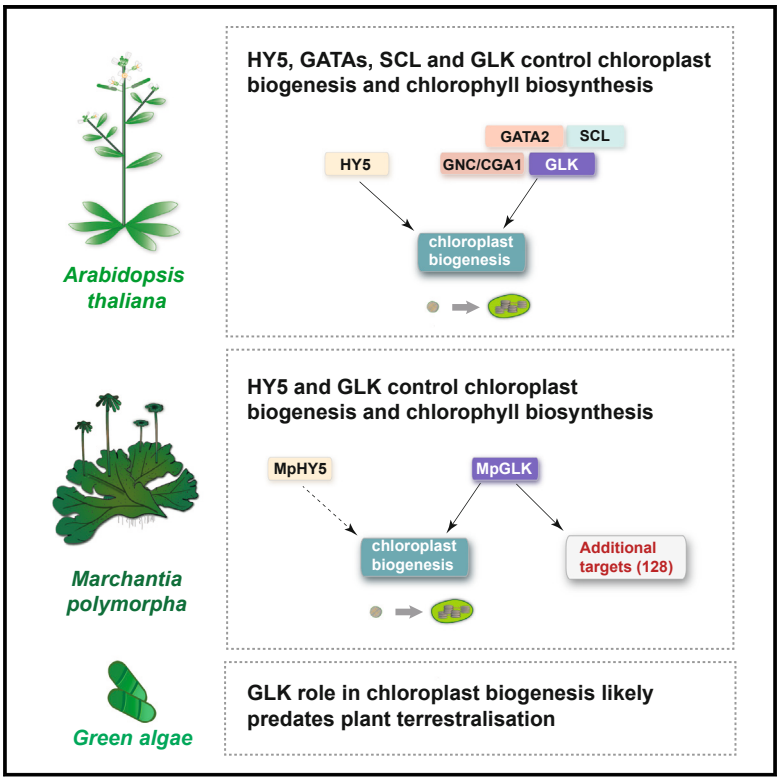
Cell Reports

Article

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

Graphical abstract



Authors

Nataliya E. Yelina, Eftychios Frangedakis, Zhemin Wang, ..., Jim Haseloff, Silin Zhong, Julian M. Hibberd

Correspondence

efragedakis@gmail.com (E.F.), jmh65@cam.ac.uk (J.M.H.)

In brief

Yelina et al. determine the degree of functional conservation of known regulators of chloroplast biogenesis (GLK, GATAs, HY5, and SCL-MIR171 module) in the model liverwort *Marchantia polymorpha*. Only HY5 and GLK play conserved roles in chloroplast biogenesis. GLK function likely evolved before the emergence of land plants.

Highlights

- GATAs do not have a detectable role in chloroplast biogenesis in *M. polymorpha*
- GLK role in chloroplast biogenesis is conserved in *M. polymorpha*
- Targets of GLK diverged during angiosperm evolution
- Green algae have a functional GLK homolog

Yelina et al., 2024, Cell Reports *43*, 114696
September 24, 2024
https://doi.org/10.1016/j.celrep.2024.114696

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

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

Nataliya E. Yelina,[1,3,5] Eftychios Frangedakis,[1,5,*] Zhemin Wang,[2] Tina B. Schreier,[1,4] Jenna Rever,[1] Marta Tomaselli,[1] Edith C.F. Forestier,[1] Kumari Billakurthi,[1] Sibo Ren,[2] Yahui Bai,[2] Julia Stewart-Wood,[1] Jim Haseloff,[1] Silin Zhong,[2] and Julian M. Hibberd[1,6,*]

[1]Department of Plant Sciences, University of Cambridge, Cambridge CB3 EA, UK
[2]The State Key Laboratory of Agrobiotechnology, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China
[3]Present address: Crop Science Centre, University of Cambridge, 93 Lawrence Weaver Road, Cambridge CB3 0LE, UK
[4]Present address: Department of Biology, University of Oxford, South Parks Road, Oxford OX1 3RB, UK
[5]These authors contributed equally
[6]Lead contact
*Correspondence: efragedakis@gmail.com (E.F.), jmh65@cam.ac.uk (J.M.H.)
https://doi.org/10.1016/j.celrep.2024.114696

