLK* controls similar sets of genes in *M. polymorpha*
99 and *A. thaliana* we generated constitutive over-expressors and performed RNA sequencing on
100 over-expressing lines as well as *glk* mutant alleles. As previously reported in *A. thaliana*, the
101 abundance of transcripts derived from genes associated with chlorophyll biogenesis and
102 Photosystems I and II were impacted when *GLK* was mis-expressed in *M. polymorpha*. We

103 conclude that GLK function is conserved between *M. polymorpha* and angiosperms but that when
104 knocked out in *M. polymorpha* other regulators such as *GNC/CGA1, GATA2, SCL* have no impact
105 on chloroplast development. Thus, compared with angiosperms, the regulation of chloroplast
106 biogenesis in *M. polymorpha* is distinct and this may reflect more streamlined regulatory
107 pathways.
108
109 **Results**
110
111 ***M. polymorpha* possesses orthologs of the *GATA, SCL* and *GLK* transcription factors that**
112 **regulate chloroplast biogenesis in angiosperms**
113 We used phylogenetic analysis to search for *GATA* (*GNC*, *CGA1* and *GATA2*), *SCL* and *GLK*
114 orthologs in *M. polymorpha*. To do so, we examined twenty-one representative species from the
115 seven main groups of land plants as well as four green algae for which high quality genome
116 assemblies are available (for clarity a subset of species is shown in Figure 1B, full set of genes in
117 Supplemental Figure 1). These included three mosses, one liverwort, one hornwort, one
118 lycophyte, three ferns, two gymnosperms and five angiosperms, and in the green algae two
119 species of the Zygnematophyceae that are proposed to represent the sister group of land plants,
120 as well as one member from each of the Klebsormidiophyceae, Charophyte and Chlorophyte
121 lineages (Supplemental Figure 1).
122 *GNC*, *CGA1* and *GATA2* belong to the GATA superfamily of transcription factors that comprise
123 a family of zinc finger proteins present in all eukaryotes. Plant GATAs can be categorised into
124 four distinct classes (A, B, C and D) based on sequence conservation in the zinc finger motif,
125 additional protein domains, and intron positions (Reyes et al. 2004) (Figure 1B, C and
126 Supplemental Figure 2). *GATA2* is an A-Class GATA, which is characterized by the presence of
127 a conserved glutamine (Q) and a threonine (T) within the zinc finger motif (Figure 1C and
128 Supplemental Figure 2A). *GNC* and *CGA1* belong to the LLM subclass of B-Class GATAs due to
129 a conserved Leucine–Leucine–Methionine (LLM) domain at their C-terminus (Behringer and
130 Schwechheimer 2015) (Figure 1B, C and Supplemental Figure 2B). Our phylogenetic analysis
131 showed that *M. polymorpha* has a total of six GATA genes, among which we identified
132 *Mp7g03490* (annotated as Mp*GATA4*) and *Mp1g03950* (annotated as Mp*GATA2*) as single
133 orthologs of *GNC/CGA1* and *GATA2* respectively (Figure 1B, C, Supplemental Figure 2C and D,
134 and Supplemental Tables file 1). *SCL* is a member of the GRAS family of transcription factors
135 which have a conserved C-terminal domain (Figure 1B, C). *A. thaliana AtSCL6, AtSCL22* and
136 *AtSCL27* redundantly control chlorophyll biosynthesis and are regulated by miR171. The *M.*

137 *polymorpha* genome encodes eleven GRAS transcription factors. Genome-wide bioinformatic
138 and next generation sequencing analysis of *M. polymorpha* miRNAs and their potential targets
139 suggests that *Mp8g03980* (annotated as Mp*GRAS10*) is a target of miR171 (Lin and Bowman
140 2018). Although *Mp8g03980* is located in a sister clade to *A. thaliana* At*SCL6*, At*SCL22* and
141 At*SCL27* (Lin and Bowman 2018), based on the conservation of its miR171-mediated regulation
142 we refer to *Mp8g03980* as Mp*SCL*. Lastly, *GLK* genes belong to the GARP (GOLDEN2-LIKE,
143 ARR-B, Psr1) family of transcription factors (Safi et al. 2017) and can be divided into two main
144 subfamilies comprising the GARP-G2-like and GARP-ARR-B transcription factors. *GLK* genes
145 belong to the GARP-G2-like group and are distinguished by a characteristic C-terminal domain
146 called GOLDEN2 C-terminal (GCT) box (Figure 1C and Supplemental Figure 3A) (Fitter et al.
147 2002). Our phylogenetic analysis showed that *GLK* is present in one or two copies in most land
148 plants, but multiple copies are present in some moss and fern species. *GLK* genes were thought
149 to be land plant specific, however, we were able to detect *GLK* orthologs in two
150 Zygnematophyceae green algae for which genome assemblies became recently available (Cheng
151 et al. 2019) (two genes in *Spirogloea muscicola* [SM000086S23019 and SM000015S01186] and
152 one gene in *Mesotaenium endlicherianum* [ME000101S10682]) (Supplemental Fig3B-C).
153 Although *M. polymorpha* has eight GARP-G2 genes, we identified *Mp7g*09740 (annotated as
154 Mp*GARP8*) as a single ortholog of *GLK* (Figure 1B, C, Supplemental Figure 3A-C, and
155 Supplemental Tables file 1). In summary, our analysis indicates that *M. polymorpha* contains
156 single orthologs of *GNC/CGA1*, *GATA2*, *SCL* and *GLK* that we hereafter refer to as Mp*GATA4*,
157 Mp*GATA2*, Mp*SCL* and Mp*GLK* respectively.
158
159 **Mp*glk* mutants show compromised chlorophyll accumulation**
160 To test whether the Mp*GATA4*, Mp*GATA2*, Mp*SCL* and Mp*GLK* orthologs control chloroplast
161 biogenesis in *M. polymorpha* we used CRISPR/Cas9 editing to generate knockout mutant alleles
162 for each gene. For Mp*GATA4* two guide RNAs (gRNAs) were used to target the second and third
163 exons, for Mp*GATA2* one gRNA targeted the second exon, for Mp*SCL* two gRNAs targeted the
164 middle of its single exon, and for Mp*MIR171* two gRNAs flanking the miR171 gene were designed
165 (Figure 1C). Finally, for Mp*GLK* a single gRNA was designed to target the sixth exon (Figure 1C).
166 Plants transformed with the same vector but without a gRNA sequence were used as 'no guide
167 RNA' controls. Each mutant line was clonally propagated through one gemmae generation to
168 obtain isogenic plants, and for each targeted gene three independent lines selected for analysis.
169 Mutant alleles of Mp*GATA4* contained deletions of 93 bp (line 1 /201), 862bp (line 2 /215) and
170 248bp (line 3/ 227) (Supplemental Figure 4). For Mp*GATA2* mutant alleles with two distinct 12bp

171 base pair deletions (lines 1 and 2 /106&109) and a 16bp insertion (line 3/ 62) were isolated
172 (Supplemental Figure 4). Mutant alleles of Mp*SCL* contained 59bp or 7bp deletions that
173 introduced frameshifts resulting in premature stop codons. miRNA genes contain a ~21-bp
174 miRNA sequence and its reverse complement called a miRNA* separated by a spacer. With the
175 exception of 10 bp of the Mp*MIR171** sequence, Mp*mir171* mutants had a 139 bp deletion and 1
176 bp insertion resulting in a deletion of the entire Mp*MIR171* gene (Supplemental Figure 4). When
177 transcribed, miRNA genes produce a stem-loop or a 'hairpin' miRNA precursor which is cleaved
178 to release a mature miRNA. Deletion of the miRNA171 sequence, the spacer and ~half of the
179 miR171* sequence in our Mp*mir171* mutants, therefore, leads to a miR171 loss. Finally, for the
180 Mp*GLK* mutant allele, a frameshift and a premature stop codon are predicted to generate a
181 truncated protein of 201 amino acids (compared to the full-length 585 amino acid protein)
182 (Supplemental Figure 4). We tested whether mRNA levels from the mutated genes were affected
183 in the respective mutants but did not observe a clear correlation between mRNA levels and mutant
184 phenotypes (Supplemental Figure 5).
185 Mp*mir171* and Mp*gata2* mutants did not show any morphological or developmental
186 phenotypes and had similar chlorophyll levels and chlorophyll a/b ratios compared with wild type
187 and 'no guide RNA' controls (Figure 1D, F, H, K and L). Mp*gata4* and Mp*scl* mutants showed
188 altered thallus morphology. For example, Mp*gata4* mutants had narrower thallus lobes, while thalli
189 of Mp*scl* mutants were 'stunted' with inward curling edges (Figure 1E and G). However, in both
190 cases chlorophyll levels were not statistically different from controls with the exception of one
191 Mp*gata4* line that had an ~10% increase. (Figure 1J, M). Mp*gata4* and Mp*scl* mutant chlorophyll
192 a/b ratios were slightly reduced. In contrast, Mp*glk* mutants had an obvious pale green phenotype
193 (Figure 1I). Quantitative assays showed that chlorophyll content was ~90% lower and chlorophyll
194 a/b ratios higher than 'no guide RNA' controls (Figure 1N). Mp*glk* mutants also showed
195 morphological changes with narrower thallus lobes and upward curling lobe edges (Figure 1I).
196 In summary, absence of functional Mp*gata4* and Mp*gata2* genes did not appear to affect
197 chlorophyll biosynthesis or chloroplast biogenesis in *M. polymorpha*. As *GNC/CGA1* mutants of
198 *A. thaliana* contain lower chlorophyll levels (Zubo et al. 2018), our data argue for lack of functional
199 conservation between *A. thaliana* and *M. polymorpha* GNC/CGA1 and Mp*GATA4* in regulating
200 chloroplast biogenesis. Similarly, we found a lack of conservation between the function of
201 At*GATA2* and Mp*GATA2*. Moreover, despite conservation of a miR171 target site in the Mp*SCL*,
202 it does not appear to control chlorophyll content. Again, this contrasts with the role in *A. thaliana*
203 where triple At*scl6* At*scl22* At*scl27* mutants show increased chlorophyll levels and At*MIR171*
204 mimic lines which are functional equivalents of At*mir171* knockouts have reduced chlorophyll

205 levels (Todesco et al. 2010; Ma et al. 2014). In contrast to *M. polymorpha GNC/CGA1, GATA2*
206 and *SCL* orthologs, Mp*glk* mutants accumulated less chlorophyll indicating that the Mp*GLK*
207 ortholog is required for proper chloroplast biogenesis. This implies a conserved function for GLK
208 in chloroplast biogenesis in land plants from liverworts to angiosperms.
209
210 **An allelic series of Mp*GLK* mutants**
211 The Mp*GLK* gene has 11 exons with the conserved DNA-binding GARP domain encoded by
212 exons 6 to 8 and the GCT domain by exon 11 (Figure 2A). To investigate the importance of
213 different MpGLK domains we used CRISPR/Cas9 gene editing to mutate either the N-terminal or
214 the DNA binding domains (Figure 2A). All generated Mp*glk* mutants were pale green compared
215 with controls and had reduced chlorophyll levels (Figure 2B-F). However, they were still able to
216 grow and produce gemmae and reproductive organs (Supplemental Figure 6A, B) without the
217 need for supplemental carbon sources. Disruption of the Mp*GLK* genomic sequence between
218 exons 1 and 3 resulted in 'weak' Mp*glk* alleles (Figure 2A-C, Mp*glk* g3+7 *lines 1-3*) with chlorophyll
219 content from 72 to 55 % lower than in the controls (Figure 2F). In contrast, frameshifts in exon 6
220 (upstream of the GARP DNA-binding domain) leading to a premature stop codon (Figure 2A, B,
221 D, Mp*glk* g7 *lines 1-3*), or deletion of a gene fragment between exons 6 and 8 encoding the
222 conserved GARP DNA-binding domain (Figure 2A, E, Mp*glk* g14+g17 *lines 1-3*) resulted in
223 'strong' alleles with chlorophyll ~90% lower than in the controls (Figure 2F). We found that lower
224 chlorophyll content in *M. polymorpha* Mp*glk* mutants was associated with increased chlorophyll
225 a/b ratios (Figure 2F) consistent with previous reports for Arabidopsis At*glk1* At*glk2* mutants
226 (Waters et al. 2009). We hypothesise that in the 'weak' Mp*glk* alleles disruption of Mp*GLK*
227 genomic sequence between exons 1 and 3 allowed translation from an alternative start codon
228 resulting in an N-terminally truncated but partially functional MpGLK protein, while in the 'strong'
229 Mp*glk* alleles frameshifts in exon 6 resulting in a premature stop codon or deletion of an Mp*GLK*
230 gene portion encoding a conserved GARP DNA binding domain, led to a loss-of-function MpGLK
231 protein.
232 To better understand the effects of Mp*glk* mutations on chloroplast size and number we
233 regenerated 'strong' Mp*glk* mutants combined with a transgene to mark cell boundaries via a
234 plasma membrane-targeted eGFP fluorescent protein. These lines (Mp*glk* g7 MM *lines 1-3*)
235 contained only ~85% chlorophyll compared with wild type (Figure 3A-D and Supplemental Figure
236 6C-E). We found that *Mpglk* g7 MM mutants had smaller chloroplasts compared with the controls
237 while chloroplast number per cell was slightly increased (Figure 3E-H). We used Mp*glk* g7 (lines
238 1 and 2) as well as Mp*glk* g14+g17 (lines 1 and 2) that had a ~90% reduction in chlorophyll content

