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

Nataliya E. Yelina[1,2,*], Eftychios Frangedakis[1,*,$], Zhemin Wang[3], Tina B. Schreier[1,4], Jenna Rever[1], Marta Tomaselli[1], Edith Forestier[1], Kumari Billakurthi[1], Sibo Ren[3], Yahui Bai[3], Julia Stewart-Wood[1], Jim Haseloff[1], Silin Zhong[3] and Julian M. Hibberd[1,$]

[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

[3]The State Key Laboratory of Agrobiotechnology, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China

[4]Present address: Department of Biology, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

* equal contributions

$ correspondence:

Julian M. Hibberd jmh65@cam.ac.uk

Eftychios Frangedakis ef391@cam.ac.uk

Keywords: Chloroplast biogenesis, chlorophyll, Marchantia, transcription factors

Summary

Photosynthesis in eukaryotic cells takes place in chloroplasts that develop from undifferentiated plastids in response to light. 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. Additionally, the Elongated Hypocotyl 5 (HY5) transcription factor plays a critical role in photomorphogenesis. 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, SCL and HY5 proteins in controlling chloroplast development in the model liverwort *Marchantia polymorpha*. Our results indicate that GATA, SCL and HY5 have a minimal or undetectable role in chloroplast biogenesis but loss of GLK function leads to reduced chloroplast size, underdeveloped thylakoid membranes and lower chlorophyll accumulation. ChIP-seq and RNA-seq analyses revealed that GLK can directly regulate many photosynthetic and chloroplast development-related genes. But there is extensive divergence between its function in *M. polymorpha* and flowering plants. Collectively our findings suggest that the function of GATA as well as SCL in chloroplast development and the more specialised role of HY5 in photomorphogenesis, either evolved after the divergence of vascular plants from bryophytes, that were 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(Cackett et al., 2022). 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 & López-Juez, 2013). 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 & 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 that allow the integration of these responses 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 (Bravo-Garcia et al., 2009; Fitter et al., 2002; Waters et al., 2008, 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 (Bi et al., 2005; Chiang et al., 2012; Hudson et al., 2011; Naito et al., 2007). 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* miR171 targeted SCL transcription factors redundantly regulate chlorophyll biosynthesis by repressing the expression of *Protochlorophyllide Oxidoreductase C,* a key enzyme for chlorophyll production (Ma et al., 2014). Particularly crucial is also the Elongated Hypocotyl 5 (HY5), a bZIP transcription factor that plays a primary role in de-etiolation and chloroplast development (Burko et al., 2020; Lee et al., 2007; Oyama et al., 1997). HY5 acts antagonistically with Phytochrome Interacting Factors (PIFs) to activate light-regulated genes in the presence of light (Xiao et al., 2021). Notably, a recent detailed study has shown that in *A. thaliana*, GLK and GATAs are direct targets of HY5 (T. Zhang et al., 2023). Land plants evolved from aquatic green algae, and approximately 500 MYA diverged into two major monophyletic clades, the vascular plants (which include angiosperms, gymnosperms, ferns and lycophytes) and the bryophytes (which include hornworts, liverworts and mosses) (**Figure 1B**)(de Vries & Archibald, 2018; Harris et al., 2022; Li et al., 2020). Despite our detailed knowledge of chloroplast biogenesis in angiosperms, our understanding of this process in other land lineages including the bryophytes is poor. With the exception of GLK function in the moss *Physcomitrium patens* (Yasumura et al., 2005) little is known about chloroplast biogenesis in bryophytes. Consequently, how this fundamental biological process may have evolved is unclear. Bryophytes are of particular importance to infer more accurately the ancestral state of a trait. Whilst earlier evolutionary hypotheses proposed that bryophytes represent a collection of paraphyletic lineages (Qiu et al., 2006), current phylogenetic analyses indicate that the liverworts and mosses form a monophyletic group that split from hornworts approximately 400-500 MYA (Harris et al., 2022; Li et al., 2020; One Thousand Plant Transcriptomes Initiative, 2019). This revised land plant phylogeny supports the notion that traits of the common ancestor of land plants could have diversified not only within the vascular plant clade but also in the three deeply divergent groups of bryophytes. Thus, studying representative species from more than one bryophyte clade is useful to gain insight into the likely ancestral state of any trait of interest. We therefore selected the model liverwort *Marchantia polymorpha* as a system to further investigate processes underpinning the evolution of chloroplast biogenesis. *M. polymorpha* has a small and well-annotated genome, key steps in its development are easily

accessible for observation, and an extensive set of genetic manipulation tools are available (Bowman et al., 2022; Sauret-Güeto et al., 2020).

Phylogenetic analysis identified putative homologs for GATAs (GNC/CGA1 and GATA2), the SCL-MIR171 module, as well as GLK. Additionally, a single HY5 homolog has been previously identified (Clayton et al., 2018). Thus, the regulatory network controlling chloroplast biogenesis in *A. thaliana* may be conserved in *M. polymorpha*. To test this, we generated knock-out mutants of each gene, constitutive over-expressors and performed RNA sequencing of mutant alleles. Thallus growth was perturbed by editing the Mp*GATA,* Mp*SCL*, and Mp*HY5* genes, however, no differences in chlorophyll content was detectable when plants were grown under normal conditions. In contrast, when Mp*GLK* was mutated chlorophyll accumulation was reduced and chloroplast development perturbed. Similarly, only constitutive overexpression of Mp*GLK* led to increased chlorophyll content and chloroplast size. To better understand the extent to which *GLK* controls similar sets of genes in *M. polymorpha* and *A. thaliana* we performed RNA sequencing analysis on Mp*glk* mutants and on constitutive Mp*GLK* over-expressors as well as chromatin immunoprecipitation assays followed by sequencing (ChIP-seq) on constitutive Mp*GLK* over-expressors. As previously reported in *A. thaliana*, the abundance of transcripts derived from genes associated with chlorophyll biogenesis and Photosystems I and II were impacted when *GLK* was mis-expressed in *M. polymorpha*. Our ChIP-seq analysis confirmed this, however, it also uncovered that several MpGLK targets are distinct from those identified in flowering plants.

We conclude that GLK function is conserved between *M. polymorpha* and angiosperms but that when knocked out in *M. polymorpha* other regulators such as *GNC/CGA1, GATA2, HY5* and *SCL* have no or minimal impact on chloroplast development. Thus, compared with angiosperms, the regulation of chloroplast biogenesis in *M. polymorpha* is distinct and this may reflect more streamlined regulatory pathways.

Results

M. polymorpha possesses homologs of the *GATA, SCL, HY5* and *GLK* transcription factors that regulate chloroplast biogenesis in angiosperms

We used phylogenetic analysis to search for *GATA* (*GNC*, *CGA1* and *GATA2*), *SCL* and *GLK* orthologs in *M. polymorpha*. To do so, we examined twenty-one representative species from the seven main groups of land plants as well as four green algae for which high quality genome assemblies are available (for clarity a subset of species is shown in **Figure 1B, full set of genes in Figure S1A**). These included three mosses, one liverwort, one hornwort, one lycophyte, three ferns, two gymnosperms and five angiosperms, and in the green algae two species of the Zygnematophyceae that are proposed to represent the sister group of land plants, as well as one member from each of the Klebsormidiophyceae, Charophyte and Chlorophyte lineages (**Figure S1A**). For *HY5* a single *M. polymorpha* ortholog has been identified in a previous study (Clayton et al., 2018).

GNC, *CGA1* and *GATA2* belong to the GATA superfamily of transcription factors that comprise a family of zinc finger proteins present in all eukaryotes. Plant GATAs can be categorised into four distinct classes (A, B, C and D) based on sequence conservation in the zinc finger motif, additional protein domains, and intron positions (Reyes et al., 2004). *GNC* and *CGA1* belong to the LLM subclass of B-Class GATAs due to a conserved Leucine–Leucine–Methionine (LLM) domain at their C-terminus (Behringer & Schwechheimer, 2015) (**Figure 2A and Figure S1B**). *GATA2* is an A-Class GATA, which is characterised by the presence of a conserved glutamine (Q) and a threonine (T) within the zinc finger motif (**Figure 2A and Figure S1C**). Our phylogenetic analysis showed that *M. polymorpha* has a total of six GATA genes, among which we identified *Mp7g03490* (annotated as Mp*GATA4*) and *Mp1g03950* (annotated as Mp*GATA2*) as single orthologs of *GNC/CGA1* and *GATA2* respectively (**Figure 1B, Figure S1B,S1C, S2 S3A, and Dataset S1**). *SCL* is a member of the GRAS family of transcription factors which have a conserved C-terminal domain (**Figure 1B and Figure 2A**). *A. thaliana AtSCL6, AtSCL22* and *AtSCL27* redundantly control chlorophyll biosynthesis and are regulated by MIR171. The *M. polymorpha* genome encodes eleven GRAS transcription factors. Genome-wide bioinformatic and next-generation sequencing analysis of *M. polymorpha* miRNAs and their potential targets suggests that

Mp8g03980 (annotated as Mp*GRAS10*) is a target of MIR171(Lin & Bowman, 2018). Although *Mp8g03980* is located in a sister clade to *A. thaliana* At*SCL6*, At*SCL22* and At*SCL27* (Lin & Bowman, 2018), based on the conservation of its miR171-mediated regulation we refer to *Mp8g03980* as Mp*SCL*.

Lastly, *GLK* genes belong to the GARP (GOLDEN2-LIKE, ARR-B, Psr1) family of transcription factors (Safi et al., 2017) and can be divided into two main subfamilies comprising the GARP-G2-like and GARP-ARR-B transcription factors. *GLK* genes belong to the GARP-G2-like group and are distinguished by a characteristic C-terminal domain called GOLDEN2 C-terminal (GCT) box (**Figure 1B, 2A, Figure S1D and S3B**) (Fitter et al., 2002). Our phylogenetic analysis showed that *GLK* is present in one or two copies in most land plants, but multiple copies are present in some moss and fern species. *GLK* genes were thought to be land plant specific, however, we were able to detect *GLK* orthologs in two Zygnematophyceae green algae for which genome assemblies became recently available (Cheng et al., 2019) (two genes in *Spirogloea muscicola* [SM000086S23019 and SM000015S01186] and one gene in *Mesotaenium endlicherianum* [ME000101S10682]) (**Figure S1E**). Although *M. polymorpha* has eight GARP-G2 genes, we identified *Mp7g*09740 (annotated as Mp*GARP8*) as a single ortholog of *GLK* based on the presence of the characteristic C-terminal GOLDEN2 C-terminal (GCT) box (Fitter et al., 2002) (**Figure S1D,E and Dataset S1**). In summary, our analysis indicates that *M. polymorpha* contains single orthologs of *GNC/CGA1*, *GATA2*, *SCL* and *GLK* that we hereafter refer to as Mp*GATA4*, Mp*GATA2*, Mp*SCL* and Mp*GLK* respectively.

Mp*glk* mutants show compromised chlorophyll accumulation

To test whether Mp*GATA4*, Mp*GATA2*, the Mp*SCL-MIR171* module, Mp*HY5* and Mp*GLK* orthologs control chloroplast biogenesis in *M. polymorpha* we used CRISPR/Cas9 editing to generate knockout mutant alleles for each. For Mp*GATA4* two guide RNAs (gRNAs) were used to target the second and third exons, for Mp*GATA2* one gRNA targeted the second exon, for Mp*SCL* two gRNAs targeted the middle of its single exon, and for Mp*MIR171* two gRNAs flanking the MIR171 gene were designed (**Figure 2A**). For Mp*HY5* two gRNAs were designed to target the third exon (**Figure 2A**). Finally, for Mp*GLK* a single gRNA was designed to target the sixth exon (**Figure**

2A). Plants transformed with the same vector but without a gRNA sequence were used as 'no guide RNA' controls. Each mutant line was clonally propagated through one gemmae generation to obtain isogenic plants, and for each targeted gene three independent lines were selected for analysis. In all protein coding genes, mutations led to premature stop codons (**Figure S4**) and Mp*mir171* mutants had a 139 bp deletion and 1 bp insertion resulting in a deletion of the entire Mp*MIR171* gene (**Figure S4D**).

Next, we analysed the phenotypes of the different mutants. In Mp*gata4* mutants, gametophyte morphology was perturbed with narrower thallus lobes (**Figure 2B, C**). However, chlorophyll levels were only ~10% higher in one out of the three lines we examined, compared to control plants (**Figure 2J**), which is the opposite of an expected reduction in the case of functional conservation. We analysed chloroplast morphology in the central part of the gemmae of Mp*gata4* mutant lines via confocal laser scanning microscopy (**Figure 2I**). In Mp*gata4* mutants, chloroplast size in cells of the central part of gemmae, excluding the rhizoid precursors, was similar to that of control plants (**Figure 2C, R**). Mp*gata2* mutants did not show any morphological or developmental phenotypes and had similar chlorophyll levels compared with 'no guide RNA' controls (**Figure 2B, D, K**). Chloroplast size in the Mp*gata2* mutant was also similar with 'no guide RNA' controls (**Figure 2D, R**). Mp*scl* mutants showed altered thallus morphology, with thalli being 'stunted' with inward curling edges (**Figure 2B, E**). Despite chlorophyll levels being not statistically different from controls (**Figure 2L**), in two of the three Mp*scl* mutant lines we observed an increase in the size of the chloroplasts, which in some cases was statistically significant (**Figure 2E, R**). In contrast, Mp*mir171* mutants were indistinguishable from control plants (**Figure 2B, F, M, R**). Mp*hy5* gametophyte morphology was also perturbed with thallus lobes being erected (**Figure 2G**). Chloroplast size was ~15% smaller in two Mp*hy5* lines, when grown at 22 ^{o}C, compared to controls (**Figure 2G, R**). In *A. thaliana*, HY5 regulates photosynthetic capacity at cooler temperatures, and *hy5* mutant plants have a conditional phenotype: plants grown at 17 ^{o}C show a ~25-30% reduction in chlorophyll levels compared to plants grown at 22 ^{o}C (Toledo-Ortiz et al., 2014). To test whether HY5 also has a conditional phenotype in *M. polymorpha*, we grew mutant and control plants under three different temperatures, 22 ^{o}C, 17 ^{o}C, and 12 ^{o}C. Opposite to *A. thaliana*, Mp*hy5* mutants grown at 22 ^{o}C and 17 ^{o}C had similar chlorophyll levels

to wild type controls (**Figure 2G, O, P**), however, Mp*hy5* mutants grown at 12 ^{o}C showed a ~10-15% reduction in chlorophyll levels compared to controls (**Figure 2G, P**), suggesting a conditional phenotype but at a lower temperature Interestingly, we found that in Mp*hy5* mutants the content of several sesquiterpenes secondary metabolites was increased compared to controls (**Dataset S2**). This implies that in addition to its previously reported function in regulating flavonoid biosynthesis (Clayton et al., 2018), *M. polymorpha* HY5 is likely to have a role in regulating terpenoids.