SUMMARY

Chloroplasts develop from undifferentiated plastids in response to light. In angiosperms, after the perception of light, the Elongated Hypocotyl 5 (HY5) transcription factor initiates photomorphogenesis, and two families of transcription factors known as GOLDEN2-LIKE (GLK) and GATA are considered master regulators of chloroplast development. In addition, the MIR171-targeted SCARECROW-LIKE GRAS transcription factors also impact chlorophyll biosynthesis. The extent to which these proteins carry out conserved roles in non-seed plants is not known. Using the model liverwort *Marchantia polymorpha*, we show that GLK controls chloroplast biogenesis, and HY5 shows a small conditional effect on chlorophyll content. Chromatin immunoprecipitation sequencing (ChIP-seq) revealed that MpGLK has a broader set of targets than has been reported in angiosperms. We also identified a functional GLK homolog in green algae. In summary, our data support the hypothesis that GLK carries out a conserved role relating to chloroplast biogenesis in land plants and green algae.

INTRODUCTION

Photosynthesis sustains the majority of life on Earth. In eukaryotes, this process takes place in organelles known as chloroplasts, thought to have originated from an endosymbiotic event between a photosynthetic prokaryote and eukaryotic cell.[1,2] Chloroplasts develop in response to light from undifferentiated proplastids,[3] and the key processes involved in chloroplast biogenesis include synthesis of chlorophyll, assembly of the thylakoid membranes, and accumulation of enzymes of the Calvin-Benson-Bassham cycle.[4] Since the endosymbiotic event, the majority of genes controlling chloroplast biogenesis have transferred from the plastid to the nucleus.[5] As chloroplast biogenesis needs to be responsive to the external environment in addition to the cell, nuclear-encoded photosynthesis genes are regulated by light and hormones. Key intermediaries allowing the integration of these responses are transcription factors (TFs).

Our understanding of TFs acting on photosynthesis and chloroplast biogenesis is based primarily on analysis of the model flowering plant *Arabidopsis thaliana* (Figure 1A). For example, ELONGATED HYPOCOTYL (HY5), a bZIP TF, acts antagonistically with phytochrome interacting factors (PIFs) to activate light-regulated genes in the presence of light,[6] allowing de-etiolation and chloroplast development.[7–9] TFs belonging to the GARP (GOLDEN2-LIKE, ARR-B, Psr1) and GATA families also play key roles in chloroplast biogenesis,[3] and members of the SCARECROW-LIKE (SCL) GRAS family impact on chlorophyll accumulation.[10] Within the GARP superfamily, Golden2-like (GLK) TFs are positive regulators of nuclear-encoded chloroplast and photosynthesis-related genes.[11–14] 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 PIFs as well as brassinosteroid (BR)-related genes to promote chloroplast biogenesis.[15–18] Moreover, GATA2 promotes photomorphogenesis by directly binding to light-responsive promoters[19] and in the absence of light BRASSINAZOLE RESISTANT1 (BZR1), a BR-activated TF, represses *GATA2* expression to inhibit photomorphogenesis. Notably, a recent study showed that in *A. thaliana*, *GLK* and *GATAs* are direct targets of HY5.[20] Lastly, three *A. thaliana* MIR171-targeted SCL TFs redundantly regulate chlorophyll biosynthesis by repressing the expression of *PROTOCHLOROPHYLLIDE OXIDOREDUCTASE C*.[10]

Land plants evolved from aquatic green algae, and approximately 500 mya diverged into two major monophyletic clades comprising vascular plants (angiosperms, gymnosperms, ferns, and lycophytes) and bryophytes (hornworts, liverworts, and