239 to analyse chloroplast ultrastructure via electron microscopy (Figure 3G, H). Consistent with our
240 confocal microscopy observations, Mp*glk* mutants had smaller chloroplasts compared with the
241 controls and showed perturbed ultrastructure. Specifically, chloroplasts had fewer thylakoid
242 membranes that often did not occupy the full chloroplast area and had reduced granal stacking.
243 Mp*glk* chloroplasts also showed visible signs of light stress such as increased numbers of
244 plastoglobules and stromal extrusions (stromules) (Figure 3G, H and Supplemental Figure 6F).
245 We tested whether these poorly developed chloroplasts in the Mp*glk* 'strong' mutants were
246 functional by treatment with di-chlorophenyl di-methyl urea (DCMU) which inhibits the
247 photosynthetic electron transport chain (Figure 3I, J) (Tresbst 2007). Although they contained
248 very little chlorophyll a substantial reduction in chlorophyll fluorescence following DCMU treatment
249 was evident in the mutants indicating that the photosynthetic apparatus was functional (Figure 3I,
250 J). This is consistent with the ability of 'strong' Mp*glk* alleles to grow under standard conditions
251 without a carbon supplement.
252 In summary, in our Mp*glk* mutant alleles chlorophyll content was reduced between ~55% and
253 ~90% compared with controls. To our knowledge, a ~90% chlorophyll loss represents the
254 strongest reported so far in a *glk* mutant in any plant species. Mutations that either caused a
255 deletion of the GARP DNA-binding domain or introduced a premature stop codon immediately
256 upstream of it resulted in a more severe phenotype than mutations in the 5´ end of the Mp*GLK*
257 gene. 'Strong' Mp*glk* mutants contained smaller chloroplasts with altered ultrastructure compared
258 with control plants but the photosynthetic apparatus was functional. These Mp*glk* mutant
259 phenotypes are consistent with the Mp*GLK* role in promoting chloroplast biogenesis and greening.
260
261 **The Mp*GLK* 3´ UTR reduces accumulation of Mp*GLK* transcripts**
262 Overexpression of *GLK* in angiosperms such as rice and *A. thaliana* can increase chlorophyll
263 content and chloroplast size (Nakamura et al. 2009, Wang et al. 2017, Zubo et al. 2018). Although
264 overexpression of rice *GLK* in rice is not maintained beyond the seedling stage (Nakamura et al.
265 2009) and overexpression of the maize *GLK* in rice can generate more sustained responses, it is
266 reduced in a sequence-specific manner throughout plant development suggesting the potential
267 for a post-transcriptional regulatory mechanism (Wang et al. 2017). Consistent with this
268 hypothesis, previous analysis identified that Mp*GLK* mRNA may be subject to cleavage by
269 miR11666.4 and another unidentified miRNA or siRNA via target sites in the 3´ untranslated
270 region (UTR) (Supplemental Figure 7 and Lin et al. 2016).
271 To investigate whether Mp*GLK* is subject to post-transcriptional regulation we generated three
272 constructs where the Mp*GLK* coding sequence (CDS) was driven by the strong promoter for a

273 constitutive ubiquitin-conjugating enzyme E2 gene (Mp*UBE2*) (Sauret-Güeto et al. 2020) (Figure
274 4A-E). In the first construct we used the native Mp*GLK* coding sequence, hereafter referred to as
275 Mp*GLK-CDS*. The second construct was designed to test whether Mp*GLK* coding sequence
276 contains motifs targeting its mRNA for post-transcriptional downregulation and so nucleotide
277 sequence was mutated but amino acid sequence preserved (Supplemental Figure 8). We
278 hypothesised that this would remove nucleotide sequence motifs targeting Mp*GLK* for post-
279 transcriptional degradation. The resulting 're-written' coding sequence was named Mp*GLK-*
280 *CDSrw*. In the third construct the native Mp*GLK* coding sequence was fused to the Mp*GLK* 3´
281 UTR sequence (Mp*GLK-CDS+3´UTR*) to test whether this would reduce greening. In all
282 constructs we included a plasma-membrane-targeted eGFP (Sauret-Güeto et al. 2021) to
283 visualise cell boundaries and quantify chloroplast numbers per cell. There were no differences in
284 chlorophyll levels or morphology between plants in which the plasma membrane was marked with
285 eGFP and those without the fluorescent marker (Supplemental Figure 9).
286 Mp*GLK-CDS* and Mp*GLK-CDSrw* overexpressing transformants were darker green compared
287 with the 'empty vector' controls and showed perturbations to morphology such as stunted growth
288 (Figure 4 C, D, G, H). qPCR analysis confirmed that the Mp*GLK-CDS* and the Mp*GLK-CDSrw*
289 transgenes were overexpressed (Figure 4J, K). Lines expressing Mp*GLK-CDS* and the Mp*GLK-*
290 *CDSrw* showed up to ~4 times higher chlorophyll levels compared with 'empty vector' controls
291 (Figure 4L). Interestingly, chlorophyll content in *MpGLK-CDSrw* was not higher than the *MpGLK-*
292 *CDS* lines (Figure 4L). Plants expressing Mp*GLK-CDS+3´UTR* showed intermediate chlorophyll
293 accumulation between 'empty vector' controls and those expressing Mp*GLK*-CDS (Figure 4L).
294 For all overexpressing lines chlorophyll a/b ratios were slightly reduced (~10%). qPCR analysis
295 confirmed that the Mp*GLK-CDS+3´UTR* transgene was over-expressed (Figure 4J). Mp*GLK*
296 mRNA levels were lower in Mp*GLK-CDS+3´UTR* compared with the *MpGLK-CDS* lines (Figure
297 4J), which is consistent with our hypothesis that Mp*GLK* is post-transcriptionally downregulated
298 via its *3´UTR.* To understand which part of Mp*GLK* 3´UTR negatively regulates greening we
299 tested a series of truncated Mp*GLK* 3´UTR versions covering positions 1-100, 1-200, 1-300, 1-
300 400, 1-500 and 1-610 as well as the full length 671 bp sequence (Figure 5A). The 1-200 bp
301 truncated version contained a predicted miR11666.4 target site located at positions 105-124 and
302 an additional putative miRNA/siRNA cleavage site located between positions 135 and 136 (Lin
303 and Bowman 2018). We then fused the truncated versions of Mp*GLK* 3´UTR to Mp*GLK*-CDS and
304 overexpressed these constructs under the control of the M*pUBE2* promoter. Only the 1-500 bp
305 and 1-610 bp fragments of the Mp*GLK* 3´UTR had the same negative effect on Mp*GLK*-mediated
306 greening as the full-length Mp*GLK* 3´UTR (Figure 5A, B). This suggests a presence of an

307 unknown regulatory motif(s) within the 400-671 bp region of Mp*GLK* 3´UTR that are responsible
308 for Mp*GLK* downregulation.
309 Finally, we analysed chloroplast morphology in gemmae of Mp*GLK* overexpression lines via
310 confocal laser scanning microscopy. A mature gemma has two peripheral meristematic regions
311 known as notches, while gemmae's central part is more mature and comprised of up to five cells
312 layers including non-photosynthetic rhizoids precursor cells. Oil body cells is another non-
313 photosynthetic cell type found on gemmae periphery (Kato et al. 2020). Confocal imaging of the
314 central part of gemmae showed that cells overexpressing Mp*GLK-CDS* and Mp*GLK-CDSrw* but
315 not Mp*GLK-CDS+3´UTR* contained more densely packed chloroplasts (Figure 5C-G and
316 Supplemental Figure 10A-D). Average chloroplast number per cell area was the same as in the
317 controls but chloroplasts were on average larger and in some case extremely large (Figure 5F, I
318 and Supplemental Figure 10E-F and Supplemental Figure 11). Interestingly, particularly in the
319 Mp*GLK-CDS* lines we observed an increase in chloroplast size in non-photosynthetic rhizoid
320 precursor and oil body cells compared with control lines (Figure 5H). In summary, we show that
321 Mp*GLK* 3´ UTR reduces greening that can be induced by Mp*GLK* overexpression in *M.*
322 *polymorpha.*
323
324 **GLK in *M. polymorpha* regulates thylakoid associated photosynthetic components and**
325 **chlorophyll biosynthesis, similarly to *A. thaliana***
326 To identify genes that respond to GLK in both *A. thaliana* and *M. polymorpha*, and to
327 understand GLK targets that may be unique to *M. polymorpha* we performed RNA sequencing in
328 four Mp*glk* mutants (two 'strong' and two 'weak' alleles) and *MpGLK* overexpression lines
329 alongside controls. We obtained ~45 million reads per accession, out of which 90-95% mapped
330 to the *M. polymorpha* transcriptome (Supplemental Tables file 2). Principal component analysis
331 of normalised counts shows that genotype background accounted for 62% of the variance
332 detected in the data (Supplemental Figure 12A). 839 differentially expressed genes were detected
333 in all four Mp*GLK* overexpressing lines compared with controls (*padj*-value =<0.01 test, LFC ≥ 1-
334 fold) of which 493 were upregulated and 346 were downregulated (Supplemental Figure 12B,
335 Supplemental Tables file 2). When only 'weak' Mp*glk* mutant alleles were assessed, there were
336 656 differentially expressed genes (*padj*-value =<0.01 test, LFC ≥ 1-fold) of which 190 and 466
337 were up- and downregulated, respectively (Figure 6C, Supplemental Tables file 2). As expected,
338 'strong' Mp*glk* mutant alleles had a greater impact on gene expression with 1536 differentially
339 expressed genes (*padj*-value =<0.01 test, LFC ≥ 1-fold), of which 471 and 1065 were up- and
340 downregulated, respectively (Supplemental Figure 12D, Supplemental Tables file 2). 32 genes

were differentially expressed in the Mp*GLK* overexpression lines as well as 'weak' and 'strong' mutant alleles (Supplemental Figure 12E). Gene Ontology analysis showed that in both overexpression lines and 'strong' Mp*glk* mutants, oxidation and reduction as well as carbohydrate metabolic processes were most impacted by mis-regulation of *GLK* (Figure 6A). When we specifically interrogated the effects of Mp*GLK* overexpression on chlorophyll and photosynthesis genes we found significant overlap with known GLK targets in *A. thaliana* (Tu et al. 2022; Waters et al. 2009). For example, in *M. polymorph*a GLK mis-expression affected nine of the 14 chlorophyll biosynthesis genes including *HEMB*, *HEMC*, *HEME*, *HEMY*, *DVR* and *CAO* (Figure 6B and C). Moreover, similarly to *A. thaliana*, changes in *M. polymorpha GLK* expression affected Photosystem II (PSII) genes including *psbP, psbR, LHCB1-3* and *LHCB*-6 (Figure 6D and E). In *A. thaliana* GLK targets include Photosystem I (PSI) light harvesting genes *LHCA1*&*4* and *psaD*; in *M. polymorpha* we identified additional GLK targets including *LHCA2&3, psaG, L, O, H, K* and *N* (Figure 6D and E). Additionally, consistent with *A. thaliana* (Tu et al. 2022) we found that genes encoding petD and petC components of the cytochrome b_6f complex were affected by GLK in *M. polymorpha* (Supplemental Figure 13). We conclude that there is significant functional overlap between photosynthesis gene types regulated by GLK in both *M. polymorpha* and *A. thaliana*. Finally, to test whether Mp*GATA4*, Mp*GATA2* and Mp*SCL* act redundantly with Mp*GLK,* we examined if their expression was affected when Mp*GLK* is misregulated. We did not find such evidence as there were no statistically significant changes in Mp*GATA4*, Mp*GATA2* and Mp*SCL* expression in either Mp*GLK* overexpression lines or Mp*glk* mutants (Supplemental Figure 14).

Discussion

Loss of GATAs and SCL alone do not impact on chloroplast biogenesis

The monophyletic group of bryophytes diverged from vascular plants approximately 400MYA. Since then, vascular plants and particularly angiosperms have undergone major morphological and physiological changes including elaborations in chloroplast biogenesis and photosynthesis. Very little is known about the evolution of the underlying genetic networks. We used *M. polymorpha* to test the extent of functional conservation for known regulators of chloroplast biogenesis compared to angiosperms and to shed light on the likely ancestral state.