Finally, Mp*glk* mutants had an obvious pale green phenotype (**Figure 2B,**). Mutations that either caused a deletion of the GARP DNA-binding domain or introduced a premature stop codon immediately upstream of it, resulted in a more severe phenotype (we refer to these as 'strong' alleles) than mutations in the 5´ end of the Mp*GLK* gene ('weak' alleles) (**Figure S5A-F**). In the 'strong' Mp*glk* mutant alleles, chlorophyll content was reduced by ~90% compared to the controls and chloroplasts were also reduced in size (**Figure 2H, Q, R**). Mp*glk* mutants also showed morphological changes with narrower thallus lobes and upward curling lobe edges (**Figure 2H**). However, they were still able to grow and produce gemmae and reproductive organs (**Figure S5G**) without the need for supplemental carbon sources.

When chloroplast ultrastructure was assessed via electron microscopy, consistent with confocal laser scanning microscopy, Mp*glk* mutants had smaller chloroplasts compared with controls and showed perturbed ultrastructure (**Figure 3A, B**). Specifically, chloroplasts had fewer thylakoid membranes that often did not occupy the full chloroplast area and had reduced granal stacking (**Figure 3A, B and Figure S5H**). Treatment with di-chlorophenyl di-methyl urea (DCMU) which inhibits the photosynthetic electron transport chain (Trebst, 2007) caused a substantial reduction in the chlorophyll fluorescence parameter F_v/F_m in the mutants indicating that their photosynthetic apparatus was functional (**Figure 3C, D**). This is consistent with the ability of 'strong' Mp*glk* alleles to grow under standard conditions without a carbon supplement.

In summary, the absence of functional Mp*GATA4* and Mp*GATA2* genes did not appear to affect chlorophyll biosynthesis or chloroplast biogenesis in *M. polymorpha*. As *GNC/CGA1* mutants of *A. thaliana* contain lower chlorophyll levels (Zubo et al., 2018), our data argue against functional conservation between *A. thaliana* and *M. polymorpha* GNC/CGA1 and Mp*GATA4* in regulating

chloroplast biogenesis. Similarly, we found a lack of conservation between the function of At*GATA2* and Mp*GATA2*. Moreover, despite conservation of a MIR171 target site in the Mp*SCL* mRNA, MpSCL does not appear to control chlorophyll content. Again, this is in contrasts with *A. thaliana* where triple At*scl6* At*scl22* At*scl27* mutants show increased chlorophyll levels and At*MIR171* mimic lines which are functional equivalents of At*mir171* knockouts have reduced chlorophyll levels (Ma et al., 2014; Todesco et al., 2010). Although chloroplast sizes were unaltered in Mpg*ata2*, Mp*gata4* or Mp*mir171,* we found that two out of three Mp*scl* lines had larger chloroplasts. Mp*hy5* mutants did not appear to affect chlorophyll biosynthesis but showed a conditional effect with lower chlorophyll levels when grown under a colder temperature. Absence of a functional Mp*HY5* gene resulted in plants with smaller chloroplasts in two out of the three mutant lines we examined. In summary, in this study, we found that MpGLK was the only gene that affected chlorophyll levels under standard growth conditions. Mp*glk* mutants accumulated less chlorophyll and had smaller chloroplasts than the wild type indicating that the Mp*GLK* ortholog is required for proper chloroplast biogenesis. Collectively our results imply a conserved function for GLK in chloroplast biogenesis in *M. polymorpha* and suggest that the conditional role of HY5 may be also conserved. Observed changes in chloroplast sizes in Mp*scl* and Mp*hy5* raises the question for their conserved roles in chloroplast biogenesis and photomorphogenesis.

No evidence of functional redundancy between MpGATA4, MpGATA2, MpSCL, MpHY5 and MpGLK

It is possible that MpGLK compensates for the lack of functional MpGATA4, MpGATA2, MpSCL and MpHY5 proteins in respective mutants and so redundancy in the gene regulatory networks masks their loss of function. To test whether Mp*GATA4*, Mp*GATA2,* Mp*SCL* and Mp*HY5* act redundantly with Mp*GLK,* we generated Mp*glk,gata4*, Mp*glk,gata2,* Mp*glk,scl* and Mp*glk,hy5* double mutants (**Figure 4A-E and Figure S5I**). Double mutants were not noticeably paler than the single Mp*glk* mutant (**Figure 4A-E**) and contained similar levels of chlorophyll (**Figure 4F**). In Mp*glk,gata4,* Mp*glk,scl* and Mp*glk,hy5* double mutants, thallus morphology was perturbed in a similar manner to single Mp*gata4,* Mp*scl* and Mp*hy5* mutants. Chloroplast sizes of double

Mp*glk,gata4* mutants was also comparable to those of single Mp*glk* mutants (**Figure 4G**). Further support that Mp*GATA4* and Mp*GLK* do not act redundantly is provided by the fact that overexpression of MpGATA4 in Mp*glk* plants cannot rescue Mp*glk* mutant phenotype (**Figure 4H and S5J**). We next asked whether, similarly to *A. thaliana*, HY5 regulates the expression of Mp*GLK* and Mp*GATA4* in *M. polymorpha*. To test this, we subjected 6-day-old Mp*hy5* and control gemmalings to 48 hours of darkness, followed by exposure to 1.5, 3 and 6 hours of light at 22°C. We found that Mp*GLK* and Mp*GATA4* expression was comparable in both Mp*hy5* and the control, suggesting that Mp*GLK* and Mp*GATA4* are unlikely to be targeted by MpHY5 (**Figure 4I**).

In summary, our analysis does not support the presence of functional redundancy between Mp*GATA4*, Mp*GATA2,* Mp*SCL*, Mp*HY5* and Mp*GLK.*

Mp*GLK* is sufficient to activate chloroplast biogenesis.

To determine whether homologs of Mp*GATA4,* Mp*GATA2,* Mp*SCL* and Mp*GLK* are sufficient to activate chloroplast biogenesis, we generated overexpression lines driven by the strong constitutive ubiquitin-conjugating enzyme E2 gene (Mp*UBE2*) (Sauret-Güeto et al., 2020). To facilitate analysis, GFP was used to mark the plasma membrane (**Figure 5A**). No differences in chlorophyll levels or morphology between plants with and without the plasma membrane marker were evident (**Figure S5K**). qPCR analysis confirmed that each transgene was over-expressed (**Figure S5L**). No differences in chlorophyll content or chloroplast size were detected between controls and overexpressing lines of Mp*GATA4,* Mp*GATA2* or Mp*SCL* (**Figure 5B-F, I, J**). Since in *A. thaliana, GATA2* over-expression led to constitutive photomorphogenesis (Luo et al., 2010) we placed gemmae overexpressing Mp*GATA2* in the dark alongside controls, and examined if there were any differences in growth after a week. Both over-expressors and controls looked indistinguishable. However, plants in which Mp*GLK* was overexpressed were darker green, and displayed stunted growth (**Figure 5G**). Lines expressing Mp*GLK* contained up to ~4 times higher chlorophyll and significantly increased chloroplast size (**Figure 5I, J**). Interestingly, we observed an increase in chloroplast size in normally non-photosynthetic rhizoid precursor and oil body cells compared with controls (**Figure 5H**). We conclude that only Mp*GLK* is sufficient to activate chloroplast biogenesis.

In angiosperms, GLK is subjected to post-transcriptional regulatory mechanisms (P. Wang et al., 2017). For example, overexpression of rice *GLK* in rice is not maintained beyond the seedling stage(Nakamura et al., 2009). To test this, we generated two additional constructs. The first construct was designed to test whether Mp*GLK* coding sequence contains motifs targeting its mRNA for post-transcriptional downregulation and so the nucleotide sequence was mutated but the amino acid sequence preserved (**File S1**). We hypothesised that this would remove nucleotide sequence motifs targeting Mp*GLK* for post-transcriptional degradation. The resulting 're-written' coding sequence was named Mp*GLK-CDSrw*. In the second construct the native Mp*GLK* coding sequence was fused to the Mp*GLK* 3´ UTR sequence (Mp*GLK-CDS+3´UTR*) to test whether this would reduce greening. Mp*GLK-CDSrw* overexpression had a similar effect to Mp*GLK* overexpression (**Figure S6A-E**). In contrast, plants expressing Mp*GLK-CDS+3´UTR* showed intermediate chlorophyll accumulation between controls and those expressing Mp*GLK,* the chloroplast size was unaffected (**Figure S6A-E**). To understand which part of Mp*GLK* 3´UTR negatively regulates greening we tested a series of truncated Mp*GLK* 3´UTR versions covering positions 1-100, 1-200, 1-300, 1-400, 1-500 and 1-610 as well as the full length 671 bp sequence (**Figure S6F**). The 1-200 bp truncated version contained a predicted miR11666.4 target site located at positions 105-124 and an additional putative miRNA/siRNA cleavage site located between positions 135 and 136 (Lin & Bowman, 2018). We then fused the truncated versions of Mp*GLK* 3´UTR to Mp*GLK* and overexpressed these constructs under control of the M*pUBE2* promoter. Only the 1-500 bp and 1-610 bp fragments of the Mp*GLK* 3´UTR had the same negative effect on Mp*GLK*-mediated greening as the full-length Mp*GLK* 3´UTR (**Figure S6G**). This suggests a yet unidentified regulatory motif(s) between nucleotides 400-671 of the 3´UTR responsible for Mp*GLK* downregulation. We conclude that the Mp*GLK* 3´ UTR reduces greening that is potentiated by overexpression of Mp*GLK* in *M. polymorpha.*

GLK in *M. polymorpha* regulates thylakoid associated photosynthetic components and chlorophyll biosynthesis

To identify genes that respond to MpSCL, MpGATA4, MpHY5 and MpGLK as well as the extent of their target overlap, we performed RNA sequencing (RNA-seq) in Mp*scl,* Mp*gata4, Mphy5* and

Mp*glk* mutants. We excluded Mp*GATA2* from our RNA-seq analysis because our prior investigation revealed no evidence for a role in chloroplast biogenesis. Knocking out Mp*SCL*, Mp*GATA4* and Mp*HY5* had a moderate effect with 243, 332 and 160 genes being downregulated respectively (*padj*-value =<0.01 test, LFC ≥ 1-fold) (**Figure S7A-C, Dataset S3**). In loss of function mutants for Mp*GLK,* 1065 genes showed a reduction in transcript abundance compared with controls (*padj*-value ≤0.01, LFC ≥ 1-fold) (**Figure S6D, Dataset S3**). Changes to transcript abundance that were common between genotypes were limited, with the largest overlap (85 genes) being detected for Mp*glk* and Mp*gata4* mutants (**Figure 6A**). In all three genotypes Gene Ontology (Kawamura et al., 2022) analysis showed that oxidative stress as well as hydrogen peroxide catabolic processes were most impacted (**Figure 6B**), but in Mp*glk* mutants, photosynthesis and chlorophyll biosynthesis process were also affected (**Figure 6B**). Analysis of the effects of loss of Mp*SCL,* Mp*GATA4,* Mp*HY5* and Mp*GLK* function on chlorophyll biosynthesis and canonical photosynthesis genes confirmed the greatest effect in Mp*glk*. For example, in Mp*scl* mutants only one chlorophyll biosynthesis gene, *CAO*, was upregulated (Log2FC~0.58). In Mp*hy5* and Mp*gata4* mutants, there is a moderate reduction in transcript abundance (Log2FC~-0.3-0.7) of 13 and five chlorophyll biosynthesis genes respectively (**Figure 6C**). In contrast, Mp*GLK* mis-expression affected 17 of the 19 chlorophyll biosynthesis genes, with Log2FC>=-1 for 15 genes.

We next checked the effects of Mp*scl*, Mp*gata4,* Mp*hy5* and Mp*glk* on the total of 68 annotated genes belonging to the light harvesting complexes A and B (LHCA and LHCB), complements of photosystem I and II (PSI and PSII) and cytochrome b_6f complex genes. Mp*SCL* loss led to the upregulation of only two genes (Log2FC~-0.37-0.55). Mp*HY5* or Mp*GATA4* loss resulted in a moderate reduction in the transcript levels (Log2FC~-0.4-0.7) of 6 and 30 of the total 68 annotated genes, respectively. The only exceptions were three and four LHCB encoding genes (Log2FC>-1) for Mp*gata4* and Mp*hy5* mutants respectively (**Figure 6D and Figure S7F**). Consistent with the effects on chlorophyll that we described above, the Mp*glk* mutant had the greatest effect on transcript abundance (**Figure 6D,E and Figure S7G**) with 33 of the total 68 genes downregulated (Log2FC>=-1).

In addition to its function in chloroplast biogenesis, HY5 is known to regulate cell elongation and proliferation, pigment accumulation as well as nutrient uptake (Xiao et al., 2021). When we examined whether homologs of genes known to control such processes in *A. thaliana* are mis-regulated in Mp*hy5* mutants, we could only observe an upregulation of a Sulphate Transporter (**Figure S7F**), which is opposite to what was observed in *A. thaliana*.

In conclusion, our RNA-seq analysis indicated that MpSCL, MpHY5 and MpGATA4 have the ability to control expression of a limited number of photosynthesis genes. However, compared with GLK, their reach is very limited and does not appear to be sufficiently extensive to have an impact on a phenotypic level.

ChIP-seq analysis uncovered a distinct and divergent binding pattern of MpGLK in comparison to flowering plants

To identify the targets of MpGLK, we performed ChIP-seq analysis using transgenic *M. polymorpha* that expressed the MpGLK fused to hemagglutinin (HA) tag (**Figure 7A**). Applying the data analysis pipeline utilised for the five flowering plant species (*A. thaliana*, tomato, rice, maize and tobacco) GLK ChIP-seq comparison (Tu et al., 2022), we detected a total of 2,654 reproducible peaks (IDR <0.01). Motif enrichment analysis of the summit regions of ChIP-seq peaks showed that MpGLK binds to a RGATTYY sequence motif, analogous to its orthologs in flowering plants (**Figure 7B**). Next, we associated these peaks with the 1.5 kb upstream region of annotated genes and found 1,326 GLK target genes (**Dataset S4**). In addition, we performed assay for transposase-accessible chromatin with sequencing (ATAC-seq) to identify the promoter accessible chromatin regions and carried out ChIP-seq for H3K4me3, a histone mark often associated with active genes near their transcription start site. We found that the MpGLK ChIP-seq signal overlaps with the ATAC-seq signal (**Figure 7A and D**), suggesting that MpGLK binds to the open chromatin regions. Furthermore, the binding sites are also found upstream of the +1 nucleosome that are often marked by H3K4me3, a pattern consistent with the observation for *A. thaliana* GLK (**Figure 7A and C**).

Comparing ChIP-seq data with the gene expression profile of Mp*glk* mutants, we found that 40.42% (536/1326) of the GLK target genes identified via ChIP-seq are down-regulated in the

Mp*glk* mutant (**Figure 7E**), while 35% (471/1326) of GLK target genes are up-regulated in the Mp*GLK* over-expression lines. Of those affected in Mp*glk* mutants and Mp*GLK* over-expression lines, 269 genes overlapped. Similarly, to *A. thaliana* and tomato, the average binding strength of MpGLK, as inferred from the ChIP-seq signal fold-change at the peak summit, was found to be higher in differentially expressed genes (DEGs) compared to non-DEGs (**Figure 7F**). We found that more MpGLK target genes showed differential expression in the Mp*glk* mutants compared to their counterparts in flowering plants. For example, only 21% (202/960) of the AtGLK1/2 targets are DEGs in the At*glk1/glk2* mutant. This is consistent with a more pronounced chlorophyll loss in the Mp*glk* mutants that we describe in this study compared to the published At*glk1/glk2* alleles (Waters et al., 2009). Alternatively, this may point to the evolutionary changes that occurred in flowering plants compared to the liverworts. In summary, our findings suggest a conserved role of GLK throughout the evolution of land plants but do not exclude that the transcriptional regulatory networks in flowering plants have accumulated changes resulting in increased redundancy and robustness.