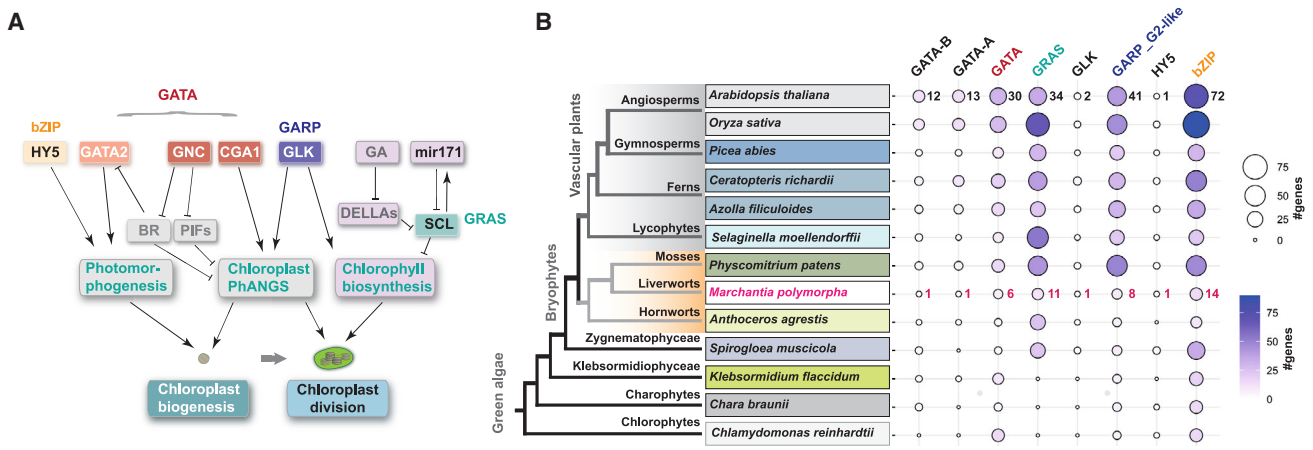


Figure 1. TFs known to regulate chloroplast development in angiosperms

(A) Simplified network illustrating families of TFs and key interacting components upstream of chloroplast biogenesis. GATA TFs include GATA2, GNC, and CGA1. GATA2 promotes photomorphogenesis in the presence of light by directly binding to light-responsive promoters. GNC and CGA1 suppress PIFs and BR-related genes, as well as promote chloroplast biogenesis and division. In the absence of light, the BR-activated TF BZR1 represses *GATA2* expression. GLK TFs are positive regulators of nuclear-encoded photosynthesis-related genes. SCL GRAS TFs are negatively regulated by miR171 and GA-DELLA signaling to control chlorophyll biosynthesis. HY5 is a bZIP TF that plays a primary role in de-etiolation and chloroplast development.

(B) Phylogenetic relationships for major lineages of land plants and green algae.[21] Gene numbers shown for TF families and subfamilies in (A) (for the full list of species, please see Figure S1A).

mosses) (Figure 1B).[21–23] Bryophytes are of particular importance to more accurately infer the ancestral state of a trait. While earlier evolutionary hypotheses proposed that they represent a collection of paraphyletic lineages,[24] it is now thought that liverworts and mosses form a monophyletic group that split from hornworts approximately 400–500 mya.[21,23,25] This revised phylogeny supports the notion that traits found in the common ancestor of land plants could have diversified not only within the vascular plant clade but also in the three deeply divergent bryophyte groups. Thus, studying representative species from more than one bryophyte clade can provide insight into the likely ancestral state. Despite our detailed knowledge of chloroplast biogenesis in angiosperms, with the exception of GLK function in the moss *Physcomitrium patens*,[26] our understanding of how chloroplast biogenesis is controlled remains unclear. The liverwort *Marchantia 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 is available.[27,28] We therefore selected *M. polymorpha* to investigate processes underpinning the evolution of chloroplast biogenesis.