Our analysis indicated that chlorophyll accumulation was not detectably perturbed in mutant alleles of Mp*gata4 or* Mp*gata2*. This contrasts with a ~30-40% decrease in chlorophyll in *A. thaliana gnc*/*cga1* mutants (Zubo et al. 2018), and reduced photomorphogenesis when At*GATA2* was suppressed via RNA silencing or artificial microRNAs (Luo et al. 2010). The simplest and

375 most parsimonious explanation for the lack of impact of Mp*gata4* or Mp*gata2* loss on chlorophyll
376 content in bryophytes such as *M. polymorpha* is that these proteins do not play a role in
377 chloroplast biogenesis, and that this represents the ancestral state. It is also possible that other
378 proteins compensate and so redundancy in the gene regulatory networks masks loss of function.
379 However, no statistically significant effects of Mp*GLK* overexpression or downregulation on
380 Mp*GATA4* and Mp*GATA2* transcript abundance most likely suggests that Mp*GLK* does not
381 control Mp*GATA4* and Mp*GATA2,* at least at the transcript level. Another possibility is that GATA
382 transcription factors regulate chloroplast biogenesis ancestrally in some but not all bryophytes.
383 This possibility is supported by some limited analysis of Pp*GATA1,* one of the two *GNC/CGA1*
384 orthologs in *P. patens*, where two Pp*GATA1* overexpression lines showed a ~10-20% increase
385 in chlorophyll content (Luan et al. 2023). A similar effect was also reported when Pp*GATA1* was
386 mis-expressed in *A. thaliana* (Luan et al. 2023). However, these increases in chlorophyll content
387 induced by Pp*GATA1* are modest compared with those reported in *A. thaliana* where *GNC* over-
388 expression increased chlorophyll tenfold in seedlings and ~30% in the leaf (Chiang et al. 2012).
389 Moreover, while *GNC/CGA1* overexpression in *A. thaliana* led to an increase in chloroplast
390 number per cell, this was not reported in *P. patens* (Luan et al. 2023). The role of *PpGATA*
391 orthologs (including both Pp*GATA1* and Pp*GATA2*), therefore, needs additional analysis to
392 confirm or refute their role in chloroplast biogenesis. We also note that editing Mp*GATA4* had an
393 effect on gametophyte morphology such that thalli of Mp*gata4* mutants had narrower lobes. This
394 contrasts with *A. thaliana* where to our knowledge there have been no reports of perturbations to
395 leaf morphology in *gnc*/*cga1* mutants. It is therefore possible that the role of this protein has been
396 repurposed from one in affecting development of photosynthetic tissue in *M. polymorpha*, to one
397 in regulating chloroplast biogenesis in *A. thaliana*.
398 Similarly, neither Mp*scl* or Mp*mir171* mutants exhibited any detectable alterations in
399 chlorophyll accumulation as would be expected if their functions were conserved between
400 bryophytes and *A. thaliana.* For example, At*scl6* At*scl22* At*scl27* triple mutants and *MIR171* over-
401 expressors lead to increased chlorophyll accumulation in *A. thaliana* (Wang et al. 2010). As no
402 other members of the *SCL* gene family in *M. polymorpha* contain a MIR171 recognition site, this
403 argues against Mp*MIR171* playing a role in the regulation of chlorophyll accumulation in this
404 liverwort. It is possible that the *M. polymorpha* ortholog of At*SCL* does regulate chlorophyll
405 biosynthesis but without the additional control derived from MIR171 that operates in *A. thaliana*.
406 SCL6, SCL22 and SCL27 also regulate proliferation of meristematic cells (Bolle 2004; Llave et al.
407 2002; Rhoades et al. 2002; Schulze et al. 2010). A similar role in *M. polymorpha* could explain
408 the perturbed thallus morphology with inward curling edges of Mp*scl* mutants. Thus, in addition

409 to GATA2 and GATA4 discussion above, it is also possible that an ancestral role of SCL relates
410 to development of photosynthetic tissue. For SCL6, SCL22 and SCL27 this function seems to be
411 retained in *A. thaliana*, and a role in repressing chlorophyll synthesis has been acquired.
412
413 **Loss of GLK limits chloroplast biogenesis in *M. polymorpha***
414 Unlike *GATAs* and the *MIR171-SCL* module, three lines of evidence indicate that GLK function
415 is conserved in *M. polymorpha*. Firstly, Mp*glk* mutants have significantly reduced chlorophyll
416 accumulation and smaller chloroplasts with underdeveloped thylakoids. These perturbations to
417 phenotype have been observed in other land plants (Bravo et al. 2009; Hall et al. 1998; Rossini
418 et al. 2001; Fitter et al. 2002; Yasumura et al. 2005; Waters et al. 2009). Secondly, constitutive
419 overexpression of Mp*GLK* resulted in increased chlorophyll accumulation, increased chloroplast
420 size and ectopic chloroplast development similar to reports in angiosperms (Kobayashi et al.
421 2013; Nakamura et al. 2009, Powell et al. 2012; Wang et al. 2017, Zubo et al. 2018). Thirdly, our
422 analysis of Mp*GLK* overexpression and mutant RNA transcriptomes revealed an overlap between
423 *GLK* targets in *M. polymorpha* and *A. thaliana* but also rice, tobacco, tomato and maize (Waters
424 et al. 2009; Tu et al. 2022). Most of these shared GLK targets are photosynthesis-related genes
425 such as the *LHCA*, *LHCB*, *PsbQ* and genes encoding chlorophyll biosynthesis enzymes (Tu et al.
426 2022). Finally, and in contrast to previous reports, our phylogenetic analysis identified *GLK*
427 orthologs in two Zygnematophyceae green algae for which genome assemblies became recently
428 available (Cheng et al. 2019). Taken together our results suggest that the GLK function is
429 ancestral to all land plants and the protein appears to have evolved before the transition of plants
430 to terrestrial ecosystems.
431 We also examined how Mp*GLK* is regulated at transcriptional and post-transcriptional levels.
432 First, we asked whether there are any motifs within Mp*GLK* CDS that could cause its
433 downregulation. To do so, we ‘re-wrote’ the coding sequence changing its nucleotide sequence
434 while preserving amino acids. The ‘re-written’ Mp*GLK* did not increase chlorophyll content over
435 and above that in plants over-expressing the native Mp*GLK* CDS. We believe that there are two
436 possible explanations for why Mp*GLK-CDSrw* did not enhance greening more strongly than the
437 native Mp*GLK*-CDS. Firstly, the Mp*GLK* coding sequence does not contain nucleotide sequence
438 motifs responsible for post-transcriptional regulation. Secondly, although we codon-optimised
439 Mp*GLK-CDSrw* for *M. polymorpha*, its expression may be less efficient than that of the native
440 Mp*GLK-CDS*. Finally, we found that although Mp*GLK* is negatively regulated by its 3´ UTR it is
441 unlikely to be regulated by a predicted miRNA/siRNA cleavage site, but rather some as yet
442 unknown motif located between 400 and 671 of Mp*GLK* 3´ UTR.

443 In summary, our results suggest that the regulation of photosynthesis gene expression is more
444 streamlined in *M. polymorpha*. We were unable to detect any alterations to chlorophyll content
445 when single copy members of three transcription factor families were mutated. It is possible that
446 the low levels genetic redundancy in regulators of photosynthesis in this species are associated
447 with low rates of photosynthesis and limited specialisation within the thallus. Angiosperms on the
448 other hand have a more complex development and morphology allowing colonisation of a broad
449 range of environments. For example, leaves with specialised tissues allow high rates of
450 photosynthesis. As a result, greater control over photosynthesis may have become necessary
451 compared with bryophytes, and could have been mediated by elaborations and specialisations to
452 pre-existing pathways present in the common ancestor of bryophytes and vascular plants. The
453 simplified genetic network in *M. polymorpha* either represents the ancestral state, or a simplified
454 version due to secondary loss. In *M. polymorpha s*uch secondary loss of traits has been reported
455 for stomata and their regulatory genetic network (Brogan et al. 2020; Rich and Delaux, 2020). To
456 elucidate the ancestral function of GLK further detailed genetic studies in other non-seed plants
457 will be necessary. Furthermore, our finding of GLK homologs in the green algal lineage
458 Zygnematophyceae, provides an exciting opportunity to test whether its role in photosynthesis
459 predates the colonisation of land.
460
461 **Materials & Methods**
462
463 **Phylogenetic analysis**
464 To identify GATA B-Class and A-Class genes, three different approaches were combined:
465 Firstly, the GATA protein sequences for 21 plant genomes were mined from the iTAK (Zheng et
466 al. 2016) and PlantTFDB databases (Jin et al. 2017), Phytozome, Fernbase, Phycozome and
467 PhytoPlaza. Sequences for each individual species were aligned with the AtGNC and AtCGA1
468 amino acid sequences using MAFFT (Katoh et al. 2005). Results were filtered manually to identify
469 GNC/CGA1 (B-Class) orthologs distinguished from other GATA family genes by the presence of
470 conserved serine (S) residue, a conserved IRX(R/K)K motif (I: Isoleucine, R: Arginine, X: any
471 amino acid and K: Lysine), and the presence or absence of conserved LLM- (leucine– leucine–
472 methionine) domain at their C-terminus (Behringer et al. 2014). GATA2 (A-Class) orthologs, were
473 distinguished by the presence of a conserved glutamine (Q) and a threonine (T) within the zinc
474 finger motif. Secondly, we performed BLASTP searches against plant genomes on Phytozome
475 v13, fern genomes (fernbase.org), hornworts genome (hwww.hornworts.uzh.ch), green algae
476 genomes on PhycoCosm (/phycocosm.jgi.doe.gov), and 1KP using the AtGNC/CGA1 amino acid

477 sequence as a query. Results were filtered manually as described above. Finally, the combined
478 results from the above to approaches were checked against Orthofinder searches (Emms and
479 Kelly, 2019). The identified GATA protein sequences were aligned using MAFFT. Alignments
480 were then trimmed using TrimAl (Capella-Gutiérrez et al., 2009). A maximum likelihood
481 phylogenetic tree was inferred using iQTree, ModelFinder (Kalyaanamoorthy et a 2017) and
482 ultrafast approximation for phylogenetic bootstrap (Hoang et al. 2018) and SH-aLRT test
483 (Guindon et al. 2010). The tree was visualised using iTOL.
484 To identify *GLK* genes, three different approaches were combined: Firstly, the G2-GARP
485 protein sequences for 21 plant genomes were mined from the iTAK (Zheng et al. 2016) and
486 PlantTFDB databases (Jin et al. 2017), Phytozome, Fernbase (Li et al. 2018), Phycozome and
487 PhytoPlaza. Sequences for each individual species were aligned with the AtGLK1/2 amino acid
488 sequences using MAFFT (Katoh et al. 2019). Results were filtered manually to identify GLK
489 orthologs distinguished from other G2-GARP family genes by three characteristic motifs:
490 AREAEAA motif (consensus motif) in the DNA-binding domain, VWG(Y/H)P and the
491 PLGL(R/K)(P/S)P in the GCT-box domain. Secondly, we performed BLASTP searches against
492 plant genomes on Phytozome v13, fern genomes (fernbase.org), hornworts genome
493 (hwww.hornworts.uzh.ch), green algae genomes on PhycoCosm (/phycocosm.jgi.doe.gov), and
494 1KP (One Thousand Plant Transcriptomes Initiative 2019) using the AtGLK1/2 amino acid
495 sequence as a query. Results were filtered manually as described above. Finally, the combined
496 results from the above two approaches were checked against Orthofinder searches (Emms and
497 Kelly 2019). The identified GLK protein sequences were aligned using MAFFT. Alignments were
498 then trimmed using TrimAl (Capella-Gutiérrez et al. 2009). A maximum likelihood phylogenetic
499 tree was inferred using iQTree (Nguyen et al. 2015), ModelFinder (Kalyaanamoorthy et al. 2017)
500 and ultrafast approximation for phylogenetic bootstrap (Hoang et al. 2018) and SH-aLRT test
501 (Guindon et al. 2010). The tree was visualised using iTOL (Letunic and Bork 2021) .
502
503 **Plant growth, transformation**, **CRISPR/Cas9 gene editing and overexpression construct**
504 **generation**
505 *Marchantia polymorpha* accessions Cam-1 (male) and Cam-2 (female) were used in this study
506 (Delmans et al. 2017). Plants were grown on half strength Gamborg B5 medium plus vitamins
507 (Duchefa Biochemie G0210, pH 5.8) and 1.2% (w/v) agar (Melford capsules, A20021), under
508 continuous light at 22 °C with light intensity of 100 µmol m^{-2} s^{-1}. Transgenic *M. polymorpha* plants
509 were obtained following an established protocol (Sauret-Güeto et al. 2020).