The dynamic nature of cis-regulatory elements in the plant genomes can give rise to divergence of TF binding. When comparing GLK binding across flowering plants such as *A. thaliana,* tomato, tobacco, rice and maize, it was found that only 10-20% of their target genes are conserved among species. However, it is worth noting that most of the conserved target genes are associated with photosynthesis (Tu et al., 2022). To investigate this further for MpGLK, OrthoFinder was employed to classify protein-coding genes in *M. polymorpha* and the aforementioned five species into ortholog groups. Only 53 GLK target genes in 48 ortholog groups were found to be conserved, and they are indeed mainly involved in photosynthesis (**Dataset S4**). For instance, 23 genes encoding CHLOROPHYLL A/B BINDING PROTEIN are all bound by MpGLK. Additionally, GLK was found to bind to CONSTANS and B-BOX transcription factor genes in all six species (**Figure 7G**), suggesting that they played an ancient regulatory role in circadian rhythm and the photomorphogenesis. However, it is important to highlight that most MpGLK targets are not conserved. For instance, *1-deoxy-D-xylulose-5-phosphate reductoisomerase* and *Pyruvate,orthophosphate (Pi) dikinase*, which are GLK targets in liverworts, do not have orthologs in *A. thaliana* that are targeted by AtGLK1/2 (**Figure 7A and C**). Even the MpGLK target genes

encoding chloroplast-located proteins do not exhibit conservation across these plant species (**Figure 7H**). These findings provide further support to our hypothesis that the transcriptional regulatory mechanism has diverged during the over 500 million years since the split between bryophyte and flowering plants.

Discussion

Loss of MpGATAs, MpHY5 or MpSCL has a minimal impact on chloroplast biogenesis

The monophyletic group of bryophytes diverged from vascular plants approximately 400-500 MYA (Harris et al., 2022; Li et al., 2020; One Thousand Plant Transcriptomes Initiative, 2019). 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). In *A. thaliana GNC* over-expression increases chlorophyll tenfold in seedlings and ~30% in the leaf (Chiang et al., 2012) and *GATA2* over-expression led to constitutive photomorphogenesis (Luo et al., 2010), The simplest and most parsimonious explanation for the lack of impact of Mp*GATA4* or Mp*GATA2* loss or overexpression on chlorophyll content in bryophytes such as *M. polymorpha* is that these proteins do not play a role in chloroplast biogenesis, and that this represents the ancestral state. It is also possible that other proteins compensate and so redundancy in the gene regulatory networks masks loss of function. However, Mp*glk* and Mp*gata4* or Mp*gata2* double mutants had similar chlorophyll levels with Mp*glk* single mutants, implying lack of redundancy, at least between MpGLK and GATAs. Another possibility is that GATA transcription factors regulate chloroplast biogenesis ancestrally in some but not all bryophytes. This possibility is supported by some limited analysis of Pp*GATA1,* one of the two *GNC/CGA1* orthologs in *P. patens*, where two Pp*GATA1* overexpression lines showed a ~10-20% increase in chlorophyll content (Luan et al., 2023). A similar effect was also reported when Pp*GATA1* was mis-expressed in *A. thaliana* (Luan et al., 2023). However, these increases in chlorophyll content induced by Pp*GATA1* are modest compared with those reported in *A. thaliana* where *GNC* over-expression increased chlorophyll tenfold in seedlings and ~30% in the leaf (Chiang et al., 2012). Moreover, while *GNC/CGA1*

overexpression in *A. thaliana* led to an increase in chloroplast number per cell, this was not reported in *P. patens* (Luan et al., 2023). The role of *PpGATA* orthologs (including both Pp*GATA1* and Pp*GATA2*), therefore, needs additional analysis to confirm or refute their role in chloroplast biogenesis. A recent study of MpGATA4 in *M. polymorpha* proposed a role in response to high-light stress (Schröder et al., 2023) with *Early Light Induced Protein* (*ELIP*) encoding genes being mis-regulated in Mp*gata4* mutants. Our transcriptome analysis of Mp*gata4* mutants revealed that only a single *ELIP* gene (Mp*ELIP-11*) is moderately downregulated (Log2FC=-0.678) alongside a small number of genes encoding components of the LHC and PS. We also noted that editing Mp*GATA4* had an effect on gametophyte morphology such that thalli of Mp*gata4* mutants had narrower lobes. This contrasts with *A. thaliana* where to our knowledge there have been no reports of perturbations to leaf morphology in *gnc*/*cga1* mutants. It is therefore possible that the role of this protein has been repurposed from one in affecting development of photosynthetic tissue and response to high light in *M. polymorpha*, to one in regulating chloroplast biogenesis in *A. thaliana*.

Similarly, neither Mp*scl* or Mp*mir171* mutants exhibited any detectable alterations in chlorophyll accumulation as would be expected if their functions were conserved between bryophytes and *A. thaliana.* For example, At*scl6* At*scl22* At*scl27* triple mutants and *MIR171* over-expressors lead to increased chlorophyll accumulation in *A. thaliana* (L. Wang et al., 2010). As no other members of the *SCL* gene family in *M. polymorpha* contain a MIR171 recognition site, this argues against Mp*MIR171* playing a role in the regulation of chlorophyll accumulation in this liverwort. It is possible that the *M. polymorpha* homolog of At*SCL* does regulate chlorophyll biosynthesis but without the additional control derived from MIR171 that operates in *A. thaliana*. SCL6, SCL22 and SCL27 also regulate proliferation of meristematic cells (Bolle, 2004; Llave et al., 2002; Rhoades et al., 2002). If that was the case, we would expect that over-expression SCL to cause a decrease in chlorophyll content as seen in *A. thaliana* (Ma et al., 2014). However, our SCL over-expression lines looked similar to control plants. A similar role in *M. polymorpha* could explain the perturbed thallus morphology with inward curling edges of Mp*scl* mutants. Thus, in addition to GATA2 and GATA4 discussion above, it is also possible that an ancestral role of SCL

relates to development of photosynthetic tissue. For SCL6, SCL22 and SCL27 this function seems to be retained in *A. thaliana*, and a role in repressing chlorophyll synthesis has been acquired.

We found limited evidence that HY5 in *M. polymorpha* plays a role in chloroplast biogenesis similar to that observed in *A. thaliana*. Chlorophyll levels were comparable to controls when plants were grown at 17 ^{o}C compared to the ~30% reduction found in *A. thaliana*. We were able to detect a minor ~10-15% reduction in Mp*hy5* chlorophyll levels only when plants were grown at 12 ^{o}C. Transcriptome and qPCR analysis of Mp*hy5* mutants also failed to identify a clear role of HY5 in the regulation of chloroplast biogenesis and photosynthesis associated genes. HY5 is also known to play a role in the response to Ultraviolet B, a major abiotic stress in land plants. In *A. thaliana* UVB activates the UV RESISTANCE LOCUS8 (UVR8) photoreceptor. UVR8 then enhances HY5 activity by preventing its degradation which is triggered by E3 ubiquitin ligase CONSTITUTIVELY PHOTOMORPHOGENIC1 (COP1) mediated ubiquitination. Consequently, HY5 regulates the transcription of numerous genes required for the UVB photomorphogenic response (Binkert et al., 2014; Ulm et al., 2004), including for DNA repair, protection of photosynthetic machinery and production of UVB-absorbing phenylpropanoids. In *M. polymorpha* UVB response included the production of UVB-absorbing flavonoids and the induction of HY5, similar to *A. thaliana* (Clayton et al., 2018). Thus, it appears that HY5 in *M. polymorpha* still responds to light signals, however in flowering plants it is part of a more complex regulatory network that integrates a broader set of light signals with chloroplast biogenesis. MpHY5 is likely to play an additional role in secondary metabolism based on our finding that Mp*hy5* mutants accumulate extra levels of sesquiterpenes.

Loss of GLK limits chloroplast biogenesis in *M. polymorpha*

Unlike *GATAs* and the *MIR171-SCL* module, three lines of evidence indicate that GLK function is conserved in *M. polymorpha*. Firstly, Mp*glk* mutants have significantly reduced chlorophyll accumulation and smaller chloroplasts with underdeveloped thylakoids. These perturbations to phenotype have been observed in other land plants (Bravo-Garcia et al., 2009; Fitter et al., 2002; Rossini et al., 2001; Waters et al., 2008; Yasumura et al., 2005). Secondly, constitutive overexpression of Mp*GLK* resulted in increased chlorophyll accumulation, increased chloroplast

size and ectopic chloroplast development similar to reports in angiosperms (Kobayashi et al., 2013; Nakamura et al., 2009; Powell et al., 2012; P. Wang et al., 2017; Zubo et al., 2018). Thirdly, our transcriptome analysis of Mp*GLK* overexpression and mutant lines revealed an overlap between *GLK* dependent genes in *M. polymorpha* and *A. thaliana* but also rice, tobacco, tomato and maize (Tu et al., 2022; Waters et al., 2009). ChIP-seq confirmed that most of them are photosynthesis-related genes directly bound by GLK such as the *LHCA*, *LHCB*, *PsbQ* and genes encoding chlorophyll biosynthesis enzymes (Tu et al., 2022). However, we also uncovered a major divergence of GLK binding pattern during land plant evolution, as less than 5% of the MpGLK target genes are conserved in flowering plants.

Finally, and in contrast to previous reports, our phylogenetic analysis identified *GLK* orthologs in two Zygnematophyceae green algae for which genome assemblies became recently available (Cheng et al., 2019). Taken together our results suggest that the GLK function is ancestral to all land plants and appears to have evolved before the transition of plants to terrestrial ecosystems.

In summary, our results suggest that the regulation of photosynthesis gene expression is more streamlined in *M. polymorpha*. It is possible that the low levels of genetic redundancy in regulators of photosynthesis in this species are associated with low rates of photosynthesis and limited specialisation within the thallus. Angiosperms on the other hand have a more complex development and morphology allowing colonisation of a broad range of environments. For example, leaves with specialised tissues allow high rates of photosynthesis. As a result, greater control over photosynthesis may have become necessary compared with bryophytes and could have been mediated by elaborations and specialisations to pre-existing pathways present in the common ancestor of bryophytes and vascular plants. The simplified genetic network in *M. polymorpha* either represents the ancestral state, or a simplified version due to secondary loss. For example, such secondary loss of traits has been reported for *M. polymorpha* stomata and their regulatory genetic network (Harris et al., 2020; Rich & Delaux, 2020). To elucidate the ancestral function of GLK further detailed genetic studies in other non-seed plants will be necessary. Furthermore, our finding of GLK homologs in the green algal lineage Zygnematophyceae, provides an exciting opportunity to test whether its role in photosynthesis predates the colonisation of land.

Materials & Methods

Phylogenetic analysis

To identify GATA B-Class and A-Class genes, three different approaches were combined: Firstly, the GATA protein sequences for 21 plant genomes were mined from the iTAK (Zheng et al., 2016) and PlantTFDB databases (Jin et al., 2017), Phytozome, Fernbase, Phycozome and PhytoPlaza. Sequences for each individual species were aligned with the AtGNC and AtCGA1 amino acid sequences using MAFFT (Katoh et al., 2019). Results were filtered manually to identify GNC/CGA1 (B-Class) orthologs distinguished from other GATA family genes by the presence of conserved serine (S) residue, a conserved IRX(R/K)K motif (I: Isoleucine, R: Arginine, X: any amino acid and K: Lysine), and the presence or absence of conserved LLM- (leucine– leucine–methionine) domain at their C-terminus (Behringer & Schwechheimer, 2015). GATA2 (A-Class) orthologs, were distinguished by the presence of a conserved glutamine (Q) and a threonine (T) within the zinc finger motif. Secondly, we performed BLASTP searches against plant genomes on Phytozome v13, fern genomes (fernbase.org) (Li et al., 2018), hornworts genome (www.hornworts.uzh.ch) (Li et al., 2020), green algae genomes on PhycoCosm (phycocosm.jgi.doe.gov), and 1KP using the AtGNC/CGA1 amino acid sequence as a query. Results were filtered manually as described above. Finally, the combined results from the above to approaches were checked against Orthofinder searches (Emms & Kelly, 2019). The identified GATA protein sequences were aligned using MAFFT. Alignments were then trimmed using TrimAl (Capella-Gutiérrez et al., 2009). A maximum likelihood phylogenetic tree was inferred using iQTree (Nguyen et al., 2015), ModelFinder (Kalyaanamoorthy et al., 2017) and ultrafast approximation for phylogenetic bootstrap (Hoang et al., 2018) and SH-aLRT test (Guindon et al., 2010). The tree was visualised using iTOL (Letunic & Bork, 2021).

To identify *GLK* genes, three different approaches were combined: Firstly, the G2-GARP protein sequences for 21 plant genomes were mined from the iTAK and PlantTFDB databases, Phytozome, Fernbase, Phycozome and PhytoPlaza. Sequences for each individual species were aligned with the AtGLK1/2 amino acid sequences using MAFFT. Results were filtered manually to identify GLK orthologs distinguished from other G2-GARP family genes by three characteristic motifs: AREAEAA motif (consensus motif) in the DNA-binding domain, VWG(Y/H)P and the

PLGL(R/K)(P/S)P in the GCT-box domain. Secondly, we performed BLASTP searches against plant genomes on Phytozome v13, fern genomes (fernbase.org), hornworts genome (hwww.hornworts.uzh.ch), green algae genomes on PhycoCosm (/phycocosm.jgi.doe.gov), and 1KP (One Thousand Plant Transcriptomes Initiative, 2019) using the AtGLK1/2 amino acid sequence as a query. Results were filtered manually as described above. Finally, the combined results from the above two approaches were checked against Orthofinder searches. The identified GLK protein sequences were aligned using MAFFT. Alignments were then trimmed using TrimAl. A maximum likelihood phylogenetic tree was inferred using iQTree, ModelFinder and ultrafast approximation for phylogenetic bootstrap and SH-aLRT test . The tree was visualised using iTOL .

Plant growth, transformation, CRISPR/Cas9 gene editing and overexpression construct generation

Marchantia polymorpha accessions Cam-1 (male) and Cam-2 (female) were used in this study (Delmans et al., 2017) with the exception of ChIP-seq and ATAC-seq for which *Marchantia polymorpha* accessions Tak-1 (male) and Tak-2 (female) accessions were used. Plants were grown on half strength Gamborg B5 medium plus vitamins #G0210, Duchefa Biochemie) and 1.2% (w/v) agar (#A20021, Melford), under continuous light at 22 °C with light intensity of 100 µmol m^{-2} s^{-1}. Transgenic *M. polymorpha* plants were obtained following an established protocol (Sauret-Güeto et al., 2020).