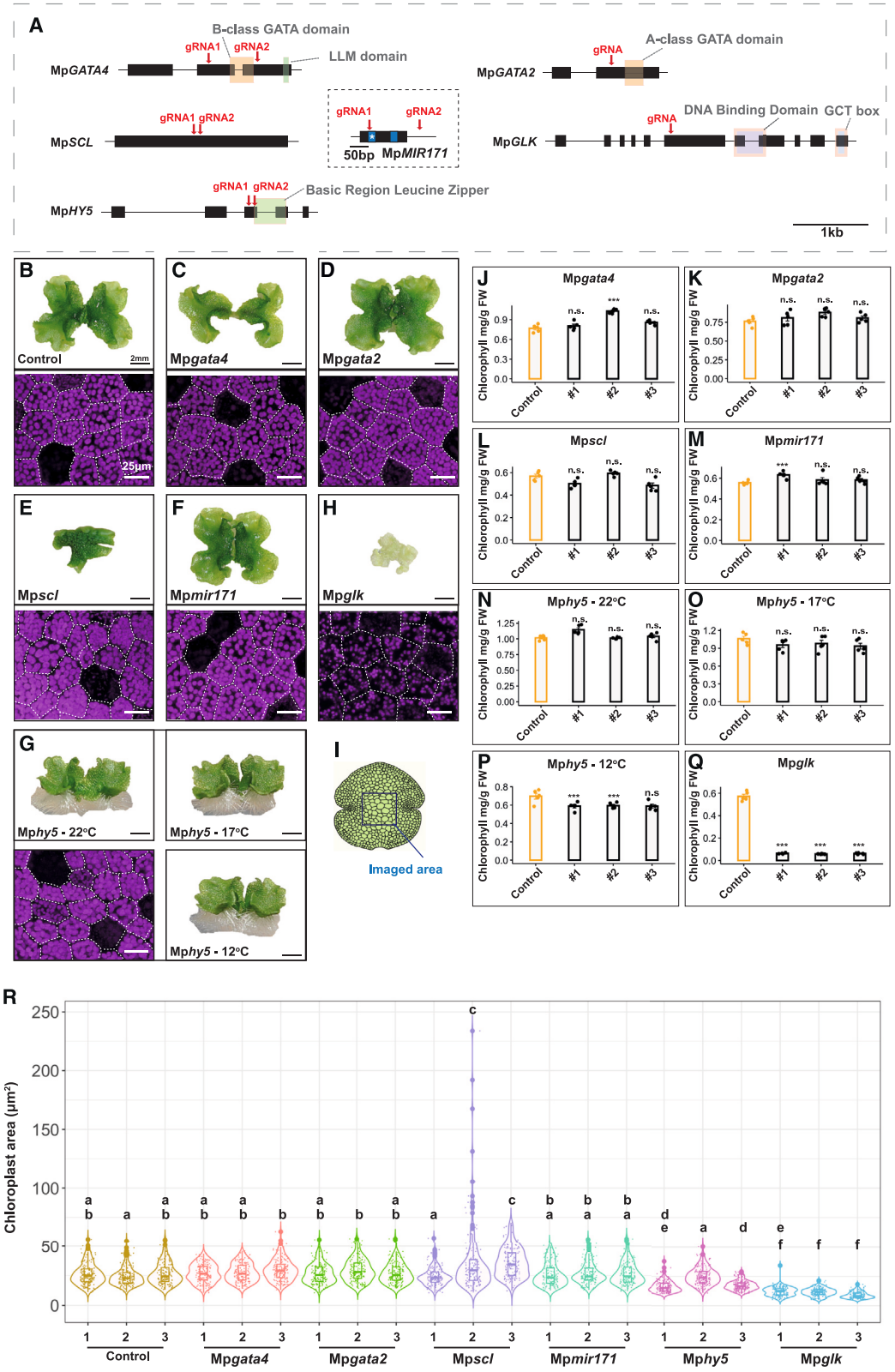
We first performed phylogenetic analysis to identify the *M. polymorpha* homologs of HY5, GLK, GATAs and the SCL-MIR171, and generated knockout mutants to test whether each component impacts on chloroplast biogenesis. We found that only HY5 and GLK are important for this process in *M. polymorpha,* but MpGATA and MpSCL are not. Moreover, we identified a functional GLK in green algae. When *GLK* was overexpressed, the abundance of transcripts derived from genes associated with chlorophyll biogenesis and photosystem I (PSI) and photosystem II (PSII) were impacted, and chromatin immunoprecipitation sequencing (ChIP-seq) confirmed that GLK can bind promoters of those genes. Intriguingly, many MpGLK targets are distinct from those documented in flowering plants. We conclude that GLK function is conserved between *M. polymorpha* and angiosperms, but that other regulators such as GNC/CGA1, GATA2, and SCL have no detectable impact on chloroplast development in the liverwort. Thus, in comparison to angiosperms, the pathway controlling chloroplast biogenesis in *M. polymorpha* is more streamlined.

RESULTS

Identification of *M. polymorpha* homologs of *HY5*, *GLK*, *GATA*, and *SCL* from angiosperms

While an HY5 homolog has previously been identified in *M. polymorpha,*[29] there is currently no systematic analysis of GATA, GLK, and SCL in this species. We therefore used phylogenetic analysis to search for GATA (GNC, CGA1, and GATA2), SCL, and GLK orthologs in *M. polymorpha*. To do so, we examined 16 representative species from the 7 main groups of land plants, as well as 5 green algae for which high-quality genome assemblies are available (Figure 1B, full set of species in Figure S1A).