510 For CRISPR/Cas9 gene editing, guide RNAs were *in silico* predicted using CasFinder tool
511 (https://marchantia.info/tools/casfinder/). Several gRNAs per target gene were *in vitro* tested as
512 described (Yelina et al. 2022) using oligonucleotides in Supplemental Tables File 3. gRNA
513 sequences that were selected to generate Mp*glk*, Mp*gata2*, Mp*gata4*, Mp*scl* and Mp*mir171*
514 mutants are listed in Supplemental Tables File 3. Single gRNA7 to mutate Mp*GLK* was cloned
515 using OpenPlant toolkit (Sauret-Güeto et al. 2020). Multiple gRNAs to mutate Mp*GATA4*, Mp*SCL*
516 and Mp*MIR171* and a single gRNA to mutate Mp*GATA2* were cloned as described in Yelina et
517 al. 2022, using oligonucleotides listed in Supplemental Tables File 3 and the destination vector
518 pMpGE013 (Sugano et al. 2018). For the overexpression constructs, Mp*GLK and* Mp*GLKrw* CDS
519 were synthesised (Integrated DNA Technologies), Mp*GLK* 3´UTR was amplified from *M.*
520 *polymorpha* genomic DNA and cloned into the pUAP4 vector (Sauret-Güeto et al. 2020).
521 Constructs were generated using the OpenPlant toolkit (Sauret-Güeto et al. 2020). OpenPlant
522 parts used: OP-023 CDS12-eGFP, OP-020 CDS_hph, OP-037 CTAG_Lti6b, OP-054 3TER,
523 _Nos-35S, OP-049 PROM_35S, OP-47 PROM_MpUBE2, OP-48 5UTR_MpUBE2, OP-063,
524 L1_HyR-Ck1, OP073 L1_Cas9-Ck4, OP-076 L2_lacZgRNA-Cas9-CsA and L1_35S_s::eGFP-
525 Lti6b (Frangedakis et al. 2021). Nucleotide sequence of Mp*GLKrw* CDS and oligonucleotide
526 sequences used for cloning are listed in Supplemental Tables File 3.
527
528 **Chlorophyll determination, fluorescence measurements and imaging analysis**
529 For chlorophyll measurements, ~30-50mg of 10-14 days old gemmalings were used, with five
530 biological replicates per genotype. The tissue was blotted on tissue paper before weighing to
531 remove excess water and then was transferred into a 1.5mL Eppendorf tube containing 1 mL of
532 dimethyl sulfoxide (DMSO) (D8418, Sigma Aldrich) and incubated in the dark at 65C with 300 rpm
533 shaking for 45 minutes. Samples were let to cool down to room temperature for approximately
534 one hour. Chlorophyll content was measured using a NanoDrop™One/One C Microvolume UV-
535 Vis Spectrophotometer (ThermoFisher) following the manufacturer's protocol. Chlorophyll
536 fluorescence measurements were carried out using a CF imager (Technologica Ltd, UK) and the
537 image processing software provided by the manufacturer, as described previously (Schreier et al.
538 2022). *M. polymorpha* plants were placed in the dark for 20 min for dark adaptation to evaluate
539 the dark-adapted minimum fluorescence (Fo), dark-adapted maximum fluorescence (*Fm*),
540 variable fluorescence *Fv* (*Fv* = *Fm*–*Fo*). All chlorophyll fluorescence images within each
541 experiment were acquired at the same time in a single image, measuring a total of three plants
542 per genotype and treatment. For the DCMU treatment, 20 µM DCMU (#45463, Sigma Aldrich)
543 was added to the half-strength MS media before it was poured into the individual petri dishes.

544 Thalli were placed for 24 h onto the DCMU-containing media before chlorophyll fluorescence
545 measurements.
546 For imaging a gene frame (#AB0576, ThermoFisher) was positioned on a glass slide. Five to
547 seven gemma were placed within the medium-filled gene frame together with 30 µL of milliQ
548 water. The frame was then sealed with a cover slip. Plants were imaged immediately using a
549 Leica SP8X spectral fluorescent confocal microscope. Imaging was conducted using either a 10×
550 air objective (HC PL APO 10×/0.40 CS2) or a 20× air objective (HC PL APO 20×/0.75 CS2).
551 Excitation laser wavelength and captured emitted fluorescence wavelength window were as
552 follows: for eGFP (488 nm, 498−516 nm) and for chlorophyll autofluorescence (488 or 515nm,
553 670−700 nm). For electron microscopy sections (~2 mm^2) of 5-6 individual 3-week-old *M.*
554 *polymorpha* thallus per genotype were harvested; and then fixed, embedded and imaged using
555 scanning electron microscopy as previously described (Schreier et al. 2022).
556

557 **RNA extraction**, **cDNA preparation, qPCR** and **RNA sequencing**

558 RNA was extracted from 3-4 two-week old gemmae, using the RNeasy Plant kit (#74903,
559 Qiagen) according to the manufacturer's protocol (RLT buffer supplemented with beta-
560 merchaptoethanol was used) and residual genomic DNA was removed using the Turbo DNA-free
561 kit (# AM1907, Invitrogen) according to the manufacturer's instructions.
562 500 ng of DNase-treated RNA was used as a template for cDNA preparation using the
563 SuperScript™ IV First-Strand Synthesis System (#18091050, Invitrogen) according to
564 manufacturer's instructions (with only modifying reverse transcriptase reaction time to 40 minutes
565 and using oligo (dT)18 primers). qPCR was performed using the SYBR Green JumpStart *Taq*
566 Ready Mix (#S4438, Sigma Aldrich) and a CFX384 RT System (Bio-Rad) thermal cycler. cDNA
567 was diluted 6 times, oligonucleotides listed in Supplemental Tables file 3 were used at a final
568 concentration of 0.5 µM and reaction conditions were as follows: initial denaturation step of 94°C
569 for 2 minutes followed by 40 cycles of 94°C for 15 seconds (denaturation) and 60°C for 1 minute
570 (annealing, extension, and fluorescence reading).
571 Library preparation and RNA sequencing was performed by Novogene (Cambridge, UK).
572 Briefly, messenger RNA was purified from total RNA using poly-T oligo-attached magnetic beads.
573 After fragmentation, the first strand cDNA was synthesized using random hexamer primers
574 followed by the second strand cDNA synthesis. cDNA end repair, A-tailing, adapter ligation, size
575 selection, amplification, and purification were performed next. Library concentration was
576 measured on a Qubit instrument following the manufacturer's procedure (Thermo Fisher

577 Scientific) followed by real-time qPCR quantification. Library size distribution was analysed on a
578 bioanalyzer (Agilent) following the manufacturer's protocol. Quantified libraries were pooled and
579 sequenced on a NovaSeq PE150 Illumina platform and 6 G raw data per sample were obtained.
580 Adapter sequences were: 5' Adapter: 5'-
581 AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-3'.
582 3' Adapter: 5'-
583 GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGATGACTATCTCGTATGCCGTCTTCTG
584 CTTG-3'
585 FastQC was used to assess read quality and TrimGalore
586 (https://doi.org/10.5281/zenodo.5127899) to trim low-quality reads and remaining sequencing
587 adapters. Reads were pseudo-aligned using Kallisto (Bray et al. 2016) to the *M. polymorpha*
588 Genome version 5 (primary transcripts only, obtained from MarpolBase) (Bowman et al. 2018).
589 Kallisto estimates abundance of each transcript in units of transcripts per million (TPM). Mapping
590 statistics for each library are provided in Supplemental Tables file 2. DGE analysis was performed
591 with DESeq2 (Love et al. 2014), with padj-values < 0.01, Supplemental Tables file 2. Plots were
592 generated using R.
593
594 **Author contributions**
595 N.Y.E., E.F., T.S., M.T. and J.R. carried out the work. N.Y.E., E.F. and J.M.H. designed the work.
596 N.Y.E, E.F. and J.M.H. wrote the manuscript. J.M.H., N.Y.E. and J.M.H. initiated and oversaw the
597 project.
598
599 **Acknowledgements**
600 This work was funded as part of the BBSRC/EPSRC OpenPlant Synthetic Biology Research
601 Centre Grant BB/ L014130/1 to J.H., BBSRC BB/F011458/1 for confocal microscopy to J.H and
602 BBP0031171 to J.M.H. T.B.S. was supported by the SNSF Postdoc Mobility Fellowship
603 (P500PB_203128) and the EMBO Long-Term Fellowship (ALTF 531-2019). For the purpose of
604 open access, the authors have applied a Creative Commons Attribution (CC BY) license to any
605 Author Accepted Manuscript version arising from this submission. We thank Karin H. Müller and
606 Filomena Gallo from the Cambridge Advanced Imaging Centre for the electron microscopy
607 sample preparation as well as the support during the image acquisition.
608
609
610

611 **REFERENCES**

612

613 Archibald, John M. 2009. "The Puzzle of Plastid Evolution." *Current Biology*. 27;19(2):R81-8.

614 Behringer, Carina, and Claus Schwechheimer. 2015. "B-GATA Transcription Factors – Insights
615 into Their Structure, Regulation, and Role in Plant Development." *Frontiers in Plant Science*.
616 https://doi.org/10.3389/fpls.2015.00090.

617 Bi, Yong-Mei, Yu Zhang, Tara Signorelli, Rong Zhao, Tong Zhu, and Steven Rothstein. 2005.
618 "Genetic Analysis of Arabidopsis GATA Transcription Factor Gene Family Reveals a Nitrate-
619 Inducible Member Important for Chlorophyll Synthesis and Glucose Sensitivity." *The Plant*
620 *Journal*.

621 Bolle, Cordelia. 2004. "The Role of GRAS Proteins in Plant Signal Transduction and
622 Development." *Planta* 218 (5): 683–92.

623 Bowman, John L., Mario Arteaga-Vazquez, Frederic Berger, Liam N. Briginshaw, Philip Carella,
624 Adolfo Aguilar-Cruz, Kevin M. Davies, et al. 2022. "The Renaissance and Enlightenment of
625 Marchantia as a Model System." *The Plant Cell* 34 (10): 3512–42.

626 Bravo-Garcia, Armando, Yuki Yasumura, and Jane A. Langdale. 2009. "Specialization of the
627 Golden2-like Regulatory Pathway during Land Plant Evolution." *The New Phytologist* 183 (1):
628 133–41.

629 Bray, Nicolas L., Harold Pimentel, Páll Melsted, and Lior Pachter. 2016. "Near-Optimal
630 Probabilistic RNA-Seq Quantification." *Nature Biotechnology* 34, 525–527.

631 Cackett, Lee, Leonie H. Luginbuehl, Tina B. Schreier, Enrique Lopez-Juez, and Julian M. Hibberd.
632 2022. "Chloroplast Development in Green Plant Tissues: The Interplay between Light,
633 Hormone, and Transcriptional Regulation." *New Phytologist*. 233 (5), 2000-2016.

634 Capella-Gutiérrez, Salvador, José M. Silla-Martínez, and Toni Gabaldón. 2009. "trimAl: A Tool for
635 Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." *Bioinformatics* 25
636 (15): 1972–73.

637 Cheng, Shifeng, Wenfei Xian, Yuan Fu, Birger Marin, Jean Keller, Tian Wu, Wenjing Sun, et al.
638 2019. "Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant
639 Evolution." *Cell* 179 (5): 1057–67.e14.

640 Chiang, Yi-Hsuan, Yan O. Zubo, Wiebke Tapken, Hyo Jung Kim, Ann M. Lavanway, Louisa
641 Howard, Marinus Pilon, Joseph J. Kieber, and G. Eric Schaller. 2012. "Functional
642 Characterization of the GATA Transcription Factors GNC and CGA1 Reveals Their Key Role

643 in Chloroplast Development, Growth, and Division in Arabidopsis." *Plant Physiology* 160 (1):
644 332–48.
645 Delmans, Mihails, Bernardo Pollak, and Jim Haseloff. 2017. "MarpoDB: An Open Registry for
646 Marchantia Polymorpha Genetic Parts." *Plant & Cell Physiology* 58 (1): e5.
647 Emms, David M., and Steven Kelly. 2019. "OrthoFinder: Phylogenetic Orthology Inference for
648 Comparative Genomics." *Genome Biology* 20 (1): 238.
649 Fitter, David W., David J. Martin, Martin J. Copley, Robert W. Scotland, and Jane A. Langdale.
650 2002. "*GLK* gene Pairs Regulate Chloroplast Development in Diverse Plant Species." *The*
651 *Plant Journal*. **31(6):713-27**
652 Frangedakis, Eftychios, Manuel Waller, Tomoaki Nishiyama, Hirokazu Tsukaya, Xia Xu, Yuling Yue,
653 Michelle Tjahjadi, et al. 2021. "An Agrobacterium-Mediated Stable Transformation Technique
654 for the Hornwort Model Anthoceros Agrestis." *The New Phytologist* 232 (3): 1488–1505.
655 Gould, Sven B., Ross F. Waller, and Geoffrey I. McFadden. 2008. "Plastid Evolution." *Annual*
656 *Review of Plant Biology* 59: 491–517.
657 Guindon, Stéphane, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and
658 Olivier Gascuel. 2010. "New Algorithms and Methods to Estimate Maximum-Likelihood
659 Phylogenies: Assessing the Performance of PhyML 3.0." *Systematic Biology* 59 (3): 307–21.
660 Hall, L. N., L. Rossini, L. Cribb, and J. A. Langdale. 1998. "GOLDEN 2: A Novel Transcriptional
661 Regulator of Cellular Differentiation in the Maize Leaf." *The Plant Cell* 10 (6): 925–36.
662 Harris, Brogan J., C. Jill Harrison, Alistair M. Hetherington, and Tom A. Williams. 2020.
663 "Phylogenomic Evidence for the Monophyly of Bryophytes and the Reductive Evolution of
664 Stomata." *Current Biology: CB* 30 (11): 2001–12.e2.
665 Hoang, Diep Thi, Olga Chernomor, Arndt von Haeseler, Bui Quang Minh, and Le Sy Vinh. 2018.
666 "UFBoot2: Improving the Ultrafast Bootstrap Approximation." *Molecular Biology and*
667 *Evolution* 35 (2): 518–22.
668 Hudson, Darryl, David Guevara, Mahmoud W. Yaish, Carol Hannam, Nykoll Long, Joseph D.
669 Clarke, Yong-Mei Bi, and Steven J. Rothstein. 2011. "GNC and CGA1 Modulate Chlorophyll
670 Biosynthesis and Glutamate Synthase (GLU1/Fd-GOGAT) Expression in Arabidopsis." *PloS*
671 *One* 6 (11): e26765.
672 Jarvis, Paul, and Enrique López-Juez. 2013. "Biogenesis and Homeostasis of Chloroplasts and
673 Other Plastids." *Nature Reviews. Molecular Cell Biology* 14 (12): 787–802.