For CRISPR/Cas9 gene editing, guide RNAs were *in silico* predicted using CasFinder tool (https://marchantia.info/tools/casfinder/). Several gRNAs per target gene were *in vitro* tested as described (Yelina et al., 2022) using oligonucleotides in **Dataset S5**. gRNA sequences that were selected to generate Mp*glk*, Mp*gata2*, Mp*gata4*, Mp*scl* and Mp*mir171* mutants are listed in **Dataset S5**. Single gRNA7 to mutate Mp*GLK* was cloned using OpenPlant toolkit (Sauret-Güeto et al., 2020). Multiple gRNAs to mutate Mp*GATA4*, Mp*SCL* and Mp*MIR171* and a single gRNA to mutate Mp*GATA2* were cloned as described in (Yelina et al., 2022), using oligonucleotides listed in **Dataset S5** and the destination vector pMpGE013 (Sugano et al., 2018). For the overexpression constructs, Mp*GLK and* Mp*GLKrw* CDS were synthesised (Integrated DNA Technologies), Mp*GLK*

3´UTR was amplified from *M. polymorpha* genomic DNA and cloned into the pUAP4 vector (Sauret-Güeto et al. 2020). Constructs were generated using the OpenPlant toolkit (Sauret-Güeto et al., 2020). OpenPlant parts used: OP-023 CDS12-eGFP, OP-020 CDS_hph, OP-037 CTAG_Lti6b, OP-054 3TER, _Nos-35S, OP-049 PROM_35S, OP-47 PROM_MpUBE2, OP-48 5UTR_MpUBE2, OP-063, L1_HyR-Ck1, OP073 L1_Cas9-Ck4, OP-076 L2_lacZgRNA-Cas9-CsA and L1_35S_s::eGFP-Lti6b (Waller et al., 2023). Nucleotide sequence of Mp*GLKrw* CDS and oligonucleotide sequences used for cloning are listed in **Dataset S5**.

Chlorophyll determination, fluorescence measurements and imaging analysis

For chlorophyll measurements, ~30-50mg of 10-14 days old gemmalings were used, with five biological replicates per genotype. The tissue was blotted on tissue paper before weighing to remove excess water and then was transferred into a 1.5mL Eppendorf tube containing 1 mL of di-methyl sulfoxide (DMSO) (#D8418, Sigma Aldrich) and incubated in the dark at 65C with 300 rpm shaking for 45 minutes. Samples were let to cool down to room temperature for approximately one hour. Chlorophyll content was measured using a NanoDrop™One/One C Microvolume UV-Vis Spectrophotometer (ThermoFisher) following the manufacturer's protocol. Chlorophyll fluorescence measurements were carried out using a CF imager (Technologica Ltd, UK) and the image processing software provided by the manufacturer, as described previously (Schreier et al., 2024). *M. polymorpha* plants were placed in the dark for 20 min for dark adaptation to evaluate the dark-adapted minimum fluorescence (Fo), dark-adapted maximum fluorescence (*Fm*), variable fluorescence *Fv* (*Fv = Fm–Fo*). All chlorophyll fluorescence images within each experiment were acquired at the same time in a single image, measuring a total of three plants per genotype and treatment. For the DCMU treatment, 20 µM DCMU (#45463, Sigma Aldrich) was added to the half-strength MS media before it was poured into the individual petri dishes. Thalli were placed for 24 h onto the DCMU-containing media before chlorophyll fluorescence measurements.

For imaging a gene frame (#AB0576, ThermoFisher) was positioned on a glass slide. Five to seven gemma were placed within the medium-filled gene frame together with 30 µL of milliQ water. The frame was then sealed with a cover slip. Plants were imaged immediately using a Leica

SP8X spectral fluorescent confocal microscope. Imaging was conducted using either a 10× air objective (HC PL APO 10×/0.40 CS2) or a 20× air objective (HC PL APO 20×/0.75 CS2). Excitation laser wavelength and captured emitted fluorescence wavelength window were as follows: for eGFP (488 nm, 498−516 nm) and for chlorophyll autofluorescence (488 or 515nm, 670−700 nm). Chloroplast area was measured using ImageJ and the Macro in **Supplemental Information**.

For electron microscopy sections (~2 mm^2) of 5-6 individual 3-week-old *M. polymorpha* thallus per genotype were harvested; and then fixed, embedded and imaged using scanning electron microscopy as previously described (Schreier et al., 2024).

Extraction of Terpenes and analysis on gas chromatograph - mass spectrometer

Approximately 200 mg of frozen plant material of *M. polymorpha* were extracted with 1 mL of cold methanol containing 5 mM of NaCl to quench enzymatic activities (Kumar et al., 2016) and 900 ng of dodecane as internal standard (#297879, Sigma Aldrich). Mechanical disruption was achieved using a TissueLyser II (Qiagen), followed by two hours of shaking on a Vibrax shaker. The resulting samples were centrifuged to isolate methanol extracts devoid of plant debris, and further extracted twice with hexane to capture non-polar and medium polar terpenes in the upper layer. Analysis of 200 µL hexane extracts was performed on a gas chromatograph (GC) Trace 1300 (ThermoFisher) coupled with a mass spectrometer (MS) ISQ 700 (ThermoFisher) and a CD-5MS column (30 m x 0.25 mm x 0.25 µm) (ThermoFisher).

Sample injection (1 µL) was conducted in splitless mode, following GC-MS oven parameters modified from (Kumar et al., 2016): initial temperature of 70°C for 3 min; first ramp of 20°C/min to 90°C; second ramp of 3°C/min to 180°C; third ramp of 5°C/min to 240°C; and final ramp of 20°C/min to 300°C, with a 6 min hold at this temperature. The MS initiated compound analysis after a 5.5-minute delay, with the transfer line and ion source temperatures set at 270°C and 230°C, respectively. Compound detection utilised the scan mode (scan time 0.17 s) within a mass detection range of 40 to 600 atomic mass units.

Chromatograms were processed using Chromeleon™ software (ThermoFisher), and quantification of main terpenes relied on comparison with the internal standard dodecane. Tentative identification of the major sesquiterpenes was achieved through comparison of mass

spectra with literature, supplemented by a single sample run with a mix of C9-C40 alkanes standard to calculate the retention indices (Adams, 2007).

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

RNA was extracted from 3-4 two-week old gemmae, using the RNeasy Plant kit (#74903, Qiagen) according to the manufacturer's protocol (RLT buffer supplemented with beta-merchaptoethanol was used) and residual genomic DNA was removed using the Turbo DNA-free kit (#AM1907, Invitrogen) according to the manufacturer's instructions.

500 ng of DNase-treated RNA was used as a template for cDNA preparation using the SuperScript™ IV First-Strand Synthesis System (#18091050, Invitrogen) according to manufacturer's instructions (with only modifying reverse transcriptase reaction time to 40 minutes and using oligo (dT)18 primers). qPCR was performed using the SYBR Green JumpStart *Taq* Ready Mix (#S4438, Sigma Aldrich) and a CFX384 RT System (Bio-Rad) thermal cycler. cDNA was diluted 6 times, oligonucleotides listed in **Dataset S5** were used at a final concentration of 0.5 µM and reaction conditions were as follows: initial denaturation step of 94°C for 2 minutes followed by 40 cycles of 94°C for 15 seconds (denaturation) and 60°C for 1 minute (annealing, extension, and fluorescence reading).

Library preparation and RNA sequencing was performed by Novogene (Cambridge, UK). Briefly, messenger RNA was purified from total RNA using poly-T oligo-attached magnetic beads. After fragmentation, the first strand cDNA was synthesised using random hexamer primers followed by the second strand cDNA synthesis. cDNA end repair, A-tailing, adapter ligation, size selection, amplification, and purification were performed next. Library concentration was measured on a Qubit instrument following the manufacturer's procedure (ThermoFisher Scientific) followed by real-time qPCR quantification. Library size distribution was analysed on a bioanalyzer (Agilent) following the manufacturer's protocol. Quantified libraries were pooled and sequenced on a NovaSeq PE150 Illumina platform and 6 G raw data per sample were obtained. Adapter sequences were: 5' Adapter: 5'-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-3'. 3' Adapter: 5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGATGACTATCTCGTATGCCGTCTTCTGCTTG-3'

FastQC was used to assess read quality and TrimGalore (https://doi.org/10.5281/zenodo.5127899) to trim low-quality reads and remaining sequencing adapters. Reads were pseudo-aligned using Kallisto (Bray et al., 2016) to the *M. polymorpha* Genome version 5 (primary transcripts only, obtained from MarpolBase) (Bowman et al. 2018). Kallisto estimates the abundance of each transcript in units of transcripts per million (TPM). Mapping statistics for each library are provided in **Dataset S3**. DGE analysis was performed with DESeq2 (Love et al., 2014), with padj-values < 0.01, **Dataset S3**. Plots were generated using R. All raw data has been deposited in NCBI under the accession number PRJNA397637.

ChIP-seq and ATAC-seq

ChIP-seq was performed using two Mp*GLK* overexpression lines transformed with a construct of proMp*EF1a*::Mp*GLK*-HA. Regenerating thalli from T1 transformants were cultured for 4 weeks and harvested at zeitgeber time 1h. Approximately 5g fresh thalli were crosslinked with 1% (w/v) formaldehyde. Nuclei were isolated as previously described (Tu et al., 2022). Chromatin was then sonicated using Bioruptor (4x 5 min cycles of 30sec ON/OFF). Sonicated chromatin was incubated with 2µg Anti-HA Tag Antibody (#26183, Invitrogen) for 6h at 4°C and an additional 2h at 4°C with 20µl Dynabeads Protein G (#10003D, ThermoFisher) blocked with 0.1% BSA. Beads with immunoprecipitated DNA were washed twice with low salt buffer (10mM Tris-HCl pH8.0, 0.15M NaCl, 1mM EDTA, 1% Triton X-100), twice with high salt buffer (10mM Tris-HCl pH8.0, 0.25M NaCl, 1mM EDTA, 1% Triton X-100) and TET (10mM Tris-HCl pH8.0, 1mM EDTA, 0.1% Tween-20). TS-Tn5 was purified and assembled with Tn5MEA/B adapter as previously described (Tu et al., 2020). Tagmentation was performed on-bead with assembled TS-Tn5 for 30 min at 37°C. After tagmentation, beads were washed with low salt, high salt buffer and TET. Reverse crosslinking were performed at 55°C for 1h and 65°C overnight with 0.3M NaCl, 0.4% SDS, 10mM Tris-HCl pH8.0, 1mM EDTA, Proteinase K. DNA was purified with Qiagen MinElute (#28004, Qiagen) and PCR amplified with primers matching the Illumina N50x and N70x sequences. Libraries were sequenced with Hiseq X using the 150 bp paired-end mode. Histone ChIP-seq and ATAC-seq were performed as previously described (Tu et al., 2020). All raw data has been deposited in NCBI

under the accession number PRJNA1043823. The processed data can be viewed at http://www.epigenome.cuhk.edu.hk/.

ChIP-seq and ATAC-seq data analysis

ChIP-seq reads and ATAC-seq reads of *A. thaliana* and *M. polymorpha* were separately mapped to the reference genomes (*A. thaliana* TAIR10, *M. polymorpha* Tak v6) with bowtie2 with option -3 100 (version 2.3.5.1). The unmapped and low quality reads were filtered with SAMtools view (version 1.10) -F 4 and -q 20, and duplicated reads were removed using SAMTools rmdup. MACS2 (Y. Zhang et al., 2008) were used for peak calling and peak summit positions were retrieved by using the "–call-summits" function in MACS2. The peaks are resized to 150 bp regions (±75 bp from the summit position) for IDR to calculate overlap. The summit's signal fold-change values from two replicates are supplied to IDR to identify reproducible peaks. The overlap regions passed the IDR 0.01 and with average summit signal foldchange >5 in two replicates were kept and resized back to 150 bp. They were then associated with gene promoters based on summit distance to the TSS (-1.5kb and + 0.5kb).

De novo motif discovery was performed with HOMER. GLK motifs in the GLK ChIP-seq summit region were extracted using findMotifsGenome.pl in HOMER, with the parameter -len 8. Othofinder (version 2.2.7) was used to identify orthologs using the peptide sequence of all protein coding genes in *M. polymorpha*, *A. thaliana, Nicotiana benthamiana, Solanum lycopersicum, Oryza sativa,* and *Zea mays*. The transcription factor prediction was performed with iTAK. TargetP 2.0 (https://services.healthtech.dtu.dk/services/TargetP-2.0/) was used to predict chloroplast localization of the protein encoded by GLK target genes.

Author contributions

N.Y.E., E.F., T.S., M.T., K. B., Ed.F., J.R., Z.M., S.R., Y.B., J.S-W., J.H. and S.Z. carried out the work. N.Y.E., E.F. and J.M.H. designed the work. N.Y.E, E.F. and J.M.H. wrote the manuscript with input from all authors.

Acknowledgements

This work was funded as part of the BBSRC/EPSRC OpenPlant Synthetic Biology Research Centre Grant BB/ L014130/1 to J.H., BBSRC BB/F011458/1 for confocal microscopy to J.H, BBP0031171 to J.M.H. and Hong Kong GRF 1411918/14109420 to S.Z.. T.B.S. was supported by the SNSF Postdoc Mobility Fellowship (P500PB_203128) and the EMBO Long-Term Fellowship (ALTF 531-2019). W.Z. was supported by the Zhujiang postdoc fellowship. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission. We thank Karin H. Müller and Filomena Gallo from the Cambridge Advanced Imaging Centre for the electron microscopy sample preparation as well as the support during the image acquisition.

REFERENCES

Adams, R. P. (2007). *Identification of Essential Oil Components by Gas Chromatography/mass Spectrometry*.

Archibald, J. M. (2009). The Puzzle of Plastid Evolution. In *Current Biology* (Vol. 19, Issue 2, pp. R81–R88). https://doi.org/10.1016/j.cub.2008.11.067

Behringer, C., & Schwechheimer, C. (2015). B-GATA transcription factors – insights into their structure, regulation, and role in plant development. In *Frontiers in Plant Science* (Vol. 6). https://doi.org/10.3389/fpls.2015.00090

Binkert, M., Kozma-Bognár, L., Terecskei, K., De Veylder, L., Nagy, F., & Ulm, R. (2014). UV-B-responsive association of the Arabidopsis bZIP transcription factor ELONGATED HYPOCOTYL5 with target genes, including its own promoter. *The Plant Cell*, *26*(10), 4200–4213.

Bi, Y.-M., Zhang, Y., Signorelli, T., Zhao, R., Zhu, T., & Rothstein, S. (2005). Genetic analysis of Arabidopsis GATA transcription factor gene family reveals a nitrate-inducible member important for chlorophyll synthesis and glucose sensitivity. In *The Plant Journal* (Vol. 44, Issue 4, pp. 680–692). https://doi.org/10.1111/j.1365-313x.2005.02568.x

Bolle, C. (2004). The role of GRAS proteins in plant signal transduction and development. *Planta*, *218*(5), 683–692.