GNC, CGA1, and GATA2 belong to the GATA superfamily of TFs that comprise a family of zinc finger proteins present in all eukaryotes[30,31] (Figure 2A, S1B, and S1C). *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 (Figures 1B, 2A, S1B, S1C, S2, and S3A; Data S1). SCL is a member of the GRAS family of TFs, which have a conserved C-terminal domain (Figures 1B and 2A). *A. thaliana* AtSCL6, AtSCL22, and AtSCL27 redundantly control chlorophyll biosynthesis and are regulated by MIR171. The *M. polymorpha* genome encodes 11 GRAS TFs. Mp8g03980



(legend on next page)

has been annotated as Mp*GRAS10* and is located in a sister clade to *A. thaliana* At*SCL6*, At*SCL22*, and At*SCL27*.[32] Based on conservation of its MIR171-mediated regulation, we refer to Mp8g03980 as Mp*SCL*. Lastly, GLK belongs to the GARP family of TFs[33] (Figures 1B, 2A, S1D, and S3B).[13] Our phylogenetic reconstructions identified Mp7g09740 (annotated as Mp*GARP8*) as a single ortholog of *GLK* due to the presence of the characteristic C-terminal GOLDEN2 C-terminal (GCT) box[13] (Figures S1D and S1E; Data S1). In summary, our analysis indicates that *M. polymorpha* contains homologs of *GNC/CGA1*, *GATA2*, *SCL*, and *GLK* that we hereafter refer to as Mp*GATA4*, Mp*GATA2*, Mp*SCL*, and Mp*GLK*, respectively.

Knockout mutant analysis of Mp*GATA*, Mp*SCL*, Mp*HY5*, and Mp*GLK*

To test whether Mp*GATA4*, Mp*GATA2*, the Mp*SCL-MIR171* module, Mp*HY5*, and Mp*GLK* homologs control chloroplast biogenesis in *M. polymorpha*, we used CRISPR-Cas9 editing to generate knockout mutant alleles for each (Figure 2A). Plants transformed with the same CRISPR-Cas9-containing vector but without a guide RNA (gRNA) sequence were used as "no gRNA" controls. Each mutant line was clonally propagated through one gemmae generation to obtain isogenic plants, and for each gene targeted, three independent lines were selected for analysis. For all protein-coding genes, mutations led to premature stop codons (Figure S4), and for Mp*mir171* mutants, the entire Mp*MIR171* gene was deleted (Figure S4D).

In Mp*gata4* mutants, gametophyte morphology was perturbed with narrower thallus lobes (Figures 2B and 2C). Compared with controls, the chlorophyll content was ~10% higher in one of the three lines examined (Figure 2J), a response opposite of that expected if Mp*GATA4* played a conserved role. We also analyzed chloroplast morphology 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 controls (Figures 2C and 2R). Mp*gata2* mutants did not show any morphological or developmental phenotypes and had similar chlorophyll levels compared with no gRNA controls (Figures 2B, 2D, and 2K). Chloroplast size in Mp*gata2* mutants was also not perturbed (Figures 2D and 2R). Mp*scl* mutants showed altered morphology, with thalli being stunted, with inward curling edges (Figures 2B and 2E). Despite chlorophyll levels not being 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 (Figures 2E and 2R) and a decrease in the number of chloroplasts per cell (Figure S4G), which was statistically significant in some cases. In contrast, Mp*mir171* mutants were indistinguishable from controls (Figures 2B, 2F, 2M, and 2R). Mp*hy5* gametophyte morphology was also perturbed, with thallus lobes being more erect (Figures 2B and 2G). Compared with controls, at 22°C, chloroplast size was ~15% smaller in two of the three Mp*hy5* lines (Figures 2G and 2R). In *A. thaliana*, *hy5* mutants show a conditional phenotype, with greater impact at lower temperatures.[34] To test whether HY5 also has a conditional phenotype in *M. polymorpha*, we grew plants at 22°C, 17°C, or 12°C. At 22°C and 17°C, mutants and controls had similar chlorophyll content (Figures 2G, 2N, and 2O), but at 12°C, a small but statistically significant reduction in chlorophyll was evident (Figures 2G and 2P). Lastly, several sesquiterpenes were more abundant in Mp*hy5* compared with controls (Data S2). We conclude that *hy5* mutants from *M. polymorpha* and *A. thaliana* have similar temperature-dependent phenotypes relating to chloroplast biogenesis, but in *M. polymorpha*, the protein also appears to impact on terpenoid accumulation, a role distinct from that previously reported for flavonoid biosynthesis.[29]