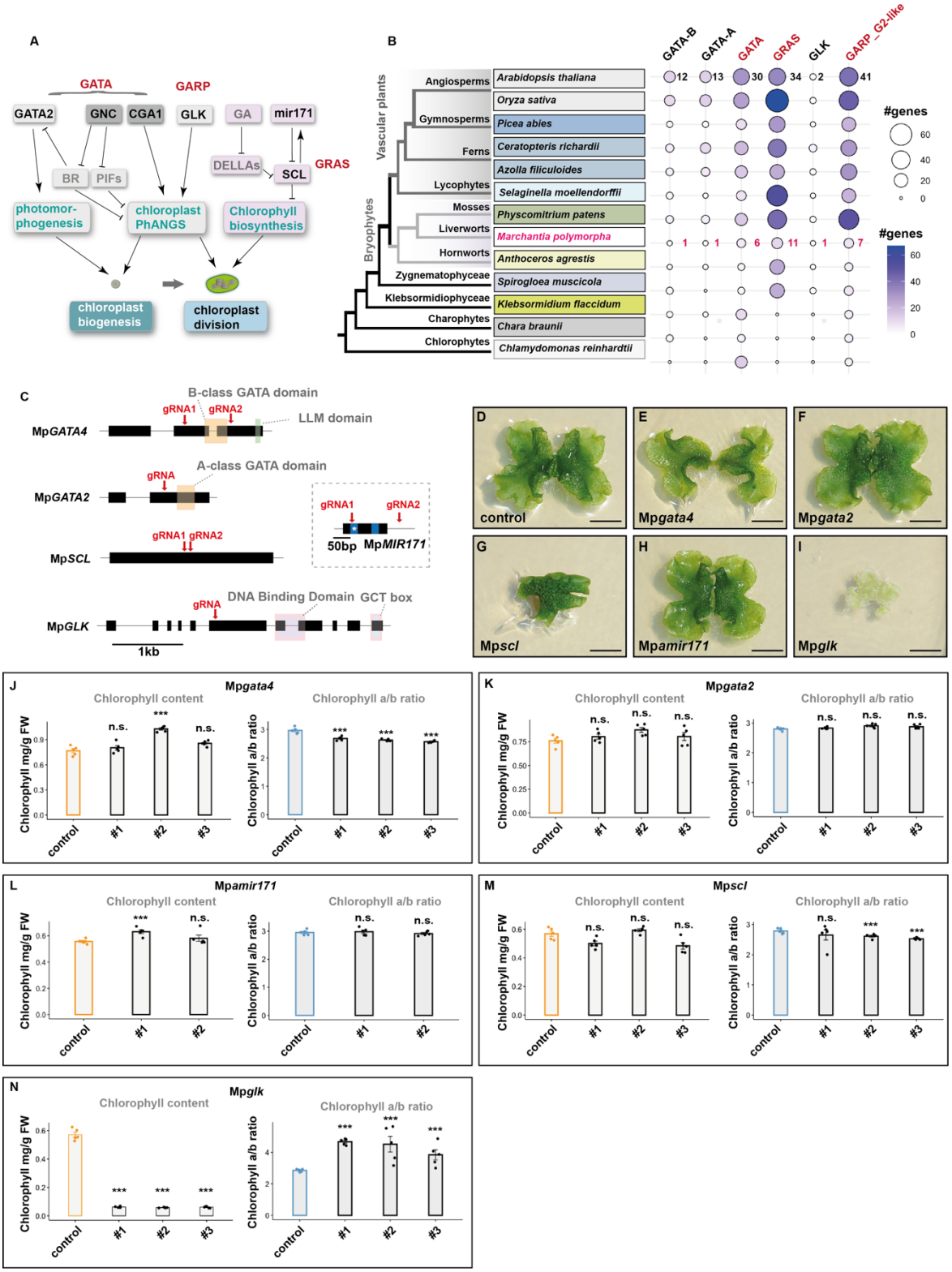
674 Jin, Jinpu, Feng Tian, De-Chang Yang, Yu-Qi Meng, Lei Kong, Jingchu Luo, and Ge Gao. 2017.
675 "PlantTFDB 4.0: Toward a Central Hub for Transcription Factors and Regulatory Interactions
676 in Plants." *Nucleic Acids Research* 45 (D1): D1040–45.
677 Kalyaanamoorthy, Subha, Bui Quang Minh, Thomas K. F. Wong, Arndt von Haeseler, and Lars S.
678 Jermiin. 2017. "ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates."
679 *Nature Methods* 14 (6): 587–89.
680 Katoh, Kazutaka, John Rozewicki, and Kazunori D. Yamada. 2019. "MAFFT Online Service:
681 Multiple Sequence Alignment, Interactive Sequence Choice and Visualization." *Briefings in*
682 *Bioinformatics* 20 (4): 1160–66.
683 Kobayashi, Koichi, Daichi Sasaki, Ko Noguchi, Daiki Fujinuma, Hirohisa Komatsu, Masami
684 Kobayashi, Mayuko Sato, et al. 2013. "Photosynthesis of Root Chloroplasts Developed in
685 Arabidopsis Lines Overexpressing GOLDEN2-LIKE Transcription Factors." *Plant & Cell*
686 *Physiology* 54 (8): 1365–77.
687 Letunic, Ivica, and Peer Bork. 2021. "Interactive Tree Of Life (iTOL) v5: An Online Tool for
688 Phylogenetic Tree Display and Annotation." *Nucleic Acids Research* 49 (W1): W293–96.
689 Li, Fay-Wei, Paul Brouwer, Lorenzo Carretero-Paulet, Shifeng Cheng, Jan de Vries, Pierre-Marc
690 Delaux, Ariana Eily, et al. 2018. "Fern Genomes Elucidate Land Plant Evolution and
691 Cyanobacterial Symbioses." *Nature Plants* 4 (7): 460–72.
692 Li, Fay-Wei, Tomoaki Nishiyama, Manuel Waller, Eftychios Frangedakis, Jean Keller, Zheng Li,
693 Noe Fernandez-Pozo, et al. 2020. "Anthoceros Genomes Illuminate the Origin of Land Plants
694 and the Unique Biology of Hornworts." *Nature Plants* 6 (3): 259–72.
695 Lin, Shih-Shun, and John L. Bowman. 2018. "MicroRNAs in Marchantia Polymorpha." *The New*
696 *Phytologist* 220 (2): 409–16.
697 Llave, Cesar, Kristin D. Kasschau, Maggie A. Rector, and James C. Carrington. 2002.
698 "Endogenous and Silencing-Associated Small RNAs in Plants." *The Plant Cell* 14 (7): 1605–
699 19.
700 Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold
701 Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550.
702 Luan, Ji, Jianfang Ju, Xiaochen Li, Xiuling Wang, Yufei Tan, and Guangmin Xia. 2022. "Functional
703 Identification of Moss PpGATA1 Provides Insights into the Evolution of LLM-Domain B-GATA
704 Transcription Factors in Plants." *Gene* 855: 147103.

705 Luo, Xiao-Min, Wen-Hui Lin, Shengwei Zhu, Jia-Ying Zhu, Yu Sun, Xi-Ying Fan, Menglin Cheng,
706 et al. 2010. "Integration of Light- and Brassinosteroid-Signaling Pathways by a GATA
707 Transcription Factor in Arabidopsis." *Developmental Cell* 19 (6): 872–83.
708 Ma, Zhaoxue, Xupeng Hu, Wenjuan Cai, Weihua Huang, Xin Zhou, Qian Luo, Hongquan Yang,
709 Jiawei Wang, and Jirong Huang. 2014. "Arabidopsis miR171-Targeted Scarecrow-Like
710 Proteins Bind to GT Cis-Elements and Mediate Gibberellin-Regulated Chlorophyll
711 Biosynthesis under Light Conditions." *PLoS Genetics*
712 https://doi.org/10.1371/journal.pgen.1004519.
713 Naito, Takahito, Takatoshi Kiba, Nobuya Koizumi, Takafumi Yamashino, and Takeshi Mizuno.
714 2007. "Characterization of a Unique GATA Family Gene That Responds to Both Light and
715 Cytokinin in Arabidopsis Thaliana." *Bioscience, Biotechnology, and Biochemistry* 71 (6):
716 1557–60.
717 Nakamura, Hidemitsu, Masayuki Muramatsu, Makoto Hakata, Osamu Ueno, Yoshiaki Nagamura,
718 Hirohiko Hirochika, Makoto Takano, and Hiroaki Ichikawa. 2009. "Ectopic Overexpression of
719 the Transcription Factor OsGLK1 Induces Chloroplast Development in Non-Green Rice
720 Cells." *Plant & Cell Physiology* 50 (11): 1933–49.
721 Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2015. "IQ-TREE:
722 A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies."
723 *Molecular Biology and Evolution* 32 (1): 268–74.
724 One Thousand Plant Transcriptomes Initiative. 2019. "One Thousand Plant Transcriptomes and
725 the Phylogenomics of Green Plants." *Nature* 574 (7780): 679–85.
726 Powell, Ann L. T., Cuong V. Nguyen, Theresa Hill, Kalai Lam Cheng, Rosa Figueroa-Balderas,
727 Hakan Aktas, Hamid Ashrafi, et al. 2012. "Uniform Ripening Encodes a Golden 2-like
728 Transcription Factor Regulating Tomato Fruit Chloroplast Development." *Science* 336
729 (6089): 1711–15.
730 Qiu, Yin-Long, Libo Li, Bin Wang, Zhiduan Chen, Volker Knoop, Milena Groth-Malonek, Olena
731 Dombrovska, et al. 2006. "The Deepest Divergences in Land Plants Inferred from
732 Phylogenomic Evidence." *Proceedings of the National Academy of Sciences of the United*
733 *States of America* 103 (42): 15511–16.
734 Rhoades, Matthew W., Brenda J. Reinhart, Lee P. Lim, Christopher B. Burge, Bonnie Bartel, and
735 David P. Bartel. 2002. "Prediction of Plant microRNA Targets." *Cell* 110 (4): 513–20.

736 Rich, Mélanie K., and Pierre-Marc Delaux. 2020. "Plant Evolution: When Arabidopsis Is More
737 Ancestral Than Marchantia." *Current Biology: CB*.
738 Rossini, L., L. Cribb, D. J. Martin, and J. A. Langdale. 2001. "The Maize golden2 Gene Defines a
739 Novel Class of Transcriptional Regulators in Plants." *The Plant Cell* 13 (5): 1231–44.
740 Safi, Alaeddine, Anna Medici, Wojciech Szponarski, Sandrine Ruffel, Benoît Lacombe, and Gabriel
741 Krouk. 2017. "The World according to GARP Transcription Factors." *Current Opinion in Plant*
742 *Biology* 39: 159–67.
743 Saint-Marcoux D, Proust H, Dolan L, Langdale JA (2015) Identification of Reference Genes for
744 Real-Time Quantitative PCR Experiments in the Liverwort *Marchantia polymorpha*. PLoS ONE
745 10(3): e0118678.
746 Sauret-Güeto, Susanna, Eftychios Frangedakis, Linda Silvestri, Marius Rebmann, Marta Tomaselli,
747 Kasey Markel, Mihails Delmans, Anthony West, Nicola J. Patron, and Jim Haseloff. 2020.
748 "Systematic Tools for Reprogramming Plant Gene Expression in a Simple Model,." *ACS*
749 *Synthetic Biology* 9 (4): 864–82.
750 Schreier, Tina B., Karin H. Müller, Simona Eicke, Christine Faulkner, Samuel C. Zeeman, and
751 Julian M. Hibberd. n.d. "Plasmodesmal Connectivity in C_4*Gynandropsis Gynandra*is Induced
752 by Light and Dependent on Photosynthesis." https://doi.org/10.1101/2022.12.07.519530.
753 Schulze, Silke, Barbara Nicole Schäfer, Eneida Abreu Parizotto, Olivier Voinnet, and Klaus Theres.
754 2010. "LOST MERISTEMS Genes Regulate Cell Differentiation of Central Zone Descendants
755 in Arabidopsis Shoot Meristems." *The Plant Journal: For Cell and Molecular Biology* 64 (4):
756 668–78.
757 Sugano, Shigeo S., Ryuichi Nishihama, Makoto Shirakawa, Junpei Takagi, Yoriko Matsuda, Sakiko
758 Ishida, Tomoo Shimada, Ikuko Hara-Nishimura, Keishi Osakabe, and Takayuki Kohchi. 2018.
759 "Efficient CRISPR/Cas9-Based Genome Editing and Its Application to Conditional Genetic
760 Analysis in Marchantia Polymorpha." *PloS One* 13 (10): e0205117.
761 Todesco, Marco, Ignacio Rubio-Somoza, Javier Paz-Ares, and Detlef Weigel. 2010. "A Collection
762 of Target Mimics for Comprehensive Analysis of microRNA Function in Arabidopsis Thaliana."
763 *PLoS Genetics* 6 (7): e1001031.
764 Trebst, Achim. 2007. "Inhibitors in the Functional Dissection of the Photosynthetic Electron
765 Transport System." *Photosynthesis Research* 92 (2): 217–24.