Bowman, J. L., Arteaga-Vazquez, M., Berger, F., Briginshaw, L. N., Carella, P., Aguilar-Cruz, A., Davies, K. M., Dierschke, T., Dolan, L., Dorantes-Acosta, A. E., Fisher, T. J., Flores-Sandoval, E., Futagami, K., Ishizaki, K., Jibran, R., Kanazawa, T., Kato,

H., Kohchi, T., Levins, J., … Zachgo, S. (2022). The renaissance and enlightenment of Marchantia as a model system. *The Plant Cell*, *34*(10), 3512–3542.

Bravo-Garcia, A., Yasumura, Y., & Langdale, J. A. (2009). Specialization of the Golden2-like regulatory pathway during land plant evolution. *The New Phytologist*, *183*(1), 133–141.

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*(5), 525–527.

Burko, Y., Gaillochet, C., Seluzicki, A., Chory, J., & Busch, W. (2020). Local HY5 Activity Mediates Hypocotyl Growth and Shoot-to-Root Communication. *Plant Communications*, *1*(5). https://doi.org/10.1016/j.xplc.2020.100078

Cackett, L., Luginbuehl, L. H., Schreier, T. B., Lopez-Juez, E., & Hibberd, J. M. (2022). Chloroplast development in green plant tissues: the interplay between light, hormone, and transcriptional regulation. *The New Phytologist*, *233*(5), 2000–2016.

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* , *25*(15), 1972–1973.

Cheng, S., Xian, W., Fu, Y., Marin, B., Keller, J., Wu, T., Sun, W., Li, X., Xu, Y., Zhang, Y., Wittek, S., Reder, T., Günther, G., Gontcharov, A., Wang, S., Li, L., Liu, X., Wang, J., Yang, H., … Melkonian, M. (2019). Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant Evolution. *Cell*, *179*(5), 1057–1067.e14.

Chiang, Y.-H., Zubo, Y. O., Tapken, W., Kim, H. J., Lavanway, A. M., Howard, L., Pilon,

M., Kieber, J. J., & Schaller, G. E. (2012). Functional characterization of the GATA transcription factors GNC and CGA1 reveals their key role in chloroplast development, growth, and division in Arabidopsis. *Plant Physiology*, *160*(1), 332–348.

Clayton, W. A., Albert, N. W., Thrimawithana, A. H., McGhie, T. K., Deroles, S. C., Schwinn, K. E., Warren, B. A., McLachlan, A. R. G., Bowman, J. L., Jordan, B. R., & Davies, K. M. (2018). UVR8-mediated induction of flavonoid biosynthesis for UVB tolerance is conserved between the liverwort Marchantia polymorpha and flowering plants. *The Plant Journal: For Cell and Molecular Biology*, *96*(3), 503–517.

Delmans, M., Pollak, B., & Haseloff, J. (2017). MarpoDB: An Open Registry for Marchantia Polymorpha Genetic Parts. *Plant & Cell Physiology*, *58*(1), e5.

de Vries, J., & Archibald, J. M. (2018). Plant evolution: landmarks on the path to terrestrial life. *The New Phytologist*, *217*(4), 1428–1434.

de Vries, J., & Gould, S. B. (2018). The monoplastidic bottleneck in algae and plant evolution. *Journal of Cell Science*, *131*(2). https://doi.org/10.1242/jcs.203414

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238.

Fitter, D. W., Martin, D. J., Copley, M. J., Scotland, R. W., & Langdale, J. A. (2002). *GLK*gene pairs regulate chloroplast development in diverse plant species. In *The Plant Journal* (Vol. 31, Issue 6, pp. 713–727). https://doi.org/10.1046/j.1365-313x.2002.01390.x

Gould, S. B., Waller, R. F., & McFadden, G. I. (2008). Plastid evolution. *Annual Review of Plant Biology*, *59*, 491–517.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321.

Harris, B. J., Clark, J. W., Schrempf, D., Szöllősi, G. J., Donoghue, P. C. J., Hetherington, A. M., & Williams, T. A. (2022). Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. *Nature Ecology & Evolution*, *6*(11), 1634–1643.

Harris, B. J., Harrison, C. J., Hetherington, A. M., & Williams, T. A. (2020). Phylogenomic Evidence for the Monophyly of Bryophytes and the Reductive Evolution of Stomata. *Current Biology: CB*, *30*(11), 2001–2012.e2.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522.

Hudson, D., Guevara, D., Yaish, M. W., Hannam, C., Long, N., Clarke, J. D., Bi, Y.-M., & Rothstein, S. J. (2011). GNC and CGA1 modulate chlorophyll biosynthesis and glutamate synthase (GLU1/Fd-GOGAT) expression in Arabidopsis. *PloS One*, *6*(11), e26765.

Jarvis, P., & López-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. *Nature Reviews. Molecular Cell Biology*, *14*(12), 787–802.

Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., & Gao, G. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, *45*(D1), D1040–D1045.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S.

(2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589.

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4), 1160–1166.

Kawamura, S., Romani, F., Yagura, M., Mochizuki, T., Sakamoto, M., Yamaoka, S., Nishihama, R., Nakamura, Y., Yamato, K. T., Bowman, J. L., Kohchi, T., & Tanizawa, Y. (2022). MarpolBase Expression: A Web-Based, Comprehensive Platform for Visualization and Analysis of Transcriptomes in the Liverwort Marchantia polymorpha. *Plant & Cell Physiology*, *63*(11), 1745–1755.

Kobayashi, K., Sasaki, D., Noguchi, K., Fujinuma, D., Komatsu, H., Kobayashi, M., Sato, M., Toyooka, K., Sugimoto, K., Niyogi, K. K., Wada, H., & Masuda, T. (2013). Photosynthesis of root chloroplasts developed in Arabidopsis lines overexpressing GOLDEN2-LIKE transcription factors. *Plant & Cell Physiology*, *54*(8), 1365–1377.

Kumar, S., Kempinski, C., Zhuang, X., Norris, A., Mafu, S., Zi, J., Bell, S. A., Nybo, S. E., Kinison, S. E., Jiang, Z., Goklany, S., Linscott, K. B., Chen, X., Jia, Q., Brown, S. D., Bowman, J. L., Babbitt, P. C., Peters, R. J., Chen, F., & Chappell, J. (2016). Molecular Diversity of Terpene Synthases in the Liverwort Marchantia polymorpha. *The Plant Cell*, *28*(10), 2632–2650.

Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., Tongprasit, W., Zhao, H., Lee, I., & Deng, X. W. (2007). Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *The Plant Cell*, *19*(3), 731–749.

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296.

Li, F.-W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P.-M., Eily, A., Koppers, N., Kuo, L.-Y., Li, Z., Simenc, M., Small, I., Wafula, E., Angarita, S., Barker, M. S., Bräutigam, A., dePamphilis, C., Gould, S., Hosmani, P. S., … Pryer, K. M. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants*, *4*(7), 460–472.

Li, F.-W., Nishiyama, T., Waller, M., Frangedakis, E., Keller, J., Li, Z., Fernandez-Pozo, N., Barker, M. S., Bennett, T., Blázquez, M. A., Cheng, S., Cuming, A. C., de Vries, J., de Vries, S., Delaux, P.-M., Diop, I. S., Harrison, C. J., Hauser, D., Hernández-García, J., … Szövényi, P. (2020). Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. *Nature Plants*, *6*(3), 259–272.

Lin, S.-S., & Bowman, J. L. (2018). MicroRNAs in Marchantia polymorpha. *The New Phytologist*, *220*(2), 409–416.

Llave, C., Xie, Z., Kasschau, K. D., & Carrington, J. C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science*, *297*(5589), 2053–2056.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550.

Luan, J., Ju, J., Li, X., Wang, X., Tan, Y., & Xia, G. (2023). Functional identification of moss PpGATA1 provides insights into the evolution of LLM-domain B-GATA transcription factors in plants. *Gene*, *855*, 147103.

Luo, X.-M., Lin, W.-H., Zhu, S., Zhu, J.-Y., Sun, Y., Fan, X.-Y., Cheng, M., Hao, Y., Oh, E., Tian, M., Liu, L., Zhang, M., Xie, Q., Chong, K., & Wang, Z.-Y. (2010). Integration of light- and brassinosteroid-signaling pathways by a GATA transcription factor in Arabidopsis. *Developmental Cell*, *19*(6), 872–883.

Ma, Z., Hu, X., Cai, W., Huang, W., Zhou, X., Luo, Q., Yang, H., Wang, J., & Huang, J. (2014). Arabidopsis miR171-Targeted Scarecrow-Like Proteins Bind to GT cis-Elements and Mediate Gibberellin-Regulated Chlorophyll Biosynthesis under Light Conditions. In *PLoS Genetics* (Vol. 10, Issue 8, p. e1004519). https://doi.org/10.1371/journal.pgen.1004519

Naito, T., Kiba, T., Koizumi, N., Yamashino, T., & Mizuno, T. (2007). Characterization of a unique GATA family gene that responds to both light and cytokinin in Arabidopsis thaliana. *Bioscience, Biotechnology, and Biochemistry*, *71*(6), 1557–1560.

Nakamura, H., Muramatsu, M., Hakata, M., Ueno, O., Nagamura, Y., Hirochika, H., Takano, M., & Ichikawa, H. (2009). Ectopic overexpression of the transcription factor OsGLK1 induces chloroplast development in non-green rice cells. *Plant & Cell Physiology*, *50*(11), 1933–1949.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274.

One Thousand Plant Transcriptomes Initiative. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, *574*(7780), 679–685.

Oyama, T., Shimura, Y., & Okada, K. (1997). The Arabidopsis HY5 gene encodes a

bZIP protein that regulates stimulus-induced development of root and hypocotyl. *Genes & Development*, *11*(22), 2983–2995.

Powell, A. L. T., Nguyen, C. V., Hill, T., Cheng, K. L., Figueroa-Balderas, R., Aktas, H., Ashrafi, H., Pons, C., Fernández-Muñoz, R., Vicente, A., Lopez-Baltazar, J., Barry, C. S., Liu, Y., Chetelat, R., Granell, A., Van Deynze, A., Giovannoni, J. J., & Bennett, A. B. (2012). Uniform ripening encodes a Golden 2-like transcription factor regulating tomato fruit chloroplast development. *Science*, *336*(6089), 1711–1715.

Qiu, Y.-L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., Estabrook, G. F., Hendry, T. A., Taylor, D. W., Testa, C. M., Ambros, M., Crandall-Stotler, B., Duff, R. J., Stech, M., Frey, W., … Davis, C. C. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(42), 15511–15516.

Reyes, J. C., Muro-Pastor, M. I., & Florencio, F. J. (2004). The GATA family of transcription factors in Arabidopsis and rice. *Plant Physiology*, *134*(4), 1718–1732.

Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., & Bartel, D. P. (2002). Prediction of plant microRNA targets. *Cell*, *110*(4), 513–520.

Rich, M. K., & Delaux, P.-M. (2020). Plant Evolution: When Arabidopsis Is More Ancestral Than Marchantia [Review of *Plant Evolution: When Arabidopsis Is More Ancestral Than Marchantia*]. *Current Biology: CB*, *30*(11), R642–R644.

Rossini, L., Cribb, L., Martin, D. J., & Langdale, J. A. (2001). The maize golden2 gene defines a novel class of transcriptional regulators in plants. *The Plant Cell*, *13*(5), 1231–1244.

Safi, A., Medici, A., Szponarski, W., Ruffel, S., Lacombe, B., & Krouk, G. (2017). The world according to GARP transcription factors. *Current Opinion in Plant Biology*, *39*, 159–167.

Saint-Marcoux, D., Proust, H., Dolan, L., & Langdale, J. A. (2015). Identification of reference genes for real-time quantitative PCR experiments in the liverwort Marchantia polymorpha. *PloS One*, *10*(3), e0118678.

Sauret-Güeto, S., Frangedakis, E., Silvestri, L., Rebmann, M., Tomaselli, M., Markel, K., Delmans, M., West, A., Patron, N. J., & Haseloff, J. (2020). Systematic Tools for Reprogramming Plant Gene Expression in a Simple Model,. *ACS Synthetic Biology*, *9*(4), 864–882.

Schreier, T. B., Müller, K. H., Eicke, S., Faulkner, C., Zeeman, S. C., & Hibberd, J. M. (2024). Plasmodesmal connectivity in C Gynandropsis gynandra is induced by light and dependent on photosynthesis. *The New Phytologist*, *241*(1), 298–313.

Schröder, P., Hsu, B.-Y., Gutsche, N., Winkler, J. B., Hedtke, B., Grimm, B., & Schwechheimer, C. (2023). B-GATA factors are required to repress high-light stress responses in Marchantia polymorpha and Arabidopsis thaliana. *Plant, Cell & Environment*, *46*(8), 2376–2390.

Sugano, S. S., Nishihama, R., Shirakawa, M., Takagi, J., Matsuda, Y., Ishida, S., Shimada, T., Hara-Nishimura, I., Osakabe, K., & Kohchi, T. (2018). Efficient CRISPR/Cas9-based genome editing and its application to conditional genetic analysis in Marchantia polymorpha. *PloS One*, *13*(10), e0205117.

Todesco, M., Rubio-Somoza, I., Paz-Ares, J., & Weigel, D. (2010). A collection of target mimics for comprehensive analysis of microRNA function in Arabidopsis thaliana.

PLoS Genetics, *6*(7), e1001031.

Toledo-Ortiz, G., Johansson, H., Lee, K. P., Bou-Torrent, J., Stewart, K., Steel, G., Rodríguez-Concepción, M., & Halliday, K. J. (2014). The HY5-PIF regulatory module coordinates light and temperature control of photosynthetic gene transcription. *PLoS Genetics*, *10*(6), e1004416.

Trebst, A. (2007). Inhibitors in the functional dissection of the photosynthetic electron transport system. *Photosynthesis Research*, *92*(2), 217–224.

Tu, X., Mejía-Guerra, M. K., Valdes Franco, J. A., Tzeng, D., Chu, P.-Y., Shen, W., Wei, Y., Dai, X., Li, P., Buckler, E. S., & Zhong, S. (2020). Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. *Nature Communications*, *11*(1), 5089.

Tu, X., Ren, S., Shen, W., Li, J., Li, Y., Li, C., Li, Y., Zong, Z., Xie, W., Grierson, D., Fei, Z., Giovannoni, J., Li, P., & Zhong, S. (2022). Limited conservation in cross-species comparison of GLK transcription factor binding suggested wide-spread cistrome divergence. *Nature Communications*, *13*(1), 7632.

Ulm, R., Baumann, A., Oravecz, A., Máté, Z., Adám, E., Oakeley, E. J., Schäfer, E., & Nagy, F. (2004). Genome-wide analysis of gene expression reveals function of the bZIP transcription factor HY5 in the UV-B response of Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(5), 1397–1402.

Waller, M., Frangedakis, E., Marron, A. O., Sauret-Güeto, S., Rever, J., Sabbagh, C. R. R., Hibberd, J. M., Haseloff, J., Renzaglia, K. S., & Szövényi, P. (2023). An optimized transformation protocol for Anthoceros agrestis and three more hornwort

species. *The Plant Journal: For Cell and Molecular Biology*, *114*(3), 699–718.