Mp*glk* mutants had an obvious pale green phenotype (Figures 2B and 2H). 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) (Figures S5A–S5F). In the strong Mp*glk* mutant alleles, chlorophyll content was reduced by ~90% compared with controls, and chloroplasts were smaller (Figures 2H, 2Q, and 2R). 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 as well as reproductive organs (Figure S5G) without supplemental carbon. Consistent with confocal laser scanning microscopy, electron microscopy confirmed that Mp*glk* mutants had smaller chloroplasts compared with controls and ultrastructure was perturbed (Figures 3A and 3B). Specifically, mutant chloroplasts had fewer thylakoid membranes with reduced granal stacking (Figures 3A, 3B, and S5H). Treatment with di-chlorophenyl dimethyl urea (DCMU), which inhibits the photosynthetic electron transport chain,[35] caused a substantial reduction in the chlorophyll fluorescence parameter F_v/F_m in mutants, indicating that their photosynthetic apparatus was functional (Figures 3C and 3D). This is consistent with the ability of strong Mp*glk*

Figure 2. Knockout mutants for Mp*GLK*, Mp*GATAs*, Mp*MIR171*, Mp*SCL*, and Mp*HY5*

(A) Schematic representation of Mp*GLK*, Mp*GATA4*, Mp*GATA2*, Mp*MIR171*, Mp*SCL*, and Mp*HY5* gene structure showing exons as black rectangles. Characteristic domains for each family are shown as colored boxes. Genomic locus of Mp*MIR171* represented as a black rectangle; miR171* and miR171 shown as blue rectangles; miR171* indicated with a white star. gRNA positions for CRISPR-Cas9 gene editing shown as red arrows.

(B–H) Top: representative images of control, Mp*gata4*, Mp*gata2*, Mp*scl*, Mp*mir171*, Mp*hy5*, and Mp*glk* mutants. Bottom: representative confocal microscopy images of control, Mp*gata4*, Mp*gata2*, Mp*scl*, Mp*mir171*, and Mp*glk* mutants. Chlorophyll autofluorescence is shown in magenta, and cell boundaries are marked with dashed lines. Scale bars: (top) 2 mm, (bottom) 25 μm.

(I) Schematic of a gemmae, with dark blue square indicating imaged area.

(J–Q) Bar plots of chlorophyll content for the Mp*gata4*, Mp*gata2*, Mp*scl*, Mp*mir171*, Mp*hy5*, and Mp*glk* mutant plants. Individual values are shown with dots. Error bars represent the SEM; $n = 5$. 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 25th to the 75th 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; $n = 150$.