766 Tu, Xiaoyu, María Katherine Mejía-Guerra, Jose A. Valdes Franco, David Tzeng, Po-Yu Chu, Wei
767 Shen, Yingying Wei, et al. 2020. "Reconstructing the Maize Leaf Regulatory Network Using
768 ChIP-Seq Data of 104 Transcription Factors." *Nature Communications* 11 (1): 5089.
769 Tu, Xiaoyu, Sibo Ren, Wei Shen, Jianjian Li, Yuxiang Li, Chuanshun Li, Yangmeihui Li, et al. 2022.
770 "Limited Conservation in Cross-Species Comparison of GLK Transcription Factor Binding
771 Suggested Wide-Spread Cistrome Divergence." *Nature Communications* 13 (1): 7632.
772 Vries, Jan de, and John M. Archibald. 2018. "Plant Evolution: Landmarks on the Path to Terrestrial
773 Life." *The New Phytologist* 217 (4): 1428–34.
774 Vries, Jan de, and Sven B. Gould. 2018. "The Monoplastidic Bottleneck in Algae and Plant
775 Evolution." *Journal of Cell Science* 131 (2).
776 Wang, Long, Yan-Xia Mai, Yan-Chun Zhang, Qian Luo, and Hong-Quan Yang. 2010.
777 "MicroRNA171c-Targeted SCL6-II, SCL6-III, and SCL6-IV Genes Regulate Shoot Branching
778 in Arabidopsis." *Molecular Plant* 3 (5): 794–806.
779 Wang, Peng, Roxana Khoshravesh, Shanta Karki, Ronald Tapia, C. Paolo Balahadia, Anindya
780 Bandyopadhyay, W. Paul Quick, Robert Furbank, Tammy L. Sage, and Jane A. Langdale.
781 2017. "Re-Creation of a Key Step in the Evolutionary Switch from C to C Leaf Anatomy."
782 *Current Biology: CB* 27 (21): 3278–87.e6.
783 Waters, Mark T., Elizabeth C. Moylan, and Jane A. Langdale. 2008. "GLK Transcription Factors
784 Regulate Chloroplast Development in a Cell-Autonomous Manner." *The Plant Journal: For*
785 *Cell and Molecular Biology* 56 (3): 432–44.
786 Waters, Mark T., Peng Wang, Muris Korkaric, Richard G. Capper, Nigel J. Saunders, and Jane A.
787 Langdale. 2009. "GLK Transcription Factors Coordinate Expression of the Photosynthetic
788 Apparatus in Arabidopsis." *The Plant Cell* 21 (4): 1109–28.
789 Yasumura, Yuki, Elizabeth C. Moylan, and Jane A. Langdale. 2005. "A Conserved Transcription
790 Factor Mediates Nuclear Control of Organelle Biogenesis in Anciently Diverged Land Plants."
791 *The Plant Cell* 17 (7): 1894–1907.
792 Yelina, Nataliya E., Daniel Holland, Sabrina Gonzalez-Jorge, Dominique Hirsz, Ziyi Yang, and Ian
793 R. Henderson. 2022. "Coexpression of MEIOTIC-TOPOISOMERASE VIB-dCas9 with Guide
794 RNAs Specific to a Recombination Hotspot Is Insufficient to Increase Crossover Frequency
795 in Arabidopsis." *G3* 12 (7).
796 Zheng, Yi, Chen Jiao, Honghe Sun, Hernan G. Rosli, Marina A. Pombo, Peifen Zhang, Michael
797 Banf, et al. 2016. "iTAK: A Program for Genome-Wide Prediction and Classification of Plant

798 Transcription Factors, Transcriptional Regulators, and Protein Kinases." *Molecular Plant* 9
799 (12): 1667–70.
800 Zubo, Yan O., Ivory Clabaugh Blakley, José M. Franco-Zorrilla, Maria V. Yamburenko, Roberto
801 Solano, Joseph J. Kieber, Ann E. Loraine, and G. Eric Schaller. 2018. "Coordination of
802 Chloroplast Development through the Action of the GNC and GLK Transcription Factor
803 Families." *Plant Physiology* 178 (1): 130–47.
804
805
806
807
808
809
810
811
812
813
814
815


A
GATA
GARP
GATA2
GNC
CGA1
GLK
GA
mir171
BR
PIFs
DELLAs
SCL
GRAS
photomor-phogenesis
chloroplast PhANGS
Chlorophyll biosynthesis
chloroplast biogenesis
chloroplast division
B
GATA-B
GATA-A
GATA
GRAS
GLK
GARP_G2-like
Vascular plants
Bryophytes
Angiosperms
Gymnosperms
Ferns
Lycophytes
Mosses
Liverworts
Hornworts
Zygnematophyceae
Klebsormidiophyceae
Charophytes
Chlorophytes
Arabidopsis thaliana
Oryza sativa
Picea abies
Ceratopteris richardii
Azolla filiculoides
Selaginella moellendorffii
Physcomitrium patens
Marchantia polymorpha
Anthoceros agrestis
Spirogloea muscicola
Klebsormidium flaccidum
Chara braunii
Chlamydomonas reinhardtii
#genes
C
B-class GATA domain
gRNA1
gRNA2
LLM domain
MpGATA4
A-class GATA domain
gRNA
MpGATA2
MpSCL
50bp
MpMIR171
DNA Binding Domain
GCT box
MpGLK
1kb
D
control
E
Mpgata4
F
Mpgata2
G
Mpscl
H
Mpamir171
I
Mpglk
J
Mpgata4
K
Mpgata2
L
Mpamir171
M
Mpscl
N
Mpglk
Chlorophyll content
Chlorophyll a/b ratio
Chlorophyll mg/g FW
Chlorophyll a/b ratio
control
#1
#2
#3
n.s.
***

820 **Figure 1: Mp*GLK* controls chlorophyll.** A) Transcription factors known to regulate chloroplast
821 development in angiosperms. GATA TFs - GATA NITRATE- INDUCIBLE CARBON
822 METABOLISM-INVOLVED (GNC) and CYTOKININ- RESPONSIVE GATA FACTOR 1 (CGA1)
823 control chloroplast development by suppressing phytochrome interacting factors and
824 brassinosteroid (BR) related genes, promoting chloroplast biogenesis and division. GATA2
825 promotes photomorphogenesis in the presence of light by directly binding to light-responsive
826 promoters. In the absence of light, the BR-activated transcription factor BRASSINAZOLE
827 RESISTANT1 (BZR1) represses GATA2 expression, inhibiting photomorphogenesis. GOLDEN2-
828 LIKE transcriptions factors (TFs) are positive regulators of nuclear-encoded photosynthesis
829 related genes. SCARECROW-LIKE (SCL) GRAS TFs are negatively regulated by miR171 and
830 GA-DELLA signalling, to control chlorophyll biosynthesis. B) Phylogenetic relationships of the
831 major lineages of land plants and green algae (Li et al. 2020). C) Schematic representation of
832 Mp*GLK,* Mp*GATA4,* Mp*GATA2* and Mp*SCL* gene structure showing exons as black rectangles.
833 Characteristic gene domains are shown as coloured boxes. Mp*MIR171* genomic locus is shown
834 in a dashed box. Mp*MIR171* gene is represented as a black rectangle, miR171* and miR171
835 shown as blue rectangles, miR171* indicated with a white star. gRNAs positions for
836 CRISPR/Cas9 gene editing are shown as red arrows. D-I) Images of wild type control, Mp*gata4*,
837 Mp*gata2*, Mp*scl*, Mp*mir171* and Mp*glk* mutant plants. Scale bars represent 2 mm. J-N) Barplots
838 of chlorophyll content and chlorophyll a/b ratio for the Mp*gata4*, Mp*gata2*, Mp*scl*, Mp*mir17*1 and
839 Mp*glk* mutant plants. Individual values are shown with dots. Error bars represent the standard
840 deviation of the mean $n = 5$. Asterisks indicate statistically significant difference using a two-tailed
841 *t*-test $P \leq 0.0001$ (***)
842