Wang, L., Mai, Y.-X., Zhang, Y.-C., Luo, Q., & Yang, H.-Q. (2010). MicroRNA171c-targeted SCL6-II, SCL6-III, and SCL6-IV genes regulate shoot branching in Arabidopsis. *Molecular Plant*, *3*(5), 794–806.

Wang, P., Khoshravesh, R., Karki, S., Tapia, R., Balahadia, C. P., Bandyopadhyay, A., Quick, W. P., Furbank, R., Sage, T. L., & Langdale, J. A. (2017). Re-creation of a Key Step in the Evolutionary Switch from C to C Leaf Anatomy. *Current Biology: CB*, *27*(21), 3278–3287.e6.

Waters, M. T., Moylan, E. C., & Langdale, J. A. (2008). GLK transcription factors regulate chloroplast development in a cell-autonomous manner. *The Plant Journal: For Cell and Molecular Biology*, *56*(3), 432–444.

Waters, M. T., Wang, P., Korkaric, M., Capper, R. G., Saunders, N. J., & Langdale, J. A. (2009). GLK transcription factors coordinate expression of the photosynthetic apparatus in Arabidopsis. *The Plant Cell*, *21*(4), 1109–1128.

Xiao, Y., Chu, L., Zhang, Y., Bian, Y., Xiao, J., & Xu, D. (2021). HY5: A Pivotal Regulator of Light-Dependent Development in Higher Plants. *Frontiers in Plant Science*, *12*, 800989.

Yasumura, Y., Moylan, E. C., & Langdale, J. A. (2005). A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. *The Plant Cell*, *17*(7), 1894–1907.

Yelina, N. E., Holland, D., Gonzalez-Jorge, S., Hirsz, D., Yang, Z., & Henderson, I. R. (2022). Coexpression of MEIOTIC-TOPOISOMERASE VIB-dCas9 with guide RNAs specific to a recombination hotspot is insufficient to increase crossover

frequency in Arabidopsis. *G3* , *12*(7). https://doi.org/10.1093/g3journal/jkac105

Zhang, T., Zhang, R., Zeng, X.-Y., Lee, S., Ye, L.-H., Tian, S.-L., Zhang, Y.-J., Busch, W., Zhou, W.-B., Zhu, X.-G., & Wang, P. (2023). HY5 regulates GLK transcription factors to orchestrate photomorphogenesis in Arabidopsis thaliana. In *bioRxiv* (p. 2022.09.30.510343). https://doi.org/10.1101/2022.09.30.510343

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137.

Zheng, Y., Jiao, C., Sun, H., Rosli, H. G., Pombo, M. A., Zhang, P., Banf, M., Dai, X., Martin, G. B., Giovannoni, J. J., Zhao, P. X., Rhee, S. Y., & Fei, Z. (2016). iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. *Molecular Plant*, *9*(12), 1667–1670.

Zubo, Y. O., Blakley, I. C., Franco-Zorrilla, J. M., Yamburenko, M. V., Solano, R., Kieber, J. J., Loraine, A. E., & Schaller, G. E. (2018). Coordination of Chloroplast Development through the Action of the GNC and GLK Transcription Factor Families. *Plant Physiology*, *178*(1), 130–147.

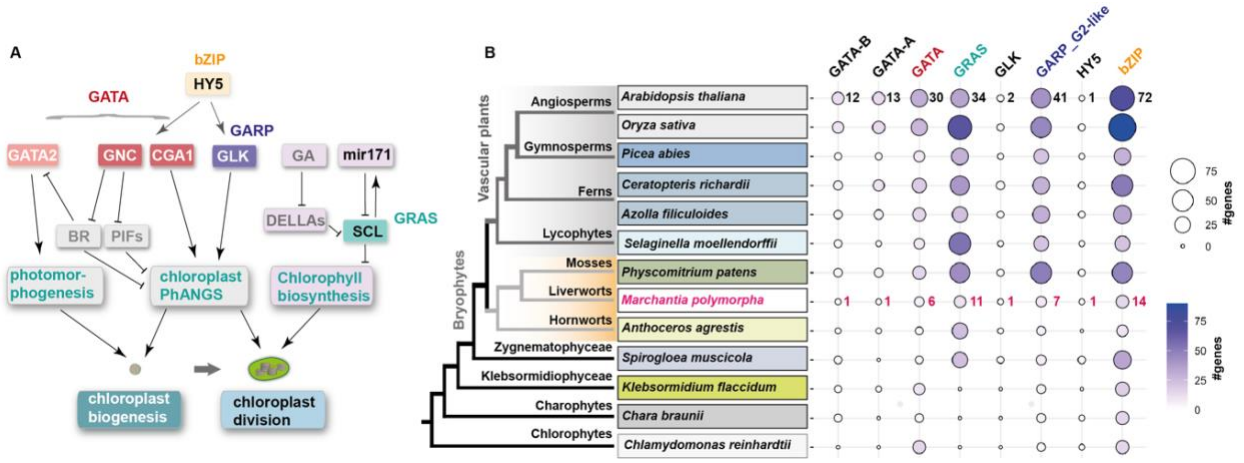
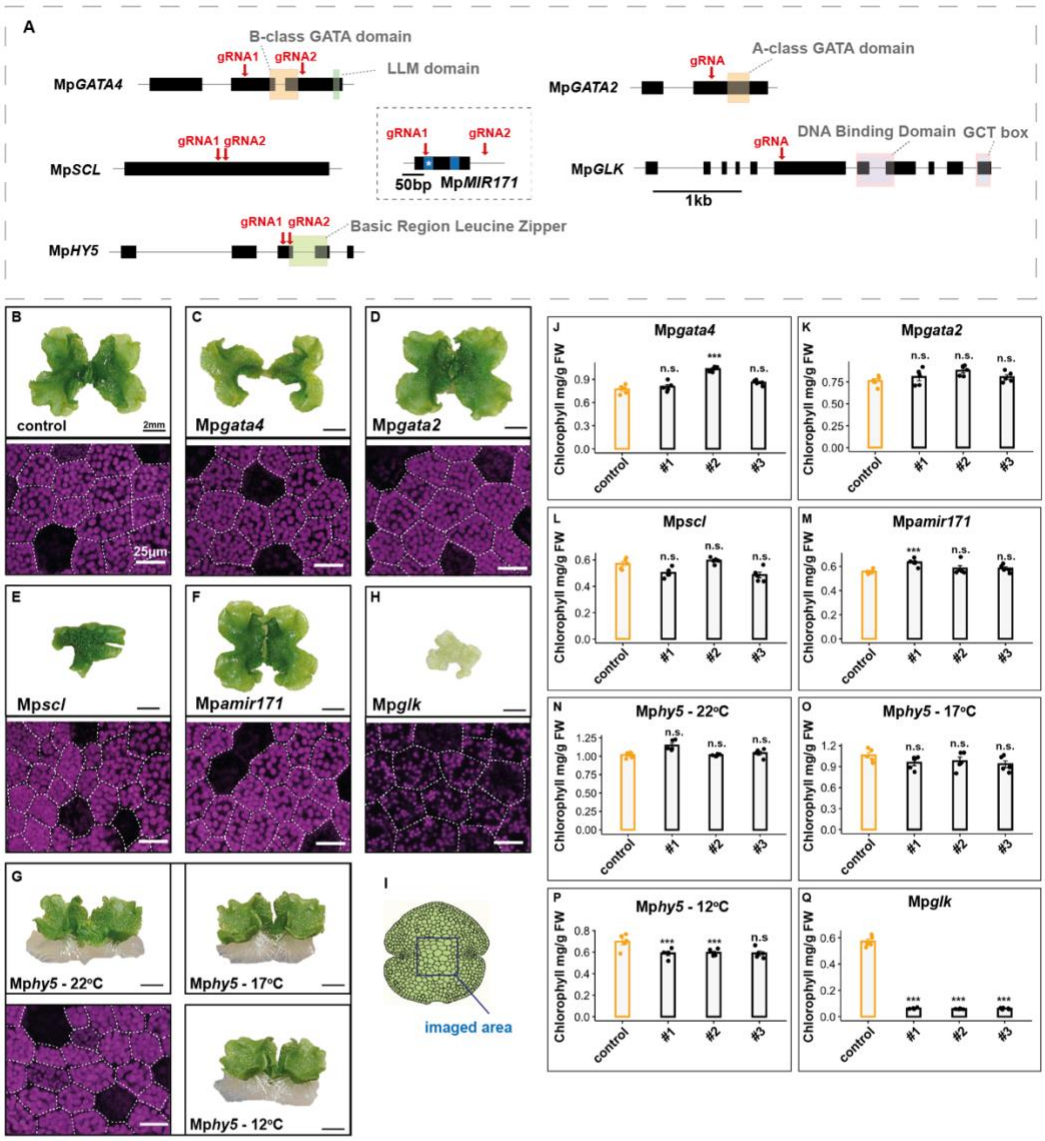


Figure 1. Transcription factors known to regulate chloroplast development in angiosperms. **A)** Simplified network illustrating families of transcription factors and key interacting components upstream of chloroplast biogenesis. GATA transcription factors include GATA2, GATA NITRATE-INDUCIBLE CARBON METABOLISM-INVOLVED (GNC), CYTOKININ- RESPONSIVE GATA FACTOR 1 (CGA1). GATA2 promotes photomorphogenesis in the presence of light by directly binding to light-responsive promoters. GNC and CGA1 suppress Phytochrome Interacting Factors (PIFs) and Brassinosteroid (BR) related genes, as well as promoting chloroplast biogenesis and division. In the absence of light, the BR-activated transcription factor BRASSINAZOLE RESISTANT1 (BZR1) represses *GATA2* expression. GOLDEN2-LIKE (GLK) transcription factors are positive regulators of nuclear-encoded photosynthesis related genes. SCARECROW-LIKE (SCL) GRAS transcription factors are negatively regulated by miR171 and GA-DELLA signalling to control chlorophyll biosynthesis. Elongated Hypocotyl 5 (HY5), a bZIP transcription factor that plays a primary role in de-etiolation and chloroplast development. **B)** Phylogenetic relationships for major lineages of land plants and green algae (Li et al. 2020). Gene numbers shown for transcription factor families and subfamilies in (A). Data derived from representative species for which high quality genome assemblies are available.


A
B-class GATA domain
gRNA1
gRNA2
LLM domain
MpGATA4
gRNA
A-class GATA domain
MpGATA2
gRNA1 gRNA2
MpSCL
gRNA1
gRNA2
50bp MpMIR171
gRNA
DNA Binding Domain
GCT box
MpGLK
1kb
gRNA1 gRNA2
Basic Region Leucine Zipper
MpHY5
B
control
2mm
25µm
C
Mpgata4
D
Mpgata2
E
Mpscl
F
Mpamir171
H
Mpglk
G
Mphy5 - 22°C
Mphy5 - 17°C
Mphy5 - 12°C
I
imaged area
J
Mpgata4
K
Mpgata2
L
Mpscl
M
Mpamir171
N
Mphy5 - 22°C
O
Mphy5 - 17°C
P
Mphy5 - 12°C
Q
Mpglk
Chlorophyll mg/g FW
control
#1
#2
#3
n.s.
***

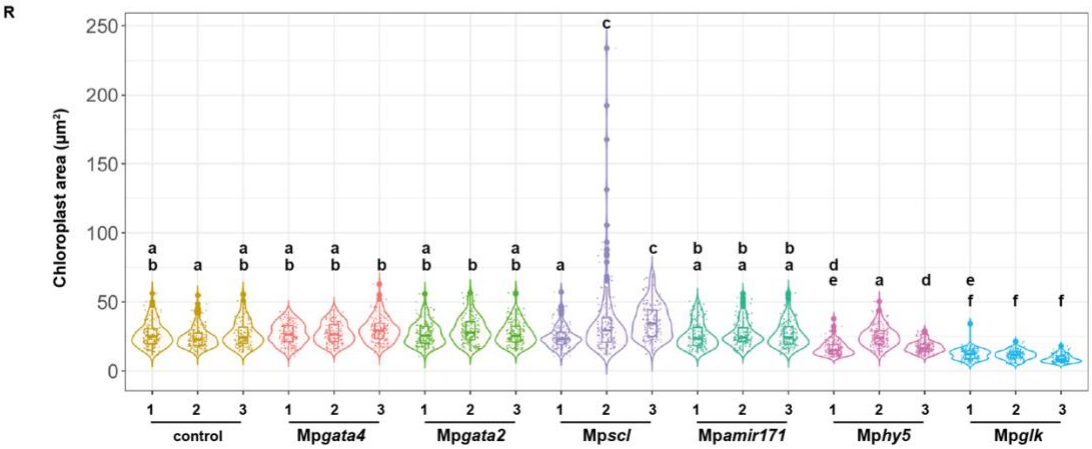

R
Chloroplast area (µm²)
control
Mpgata4
Mpgata2
Mpscl
Mpamir171
Mphy5
Mpglk

Figure 2. MpGLK controls chlorophyll and chloroplast size.

A) Schematic representation of Mp*GLK,* Mp*GATA4,* Mp*GATA2,* Mp*MIR171,* Mp*SCL* and Mp*HY5* gene structure showing exons as black rectangles. Canonical domains for each family are shown as coloured boxes. Genomic locus of Mp*MIR171* represented as a black rectangle, miR171* and miR171 shown as blue rectangles, miR171* indicated with a white star. Guide RNA positions for CRISPR/Cas9 gene editing shown as red arrows. **B-G)** Top: Representative images of wild type, Mp*gata4*, Mp*gata2*, Mp*scl*, Mp*mir171,* Mp*hy5* and Mp*glk* mutants. Scale bars represent 1 cm. Bottom: Representative confocal microscopy images of wild type, Mp*gata4*, Mp*gata2*, Mp*scl*, Mp*mir171* and Mp*glk* mutants. Chlorophyll autofluorescence is shown in magenta and cell boundaries marked with dashed lines. Scale bars represent 1 cm. **I)** Schematic of a gemmae with dark blue square indicating imaged area, **J-Q)** Barplots of chlorophyll content for the Mp*gata4*, Mp*gata2*, Mp*scl*, Mp*mir17*1, Mp*hy5* and Mp*glk* mutant plants. Individual values are shown with dots. Error bars represent the standard deviation of the mean $n = 5$. Asterisks indicate statistically significant difference using a two-tailed *t*-test $P \leq 0.0001$ (***). **R)** Boxplot showing chloroplast size range in control versus Mp*gata4*, Mp*gata2*, Mp*scl*, Mp*mir171,* Mp*hy5* and Mp*glk* mutants. 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. n = 150.