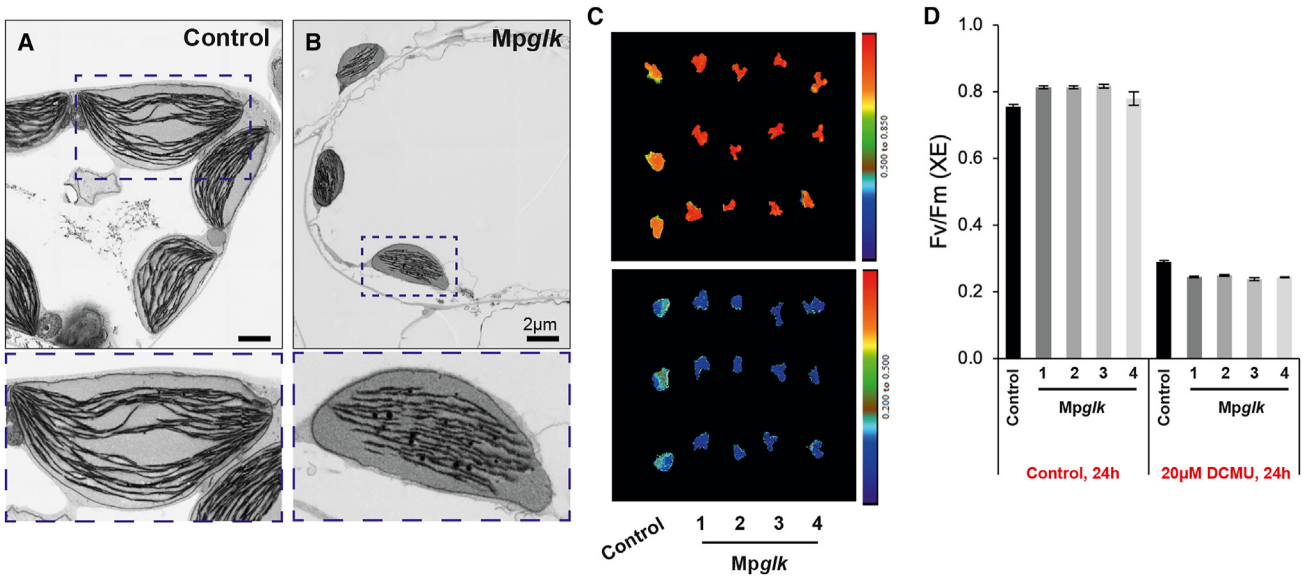


Figure 3. Mp*glk* controls chloroplast size and ultrastructure

(A and 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). Scale bars: 2 μm.

(C) Chlorophyll fluorescence images of maximum quantum efficiency of PSII (*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 SEM from $n = 3$ plants per genotype.

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 reduced chlorophyll,[36] our data argue against functional conservation between *A. thaliana 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, Mp*SCL* does not appear to control chlorophyll content. Again, this is in contrast to *A. thaliana*, where triple At*scl6,scl22,scl27* mutants showed increased chlorophyll, and At*MIR171* mimic lines, which are functional equivalents of At*mir171* knockouts, had reduced chlorophyll.[10,37,38] Although chloroplast sizes were unaltered in Mp*gata2*, Mp*gata4*, or Mp*mir171,* we found that two of the three Mp*scl* lines assessed had larger chloroplasts. In contrast, as in *A. thaliana*, Mp*hy5* mutants showed a conditional effect with lower chlorophyll at low temperatures, and Mp*glk* mutants accumulated less chlorophyll and had smaller chloroplasts than controls. Collectively, our results imply a strongly conserved function for GLK in chloroplast biogenesis in *M. polymorpha* and suggest that the conditional role of HY5 is also conserved.

Limited interplay between GATA4, GATA2, SCL, HY5, and GLK in *M. polymorpha*

It is possible that redundancy or compensatory responses mask effects on chloroplast biogenesis in the loss of function alleles for Mp*GATA4*, Mp*GATA2*, Mp*SCL*, and Mp*HY5*. To test this, we generated Mp*glk,gata4*, Mp*glk,gata2*, Mp*glk,scl*, and Mp*glk,hy5* double mutants (Figures 4A–4E and S5I). None were noticeably paler than the single Mp*glk* mutant (Figures 4A–4E), and chlorophyll content was similar (Figure 4F). In Mp*glk,gata4*, Mp*glk,scl*, and Mp*glk,hy5* double mutants, thallus morphology was perturbed in a manner similar to that seen in single Mp*gata4*, Mp*scl*, and Mp*hy5* mutants, respectively. Chloroplast size in double Mp*glk,gata4* mutants was also comparable to single Mp*glk* mutants (Figure 4G). Moreover, overexpression of Mp*GATA4* in Mp*glk* did not rescue its pale phenotype (Figures 4H and S5J). In *A. thaliana*, HY5 regulates *GLK* and *GNC/CGA1*,[20] and so we asked whether this was also the case for the homologs in *M. polymorpha*. We therefore subjected 6-day-old Mp*hy5* and control gemmalings to 48 h of darkness, followed by exposure to 1.5, 3, and 6 h of light. However, Mp*GLK* and Mp*GATA4* expression was comparable in Mp*hy5* and controls, providing no evidence for Mp*GLK* and Mp*GATA4* being targeted by MpHY5 (Figure 4I). In summary, the data provide no evidence for 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 promoter (Mp*UBE2*). To facilitate analysis and delineate borders of individual cells, EGFP was used to mark the plasma membrane (Figure 5A). No differences in chlorophyll or thallus morphology between plants with and without the plasma