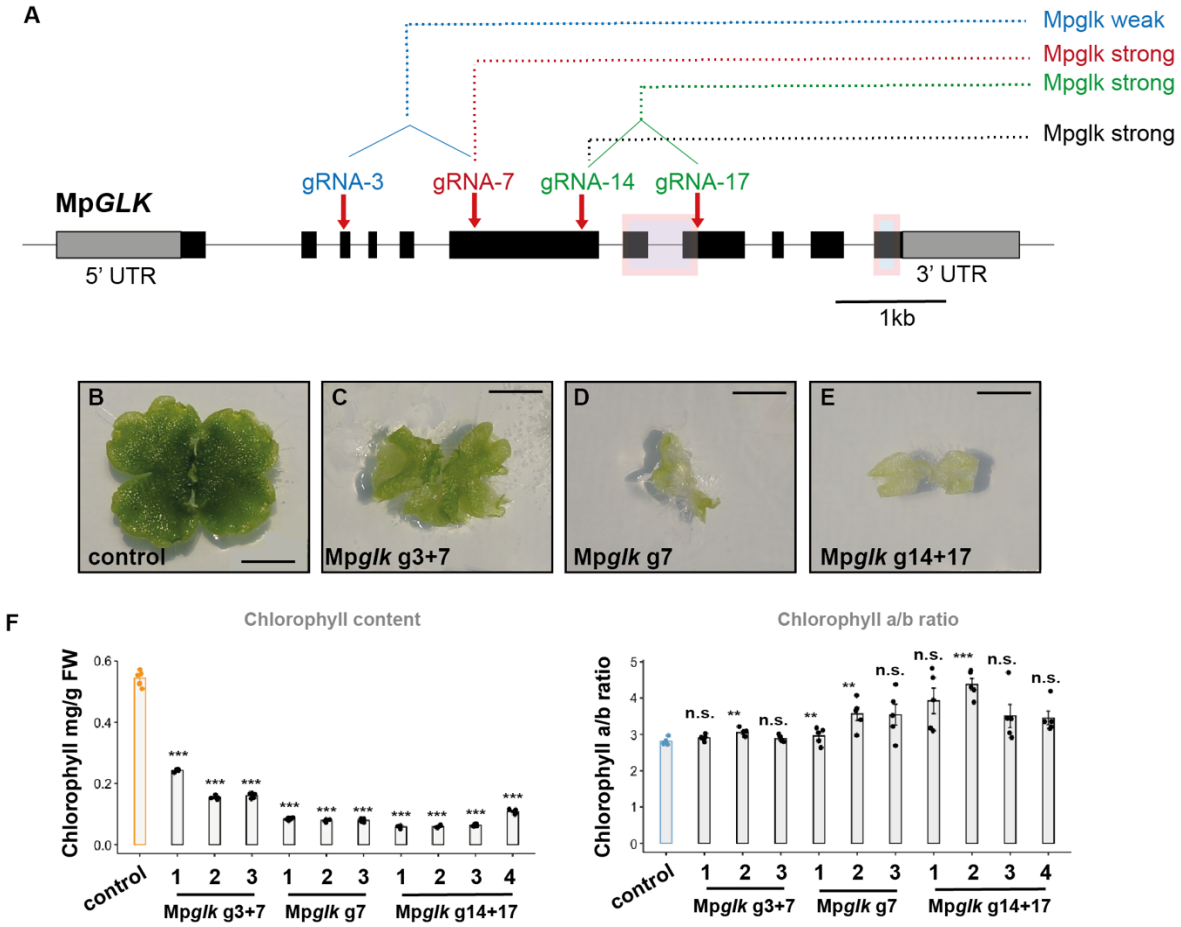
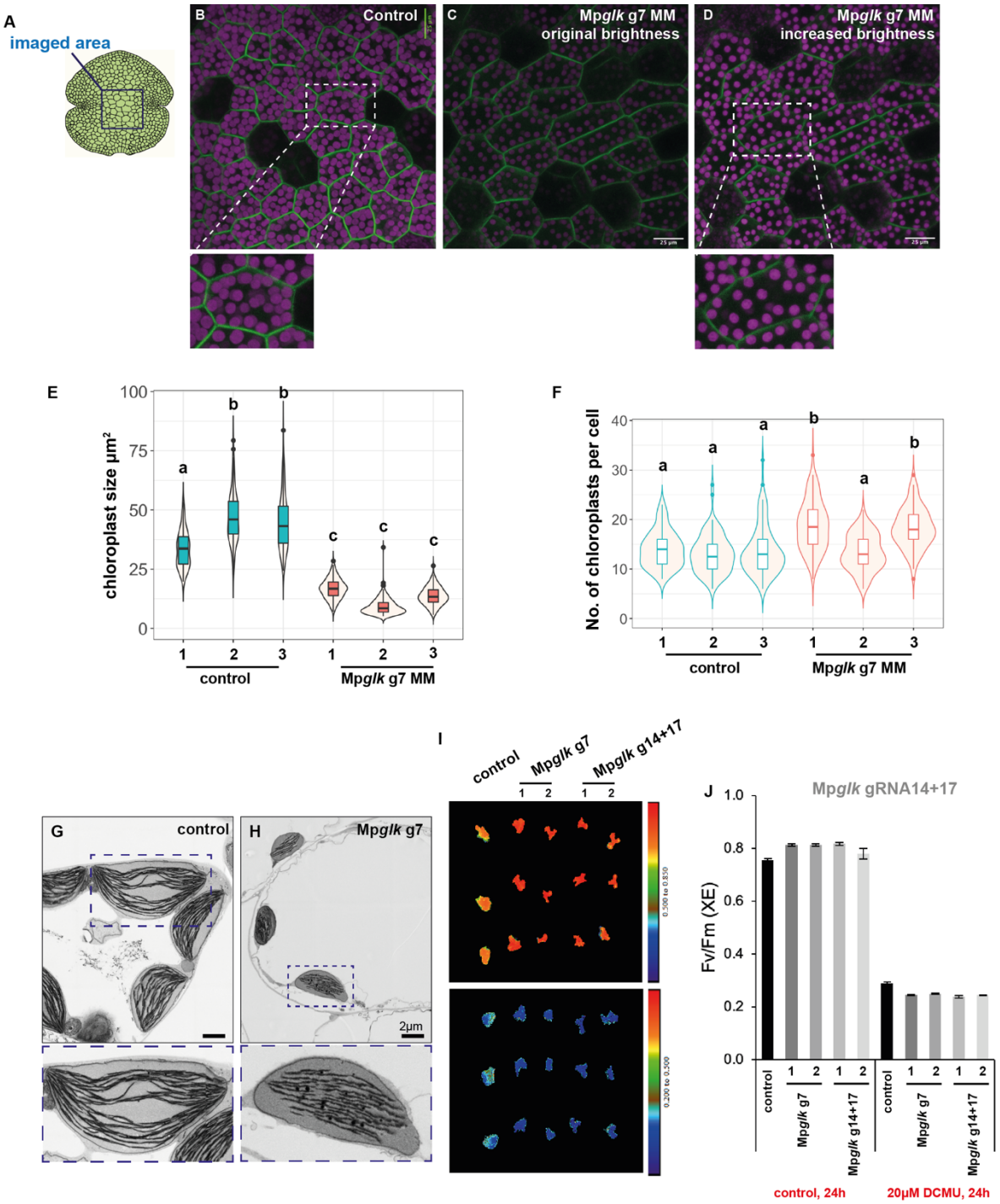


Figure 2: Mp*GLK* mutants. A) Schematic representation of Mp*GLK* gene structure showing exons as black rectangles, untranslated regions (UTRs) as grey rectangles and introns as black lines. Characteristic gene domains are highlighted by shaded boxes. Positions of gRNAs used for CRISPR/Cas9 gene editing are shown as red arrows. B-E) Images of control and Mp*glk* mutant plants. Scale bars represent 2mm. F) Barplots representing chlorophyll content and chlorophyll a:b ratio in Mp*glk* mutants. Individual values are shown as dots. Error bars, represent standard deviation of the mean form n = 5. Asterisks indicate statistically significant differences using a two-tailed t-test $P \leq 0.0001$ (***), $0.0001 \leq P \leq 0.001$ (**) and n.s.: non-significant.


A
imaged area
B
Control
C
Mpglk g7 MM
original brightness
D
Mpglk g7 MM
increased brightness
E
chloroplast size µm²
control
Mpglk g7 MM
F
No. of chloroplasts per cell
control
Mpglk g7 MM
G
control
H
Mpglk g7
2µm
I
control
Mpglk g7
Mpglk g14+17
J
Mpglk gRNA14+17
Fv/Fm (XE)
control, 24h
20µM DCMU, 24h

867 **Figure 3: Mp*glk* controls chloroplast size and ultrastructure.** A) Schematic of a gemmae, a
868 dark blue square indicates imaged area. B-D) Representative confocal microscopy images of
869 control (B) and *Mp*glk mutant (C, D) plant gemmae. Chlorophyll autofluorescence is shown in
870 magenta and plasma membrane marked with eGFP shown in green. Magnified cells are shown
871 at the bottom of panels B and D. Scale bars, represent 25 µm. E) Boxplot showing chloroplast
872 size ranges in control and Mp*glk* plants. F) Boxplot showing chloroplast numbers per cell area in
873 control and Mp*glk* plants. Box and whiskers represent the 25 to 75 percentile and minimum-
874 maximum distributions of the data. Letters show statistical ranking using a *post hoc* Tukey test
875 (with different letters indicating statistically significant differences at $P<0.01$). Values indicated by
876 the same letter are not statistically different. G, H) Scanning electron micrographs of chloroplasts
877 in control (G) and Mp*glk* mutant (H) plants. Dashed boxes highlight single chloroplasts that are
878 shown in the insets at a higher magnification (bottom panels). Scale bars represent 2 µm. I)
879 Chlorophyll fluorescence images of maximum quantum efficiency of Photosystem II
880 photochemistry (*Fv/Fm*) of untreated (top) and 24 h DCMU-treated (bottom) control and Mp*glk*
881 mutant plants. J) *Fv/Fm* measured in DCMU-treated and untreated control and Mp*glk* mutant
882 plants. Bars represent mean ± standard error from $n = 3$ plants per genotype.
883
884
885
886
887
888

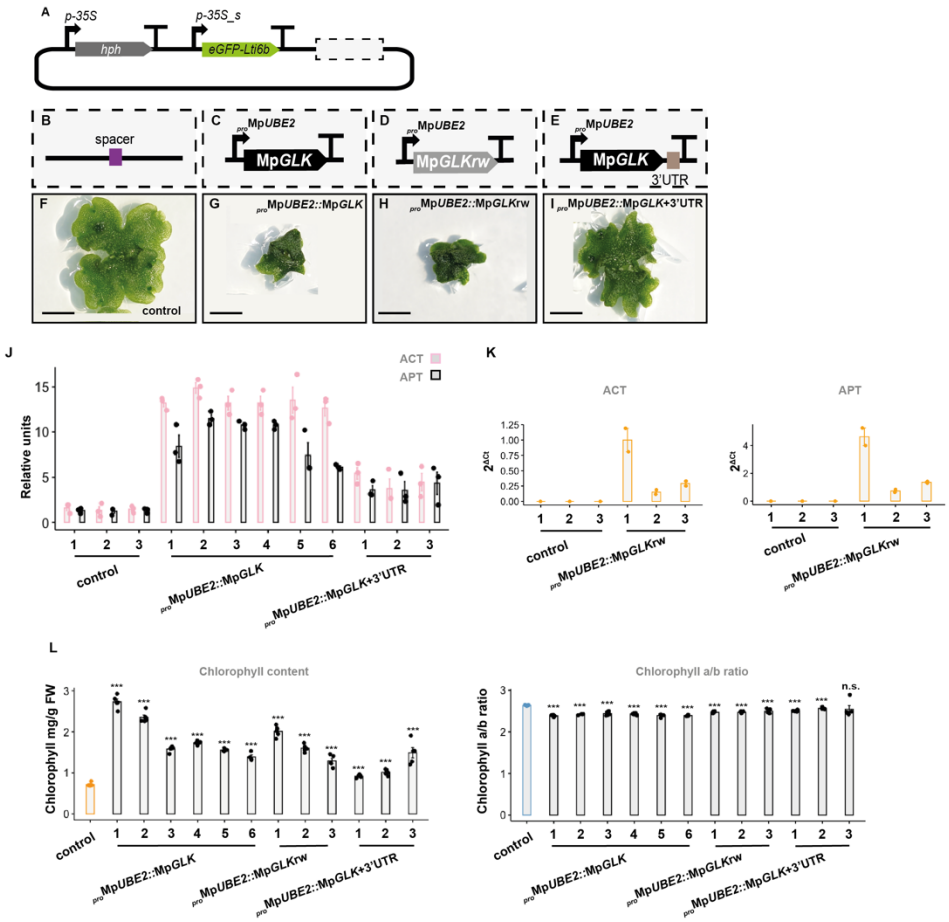


Figure 4: Mp*GLK* overexpression increases chlorophyll content. A-E) Schematic representation of transformation constructs used for Mp*GLK-CDS*, Mp*GLK-CDSrw* and Mp*GLK-CDS+3´UTR* overexpression as well as the 'empty vector' control. hph: hygromycin B phosphotransferase, 35S: cauliflower mosaic virus (CaMV) 35S promoter. F-I) Images of 'empty vector' control plants and plants transformed with Mp*GLK-CDS*, Mp*GLK-CDSrw* and Mp*GLK-CDS+3'UTR* overexpression constructs. Scale bars represent 2mm. J) Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis of Mp*GLK* levels in Mp*GLK-CDS* and Mp*GLK-CDS+3'UTR* overexpressing lines. (K) Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis of Mp*GLK*-CDSrw levels in Mp*GLK* CDSrw overexpressing lines. *ADENINE PHOSPHORIBOSYL TRANSFERASE 3* (*APT*) and *ACTIN 7* (ACT) were used as housekeeping gene controls (Saint-Marcoux et al. 2015). L) Barplots of chlorophyll content and chlorophyll a/b ratios in Mp*GLK-CDS*, Mp*GLK-CDSrw* and Mp*GLK-CDS+3'UTR* overexpressing lines compared to 'empty vector' controls. Individual values are shown with dots. Error bars, represent standard deviation of the mean from n = 3. Asterisks indicate statistically significant difference using a two-tailed t-test $P \leq 0.0001$ (***) and n.s.: non-significant.