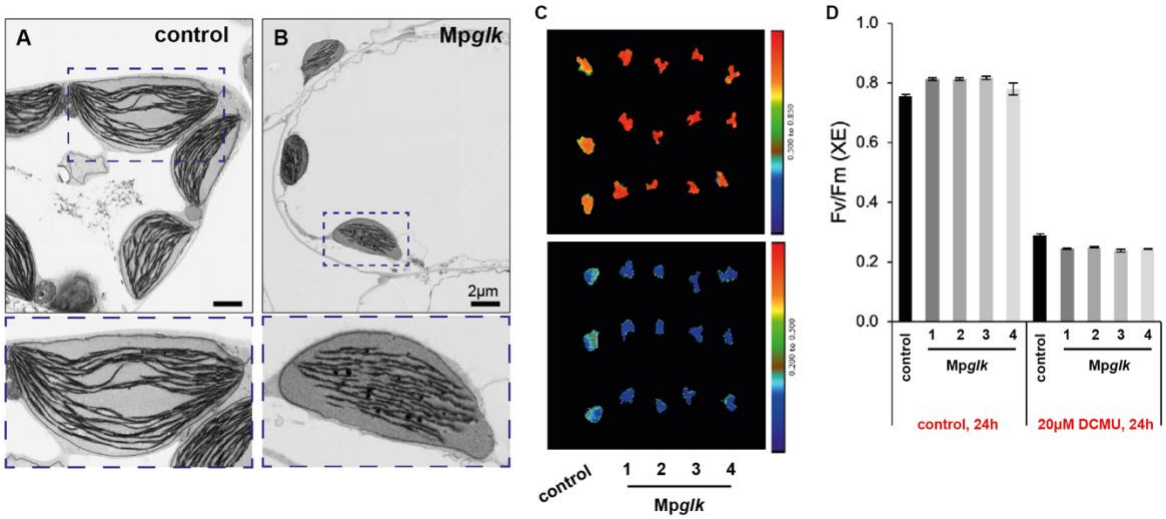


Figure 3. Mp*glk* controls chloroplast size and ultrastructure.

A-B) Scanning electron micrographs of chloroplasts from control (A) and Mp*glk* mutants (B). Dashed boxes highlight single chloroplasts that are shown in the insets at higher magnification (bottom panels). Scale bars represent 2 µm. **C)** Chlorophyll fluorescence images of maximum quantum efficiency of Photosystem II (*Fv/Fm*) of untreated (top) and 24 h DCMU-treated (bottom) control and Mp*glk* mutants. **D)** *Fv/Fm* measured in DCMU-treated and untreated control and Mp*glk* mutants. Bars represent mean ± standard error from *n* = 3 plants per genotype.

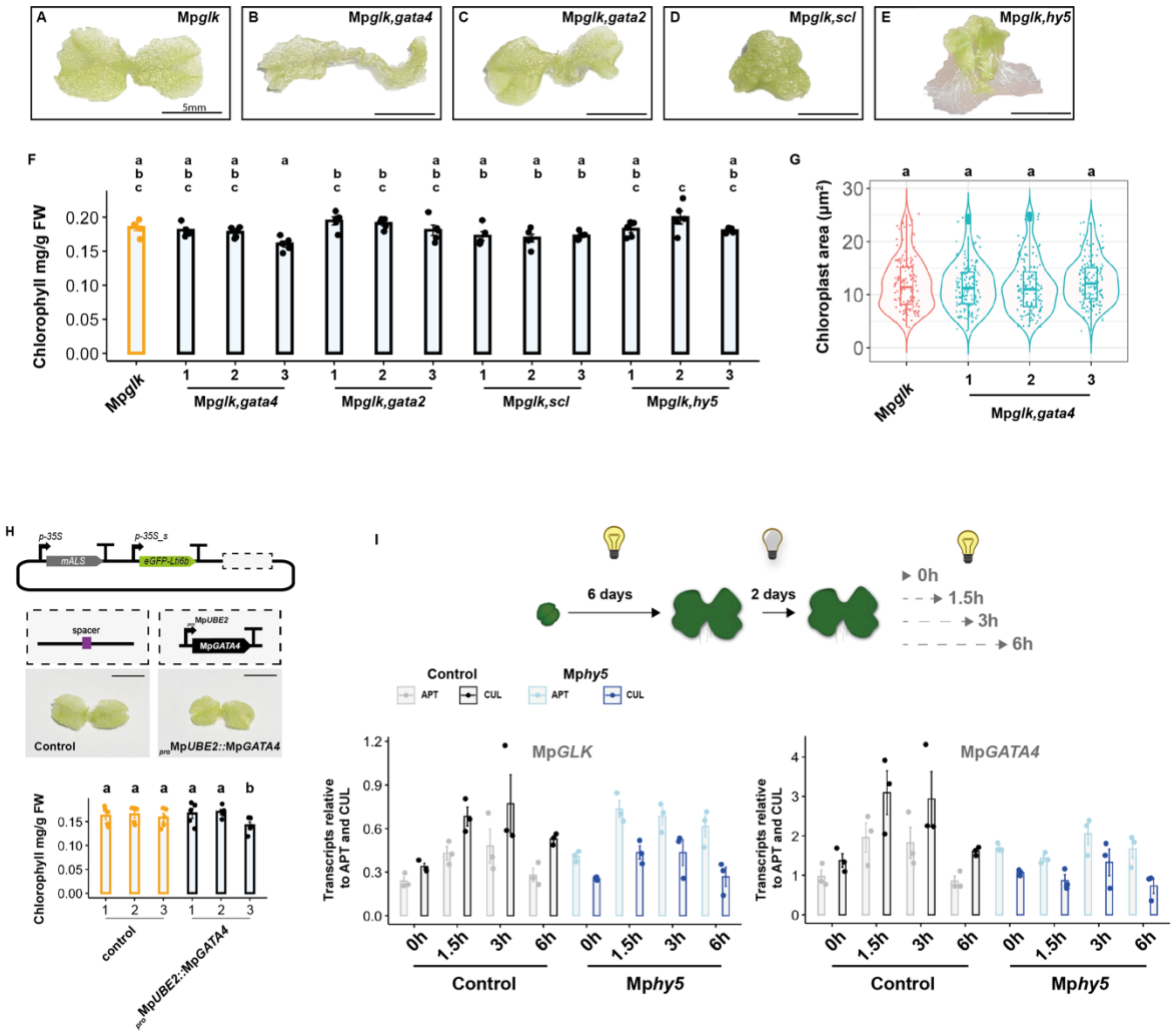


Figure 4. Loss of MpGATA4, MpGATA2, MpSCL or MpHY5 in Mp*glk* does not generate additive perturbations to phenotype.

A-E) Representative images of Mp*glk,* Mp*glk,gata4,* Mp*glk,gata2,* Mp*glk,scl* and Mp*glk,hy5* double mutants. **F)** Chlorophyll content in Mp*glk,* Mp*glk,gata4,* Mp*glk,gata2,* Mp*glk,scl* and Mp*glk,hy5* double mutants. Letters show statistical ranking using a *post hoc* Tukey test, different letters indicating statistically significant differences at P<0.01. Values indicated by the same letter are not statistically different, n=5. **G)** Chloroplast area in Mp*glk,* and Mp*glk,gata4.* Letters show statistical ranking using a *post hoc* Tukey test, different letters indicating statistically significant differences at P<0.01. Values indicated by the same letter are not statistically different, n=150. **H)** Top: Schematic representation of Mp*GATA4* overexpression construct. Bottom: Representative images of Mp*glk* mutants and Mp*glk* mutant overexpressing Mp*GATA4*. Scale bar represents 5 mm. Bottom: Chlorophyll content in Mp*glk* mutant and Mp*glk* mutant overexpressing Mp*GATA4*. Letters show statistical ranking using a *post hoc* Tukey test, different letters indicating statistically significant differences at P<0.01. Values indicated by the same letter

are not statistically different, n=5. **I)** Top: Schematic of experimental design to test the effect of lack of functional MpHY5 on Mp*GLK* and Mp*GATA4* expression in response to light. Bottom: quantitative polymerase chain reaction (qPCR) for analysis of Mp*GLK* and Mp*GATA4* expression in Mp*hy5* mutants in response to light. *ADENINE PHOSPHORIBOSYL TRANSFERASE 3* (*APT*) and *CULLIN 1* (*CUL*) were used as housekeeping gene controls (Saint-Marcoux et al., 2015).

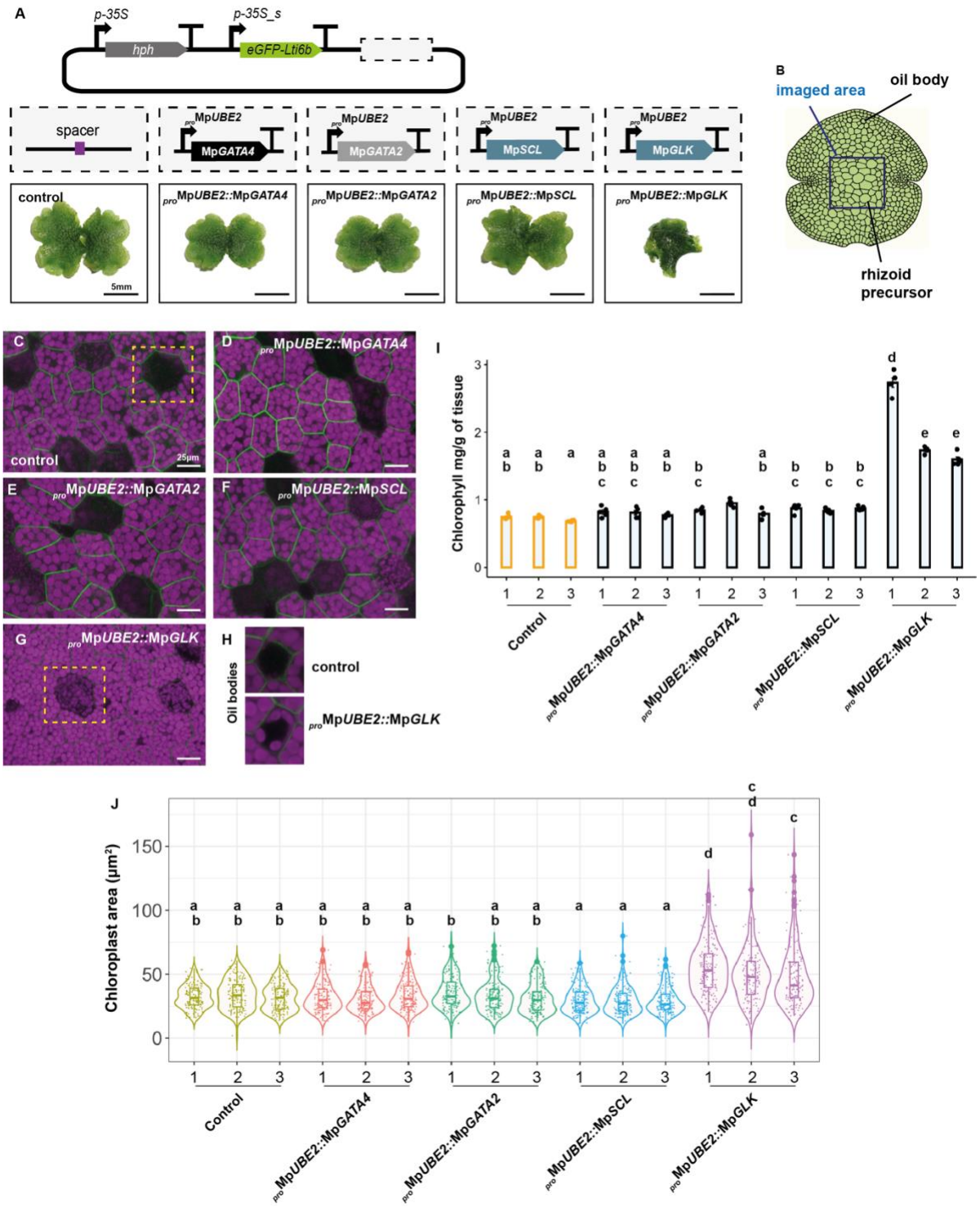

A
p-35S
p-35S_s
hph
eGFP-Lti6b
spacer
MpUBE2
MpGATA4
MpGATA2
MpSCL
MpGLK
control
5mm
MpUBE2::MpGATA4
MpUBE2::MpGATA2
MpUBE2::MpSCL
MpUBE2::MpGLK
B
imaged area
oil body
rhizoid precursor
C
control
25µm
D
E
F
G
H
Oil bodies
I
Chlorophyll mg/g of tissue
Control
J
Chloroplast area (µm²)

Figure 5. Mp*GLK* overexpression increases chlorophyll content and chloroplast size.
A) Top: Schematic representation of constructs used to overexpress Mp*GATA4,* Mp*GATA2,* Mp*SCL* and Mp*GLK.* Bottom: Representative images of controls and plants overexpressing Mp*GATA4,* Mp*GATA2,* Mp*SCL* and Mp*GLK*. Scale bar represents 5 mm. **B)** Schematic of a gemmae; a dark blue square indicates imaged area, black lines indicate approximate positions of rhizoid precursor and oil body cells. **C-G)** Representative confocal microscopy images of gemmae from controls and plants overexpressing Mp*GATA4,* Mp*GATA2,* Mp*SCL* and Mp*GLK*. Chlorophyll autofluorescence shown in magenta and plasma membrane marked with eGFP in green. Rhizoids precursors highlighted with dashed squares. Scale bars, represent 25 µm. **H)** Representative confocal microscopy images of oil body cells in control and Mp*GLK-CDS* overexpressing lines. **I)** Chlorophyll content in controls and overexpressing lines for Mp*GATA4,* Mp*GATA2,* Mp*SCL* and Mp*GLK.* Letters show statistical ranking using a *post hoc* Tukey test (different letters indicating statistically significant differences at P<0.01). Values indicated by the same letter are not statistically different, n=5. **J)** Chloroplast area in controls and lines overexpressing Mp*GATA4,* Mp*GATA2,* Mp*SCL* and Mp*GLK.* Letters show statistical ranking using a *post hoc* Tukey test (different letters indicating statistically significant differences at P<0.01). Values indicated by the same letter are not statistically different , n=150.