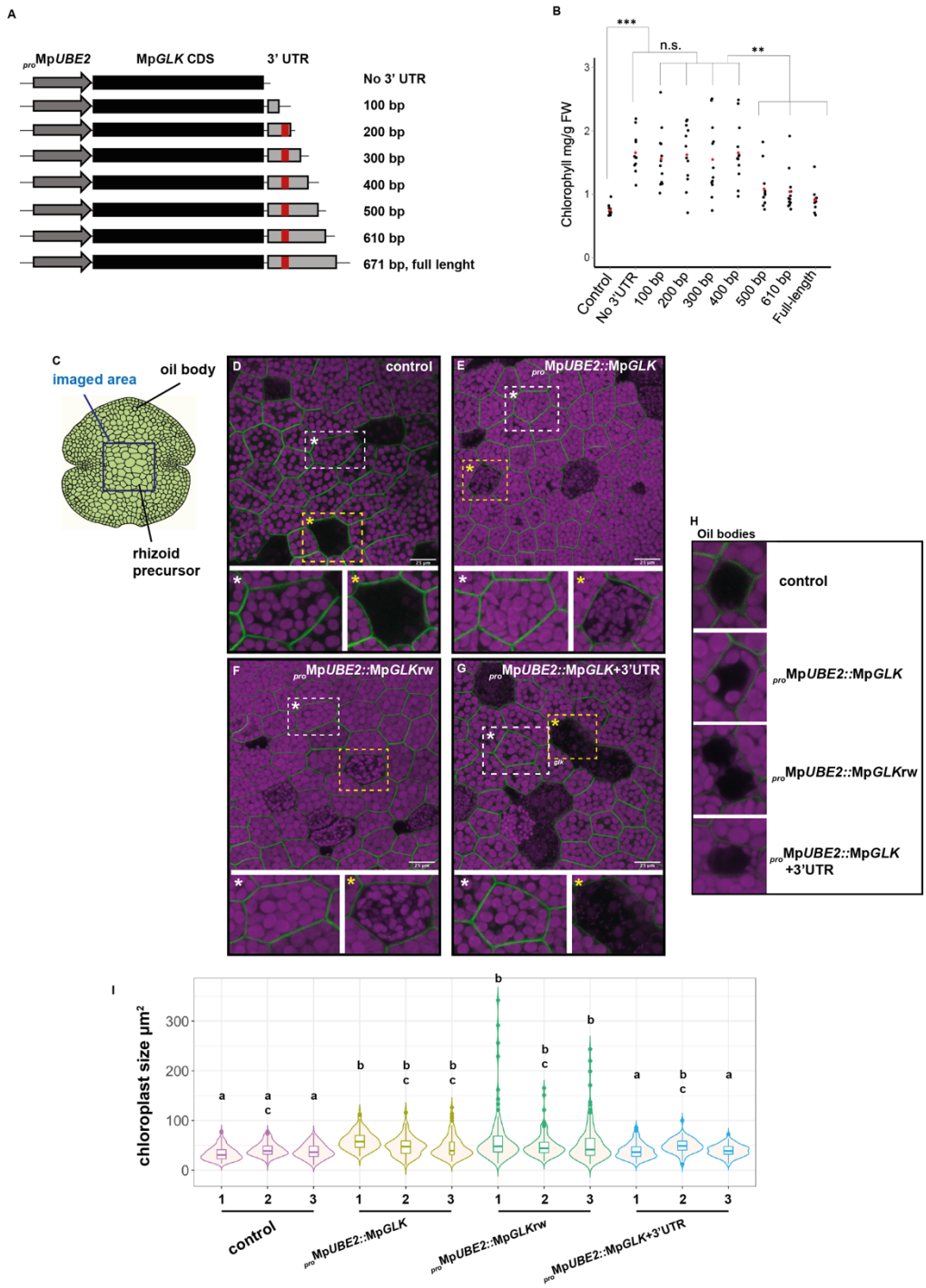
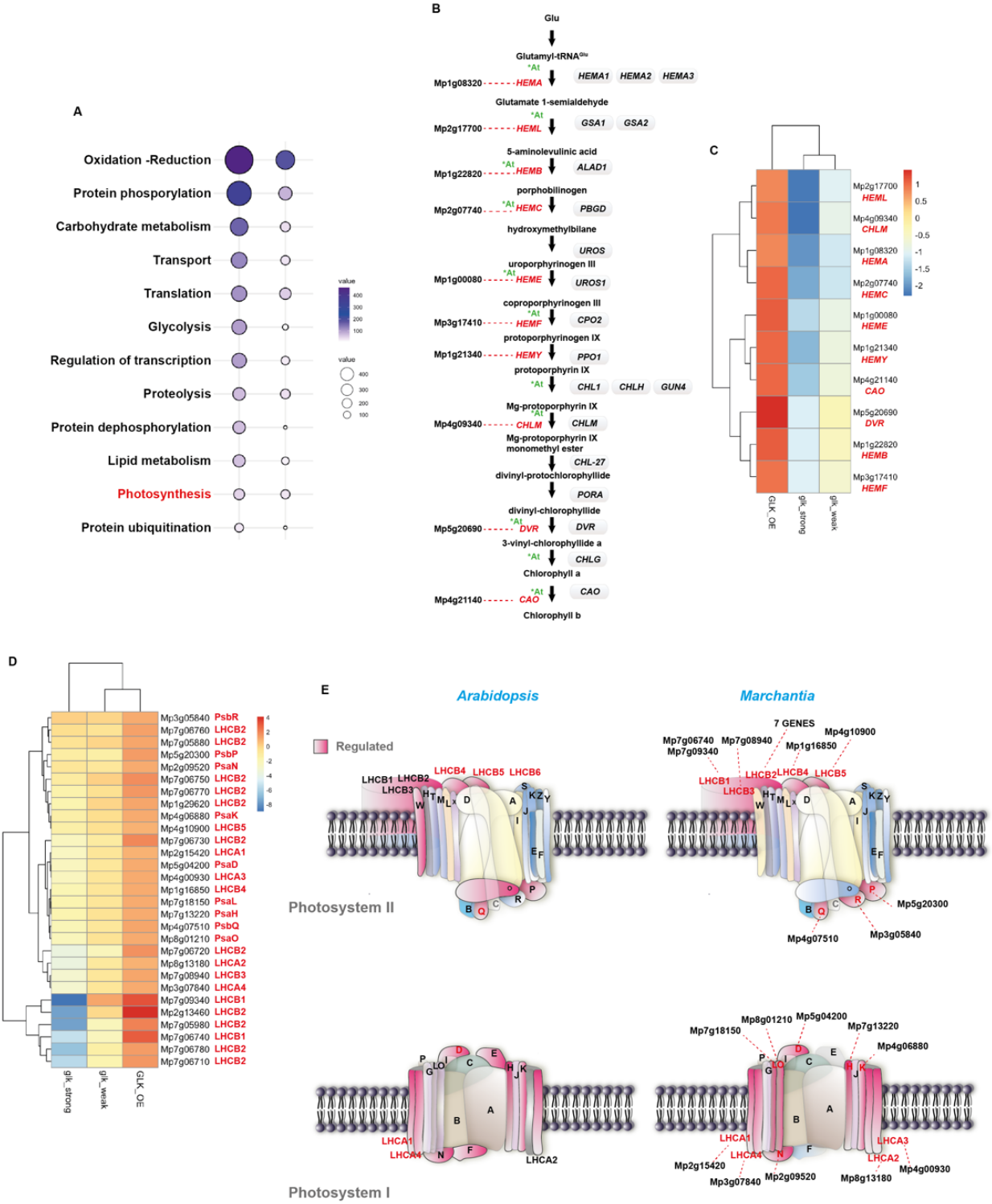

A
pro MpUBE2
MpGLK CDS
3' UTR
No 3' UTR
100 bp
200 bp
300 bp
400 bp
500 bp
610 bp
671 bp, full lenght
B
***
n.s.
**
Chlorophyll mg/g FW
Control
No 3'UTR
100 bp
200 bp
300 bp
400 bp
500 bp
610 bp
Full-length
C
imaged area
oil body
rhizoid precursor
D
control
E
pro MpUBE2::MpGLK
F
pro MpUBE2::MpGLKrw
G
pro MpUBE2::MpGLK+3'UTR
H
Oil bodies
control
pro MpUBE2::MpGLK
pro MpUBE2::MpGLKrw
pro MpUBE2::MpGLK +3'UTR
I
chloroplast size µm²
control
pro MpUBE2::MpGLK
pro MpUBE2::MpGLKrw
pro MpUBE2::MpGLK+3'UTR

Figure 5: Mp*GLK* 3´UTR controls chlorophyll content. A) Schematic representation of overexpression constructs where Mp*GLK* CDS is fused to a full-length or a series of truncated Mp*GLK* 3´UTR versions. Grey arrows represent Mp*UBE2* promoter, black rectangles - Mp*GLK* CDS, grey rectangles – a full-length or truncated Mp*GLK* 3´UTR. Red boxes indicate Mp*GLK* 3´UTR region harbouring a predicted miR11666.4 and another putative miRNA/siRNA recognition/cleavage sites. B) Jitter plot showing chlorophyll content in transgenic lines overexpressing Mp*GLK* CDS with a full-length or truncated Mp*GLK* 3´ UTRs. Black dots represent individual transformants, red dots indicate mean average. Asterisks indicate statistically significant difference using a two-tailed *t*-test $P \leq 0.0001$ (***), $P \leq 0.001$ (**) and n.s.: non-significant. C) Schematic of a gemmae; a dark blue square indicates imaged area, black arrows indicate approximate positions of rhizoid precursor and oil body cells. D-G) Representative confocal microscopy images of control plant gemmae and Mp*GLK-CDS*, Mp*GLK-CDSrw* and Mp*GLK-CDS+3´UTR* overexpressing lines. Chlorophyll autofluorescence is shown in magenta and plasma membrane marked with eGFP shown in green. Magnified cells (white asterisks) and rhizoids precursors (yellow asterisks) are shown under each panel. Scale bars, represent 25 µm. H) Representative confocal microscopy images of oil body cells in a control plant and Mp*GLK-CDS*, Mp*GLK-CDSrw* and Mp*GLK-CDS+3´UTR* overexpressing lines. I) Boxplot showing chloroplast size range in control versus Mp*GLK-CDS*, Mp*GLK-CDSrw* and Mp*GLK-CDS+3´UTR* overexpressing lines. Box and whiskers represent the 25 to 75 percentile and minimum-maximum distributions of the data. Letters show statistical ranking using a *post hoc* Tukey test (with different letters indicating statistically significant differences at $P<0.01$). Values indicated by the same letter are not statistically different.


A
Oxidation -Reduction
Protein phosporylation
Carbohydrate metabolism
Transport
Translation
Glycolysis
Regulation of transcription
Proteolysis
Protein dephosphorylation
Lipid metabolism
Photosynthesis
Protein ubiquitination
value
400
300
200
100
B
Glu
Glutamyl-tRNA^Glu
Mp1g08320 HEMA
HEMA1 HEMA2 HEMA3
Glutamate 1-semialdehyde
Mp2g17700 HEML
GSA1 GSA2
5-aminolevulinic acid
Mp1g22820 HEMB
ALAD1
porphobilinogen
Mp2g07740 HEMC
PBGD
hydroxymethylbilane
UROS
uroporphyrinogen III
Mp1g00080 HEME
UROS1
coproporphyrinogen III
Mp3g17410 HEMF
CPO2
protoporphyrinogen IX
Mp1g21340 HEMY
PPO1
protoporphyrin IX
CHL1 CHLH GUN4
Mg-protoporphyrin IX
Mp4g09340 CHLM
CHLM
Mg-protoporphyrin IX monomethyl ester
CHL-27
divinyl-protochlorophyllide
PORA
divinyl-chlorophyllide
Mp5g20690 DVR
DVR
3-vinyl-chlorophyllide a
CHLG
Chlorophyll a
Mp4g21140 CAO
CAO
Chlorophyll b
C
Mp2g17700 HEML
Mp4g09340 CHLM
Mp1g08320 HEMA
Mp2g07740 HEMC
Mp1g00080 HEME
Mp1g21340 HEMY
Mp4g21140 CAO
Mp5g20690 DVR
Mp1g22820 HEMB
Mp3g17410 HEMF
GLK_OE
glk_strong
glk_weak
D
Mp3g05840 PsbR
Mp7g06760 LHCB2
Mp7g05880 LHCB2
Mp5g20300 PsbP
Mp2g09520 PsaN
Mp7g06750 LHCB2
Mp7g06770 LHCB2
Mp1g29620 LHCB2
Mp4g06880 PsaK
Mp4g10900 LHCB5
Mp7g06730 LHCB2
Mp2g15420 LHCA1
Mp5g04200 PsaD
Mp4g00930 LHCA3
Mp1g16850 LHCB4
Mp7g18150 PsaL
Mp7g13220 PsaH
Mp4g07510 PsbQ
Mp8g01210 PsaO
Mp7g06720 LHCB2
Mp8g13180 LHCA2
Mp7g08940 LHCB3
Mp3g07840 LHCA4
Mp7g09340 LHCB1
Mp2g13460 LHCB2
Mp7g05980 LHCB2
Mp7g06740 LHCB1
Mp7g06780 LHCB2
Mp7g06710 LHCB2
glk_strong
glk_weak
GLK_OE
E
Arabidopsis
Marchantia
Regulated
LHCB1 LHCB2 LHCB3 LHCB4 LHCB5 LHCB6
7 GENES
Mp7g06740 Mp7g09340 Mp7g08940 Mp1g16850 Mp4g10900
Mp5g20300
Mp4g07510
Mp3g05840
Photosystem II
LHCA1 LHCA4 LHCA2
Mp7g18150 Mp8g01210 Mp5g04200 Mp7g13220 Mp4g06880
Mp2g15420 Mp3g07840 Mp2g09520 Mp8g13180 Mp4g00930
LHCA3
Photosystem I

942

943

944

945 **Figure 6: Mp*GLK* controls gene expression.** A) Gene ontology (GO) term enrichment of
946 Mp*GLK* overexpression plants versus 'strong' allele mutants. Differentially expressed genes with
947 p-value=<0.01 were identified and were grouped according to their GO terms corresponding to
948 biological processes. Terms with at least twenty genes are shown and the size of the dots
949 represents the number of genes corresponding to a given term. B) Summary of the chlorophyll
950 biosynthetic pathway. With "A*" genes that are regulated by *GLK1* and *GLK2* genes in *A. thaliana*.
951 C) Heatmap of differentially expressed chlorophyll biosynthetic pathway genes
952 (Log2FoldChange). D) Heatmap of differentially expressed of Photosystems I and II genes
953 (Log2FoldChange). E) Schematic representation of Photosystem I and II. Subunits showing an
954 increase (>1) in corresponding transcript levels are highlighted in pink. Genes that are
955 differentially expressed are highlighted in red. Figure modified from Water et al. 2009 and Tu et
956 al. 2022.

957
958
959