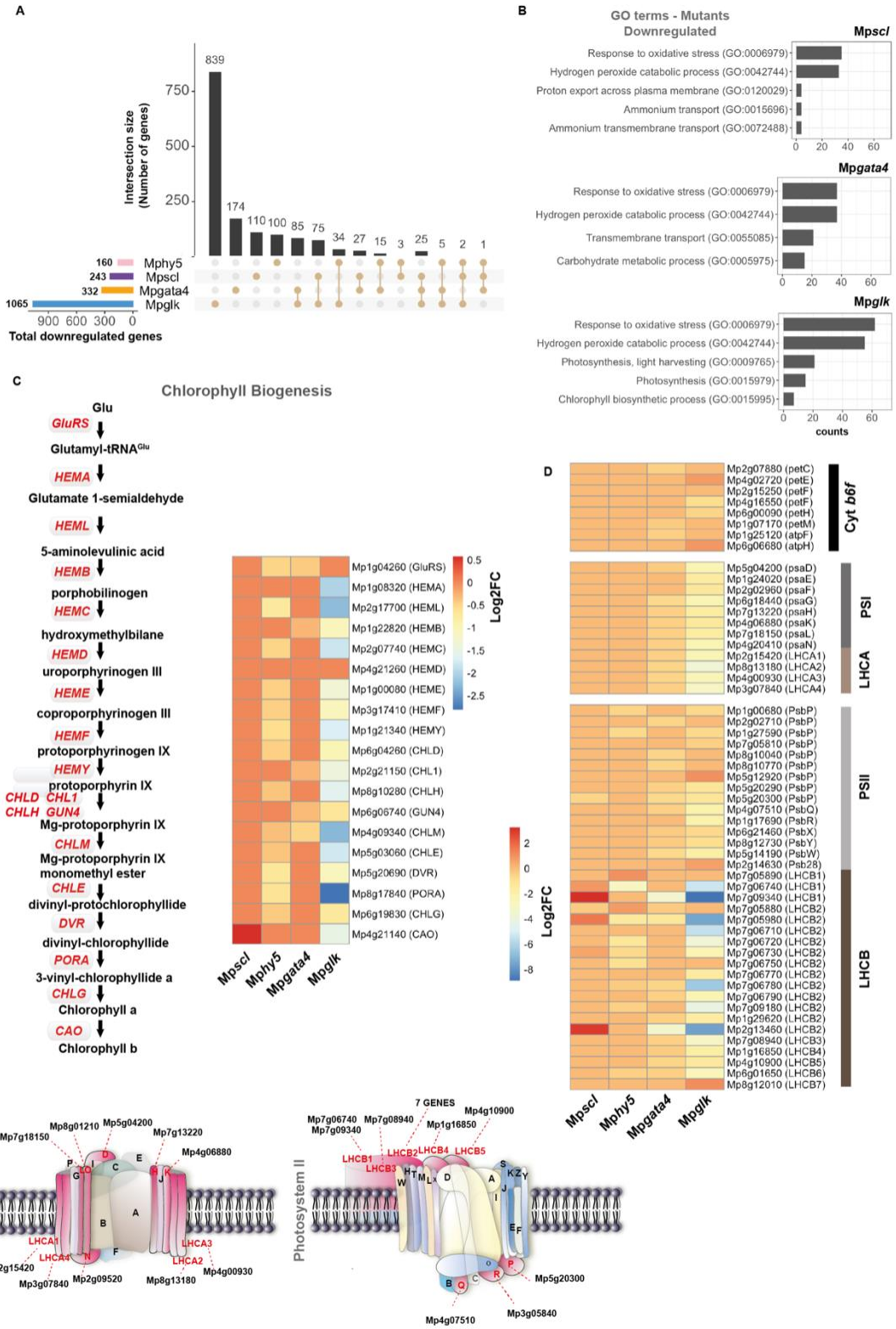

A
Intersection size (Number of genes)
839
174
110
100
85
75
34
27
15
3
25
5
2
1
160 Mphy5
243 Mpscl
332 Mpgata4
1065 Mpglk
Total downregulated genes
B
GO terms - Mutants Downregulated
Mpscl
Response to oxidative stress (GO:0006979)
Hydrogen peroxide catabolic process (GO:0042744)
Proton export across plasma membrane (GO:0120029)
Ammonium transport (GO:0015696)
Ammonium transmembrane transport (GO:0072488)
Mpgata4
Response to oxidative stress (GO:0006979)
Hydrogen peroxide catabolic process (GO:0042744)
Transmembrane transport (GO:0055085)
Carbohydrate metabolic process (GO:0005975)
Mpglk
Response to oxidative stress (GO:0006979)
Hydrogen peroxide catabolic process (GO:0042744)
Photosynthesis, light harvesting (GO:0009765)
Photosynthesis (GO:0015979)
Chlorophyll biosynthetic process (GO:0015995)
counts
C
Chlorophyll Biogenesis
Glu
GluRS
Glutamyl-tRNA^Glu
HEMA
Glutamate 1-semialdehyde
HEML
5-aminolevulinic acid
HEMB
porphobilinogen
HEMC
hydroxymethylbilane
HEMD
uroporphyrinogen III
HEME
coproporphyrinogen III
HEMF
protoporphyrinogen IX
HEMY
protoporphyrin IX
CHLD CHL1
CHLH GUN4
Mg-protoporphyrin IX
CHLM
Mg-protoporphyrin IX monomethyl ester
CHLE
divinyl-protochlorophyllide
DVR
divinyl-chlorophyllide
PORA
3-vinyl-chlorophyllide a
CHLG
Chlorophyll a
CAO
Chlorophyll b
Mp1g04260 (GluRS)
Mp1g08320 (HEMA)
Mp2g17700 (HEML)
Mp1g22820 (HEMB)
Mp2g07740 (HEMC)
Mp4g21260 (HEMD)
Mp1g00080 (HEME)
Mp3g17410 (HEMF)
Mp1g21340 (HEMY)
Mp6g04260 (CHLD)
Mp2g21150 (CHL1)
Mp8g10280 (CHLH)
Mp6g06740 (GUN4)
Mp4g09340 (CHLM)
Mp5g03060 (CHLE)
Mp5g20690 (DVR)
Mp8g17840 (PORA)
Mp6g19830 (CHLG)
Mp4g21140 (CAO)
Mpscl
Mphy5
Mpgata4
Mpglk
Log2FC
D
Mp2g07880 (petC)
Mp4g02720 (petE)
Mp2g15250 (petF)
Mp4g16550 (petF)
Mp6g00090 (petH)
Mp1g07170 (petM)
Mp1g25120 (atpF)
Mp6g06680 (atpH)
Cyt b6f
Mp5g04200 (psaD)
Mp1g24020 (psaE)
Mp2g02960 (psaF)
Mp6g18440 (psaG)
Mp7g13220 (psaH)
Mp4g06880 (psaK)
Mp7g18150 (psaL)
Mp4g20410 (psaN)
Mp2g15420 (LHCA1)
Mp8g13180 (LHCA2)
Mp4g00930 (LHCA3)
Mp3g07840 (LHCA4)
PSI
LHCA
Mp1g00680 (PsbP)
Mp2g02710 (PsbP)
Mp1g27590 (PsbP)
Mp7g05810 (PsbP)
Mp8g10040 (PsbP)
Mp8g10770 (PsbP)
Mp5g12920 (PsbP)
Mp5g20290 (PsbP)
Mp5g20300 (PsbP)
Mp4g07510 (PsbQ)
Mp1g17690 (PsbR)
Mp6g21460 (PsbX)
Mp8g12730 (PsbY)
Mp5g14190 (PsbW)
Mp2g14630 (Psb28)
Mp7g05890 (LHCB1)
Mp7g06740 (LHCB1)
Mp7g09340 (LHCB1)
Mp7g05880 (LHCB2)
Mp7g05980 (LHCB2)
Mp7g06710 (LHCB2)
Mp7g06720 (LHCB2)
Mp7g06730 (LHCB2)
Mp7g06750 (LHCB2)
Mp7g06770 (LHCB2)
Mp7g06780 (LHCB2)
Mp7g06790 (LHCB2)
Mp7g09180 (LHCB2)
Mp1g29620 (LHCB2)
Mp2g13460 (LHCB2)
Mp7g08940 (LHCB3)
Mp1g16850 (LHCB4)
Mp4g10900 (LHCB5)
Mp6g01650 (LHCB6)
Mp8g12010 (LHCB7)
PSII
LHCB
Mpscl
Mphy5
Mpgata4
Mpglk
Log2FC
E
Photosystem I
Mp7g18150
Mp8g01210
Mp5g04200
Mp7g13220
Mp4g06880
LHCA1
LHCA4
LHCA3
LHCA2
Mp2g15420
Mp3g07840
Mp2g09520
Mp8g13180
Mp4g00930
Photosystem II
7 GENES
Mp7g06740
Mp7g09340
Mp7g08940
Mp1g16850
Mp4g10900
LHCB1
LHCB3
LHCB2
LHCB4
LHCB5
Mp5g20300
Mp4g07510
Mp3g05840

Figure 6. Mp*GLK* controls expression of photosynthesis associated genes.

A) Upset diagrams showing sets of downregulated genes in Mp*scl,* Mp*gata4,* Mp*hy5* and Mp*glk* mutant alleles. **B)** Enriched Gene Ontology terms for Mp*scl,* Mp*gata4* and Mp*glk* mutants. **C)** Heatmap illustrating extent of down-regulation (Log2FoldChange) of transcripts encoding enzymes of chlorophyll biosynthesis in Mp*scl,* Mp*gata4,* Mp*hy5* and Mp*glk* mutant alleles. D) Heatmap indicating lower transcript abundance (Log2FoldChange) of genes encoding components of Photosystem II (PSII), Photosystem I (PSI), Light Harvest Complex A (LHCA), Light Harvest Complex B (LHCB) and cytochrome b_6f complex in Mp*scl,* Mp*gata4,* Mp*hy5* and Mp*glk* mutants. E) Schematic representation of Photosystem I and II. Subunits showing an increase (>1) in corresponding transcript levels are highlighted in pink. Genes that are differentially expressed are highlighted in red. Figure modified from Water et al. 2009 and Tu et al. 2022.

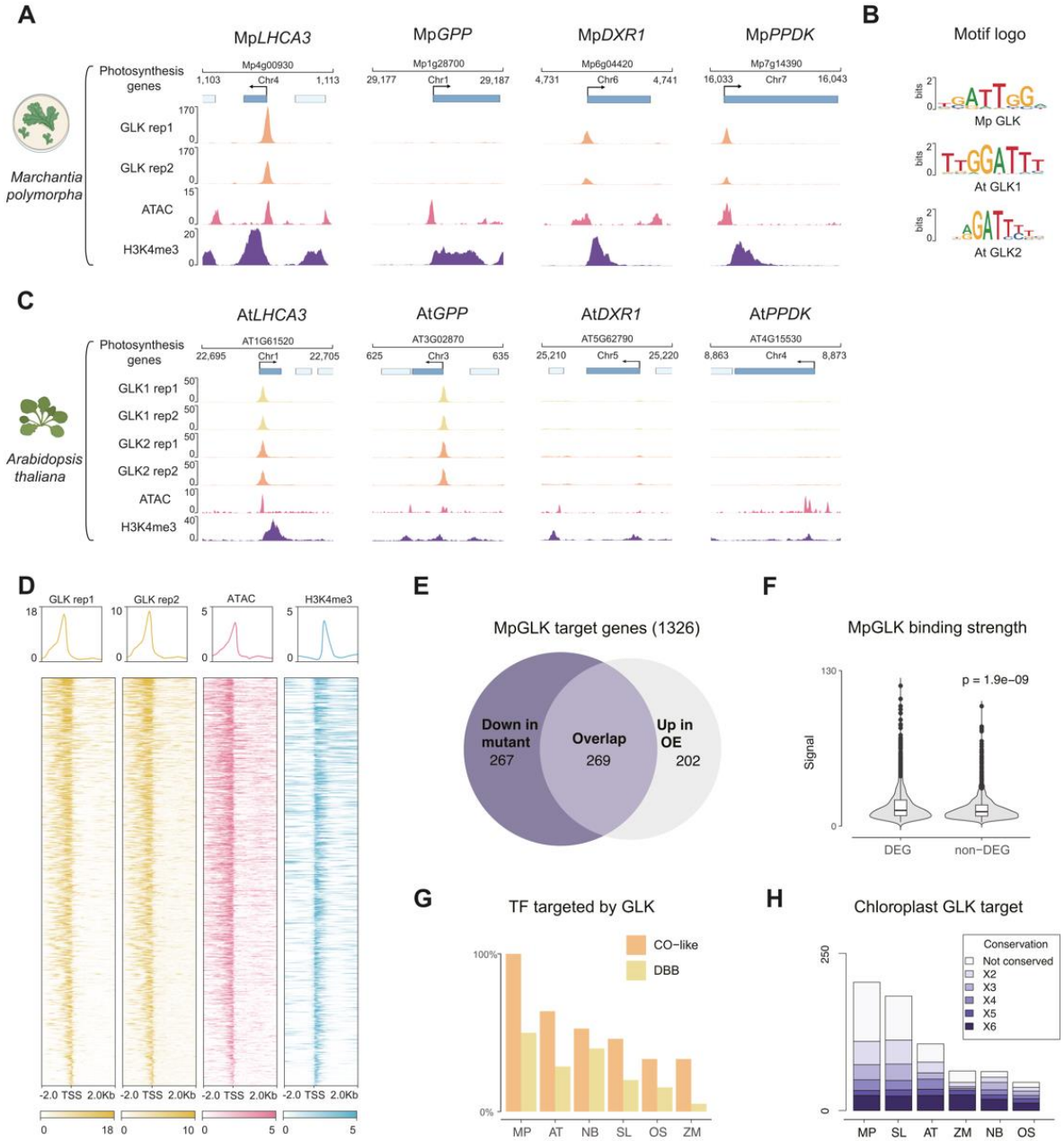

A
MpLHCA3
MpGPP
MpDXR1
MpPPDK
Photosynthesis genes
Mp4g00930
Mp1g28700
Mp6g04420
Mp7g14390
GLK rep1
GLK rep2
ATAC
H3K4me3
Marchantia polymorpha
B
Motif logo
Mp GLK
At GLK1
At GLK2
C
AtLHCA3
AtGPP
AtDXR1
AtPPDK
AT1G61520
AT3G02870
AT5G62790
AT4G15530
GLK1 rep1
GLK1 rep2
GLK2 rep1
GLK2 rep2
Arabidopsis thaliana
D
-2.0 TSS 2.0Kb
E
MpGLK target genes (1326)
Down in mutant 267
Overlap 269
Up in OE 202
F
MpGLK binding strength
p = 1.9e−09
Signal
DEG
non−DEG
G
TF targeted by GLK
CO−like
DBB
MP
AT
NB
SL
OS
ZM
H
Chloroplast GLK target
Conservation
Not conserved
X2
X3
X4
X5
X6

Figure 7. MpGLK ChIP-seq revealed divergence of GLK binding between *M. polymorpha* and flowering plants.

A) Genome browser tracks showing GLK ChIP-seq peaks, open chromatin regions (ATAC-seq), and H3K4me3 peaks in the conserved and *M. polymorpha* specific GLK target genes. *LHCB3: Light Harvest Complex 3; GPP: Galactose 1-phosphate phosphatase; DXR1; 1-deoxy-D-xylulose-5-phosphate reductoisomerase; PPDK: Pyruvate,orthophosphate (Pi) dikinase.* **B)** Motif enriched in the MpGLK and *A. thaliana* GLKs ChIP-seq peak regions. **C)** Genome browser tracks showing GLK1/2 binding, open chromatin and H3K4me3 in the *A. thaliana* orthologs of the *M. polymorpha* genes in panel A. **D)** Heatmap and average signal plot showing the chromatin features near the MpGLK binding sites. Clustering is performed using the GLK1 ChIP-seq rep1 signal. **E)** Partial overlap of MpGLK targets genes up-regulated in GLK over-expression (OE) lines and down-regulated in GLK knockout lines. **F)** Stronger ChIP-seq signals are found in the differentially expressed GLK target genes. **G)** Percentage of CONSTANS-Like (CO-like) and B-BOX (DBB) TF genes in each species' genome that are targeted by GLK. Note that all three CO-Like TFs in *M. polymorpha* are GLK targets. **H)** Number of predicted chloroplast localised GLK targets conserved between *M. polymorpha* and five flowering plants.