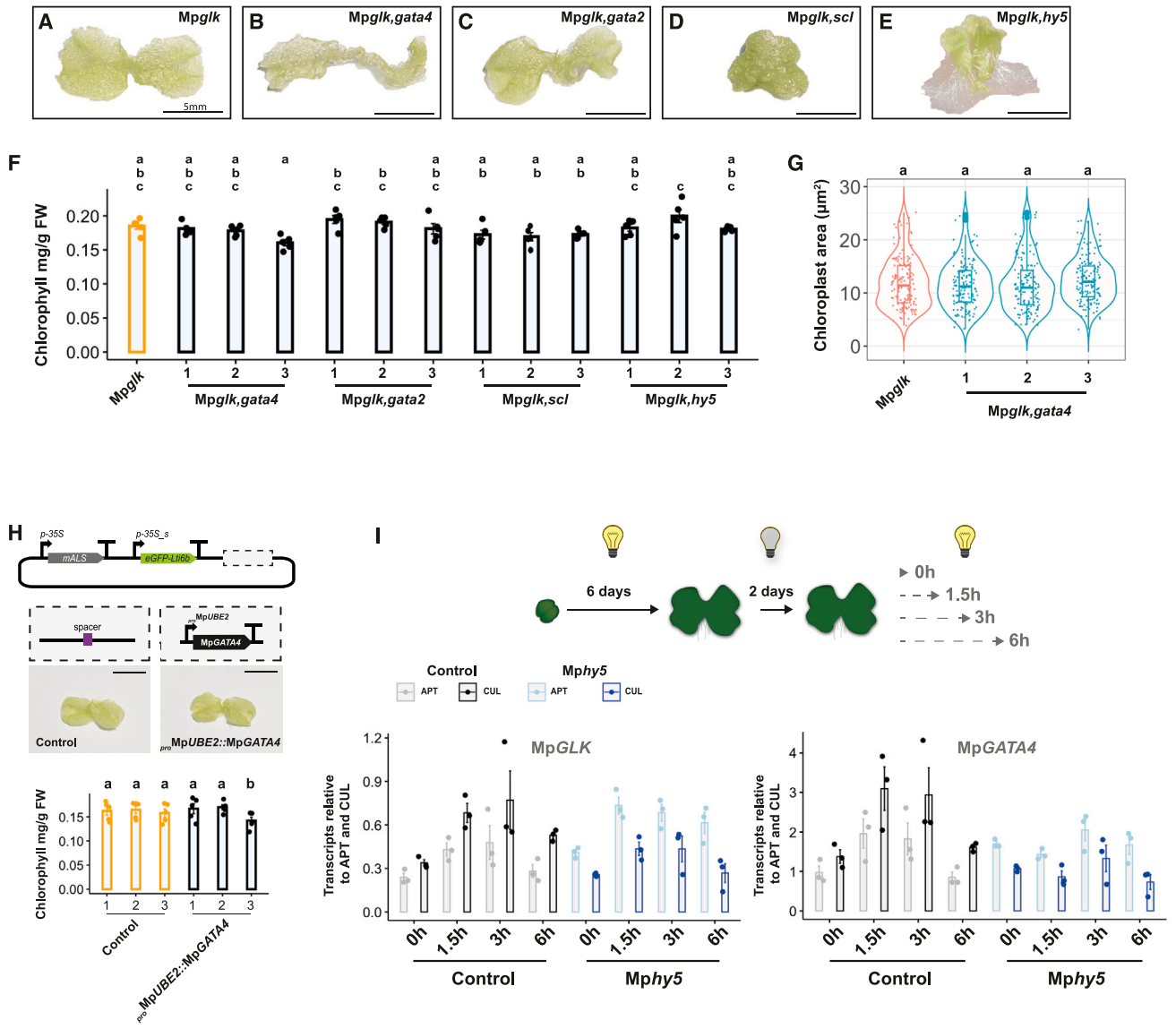


Figure 4. Mp*GATA4*, Mp*GATA2*, Mp*SCL*, and Mp*HY5* are not epistatic with Mp*GLK* for chloroplast biogenesis in *M. polymorpha*

(A–E) Representative images of (A) Mp*glk*, (B) Mp*glk,gata4*, (C) Mp*glk,gata2,* (D) Mp*glk,scl*, and (E) Mp*glk,hy5* double mutants.

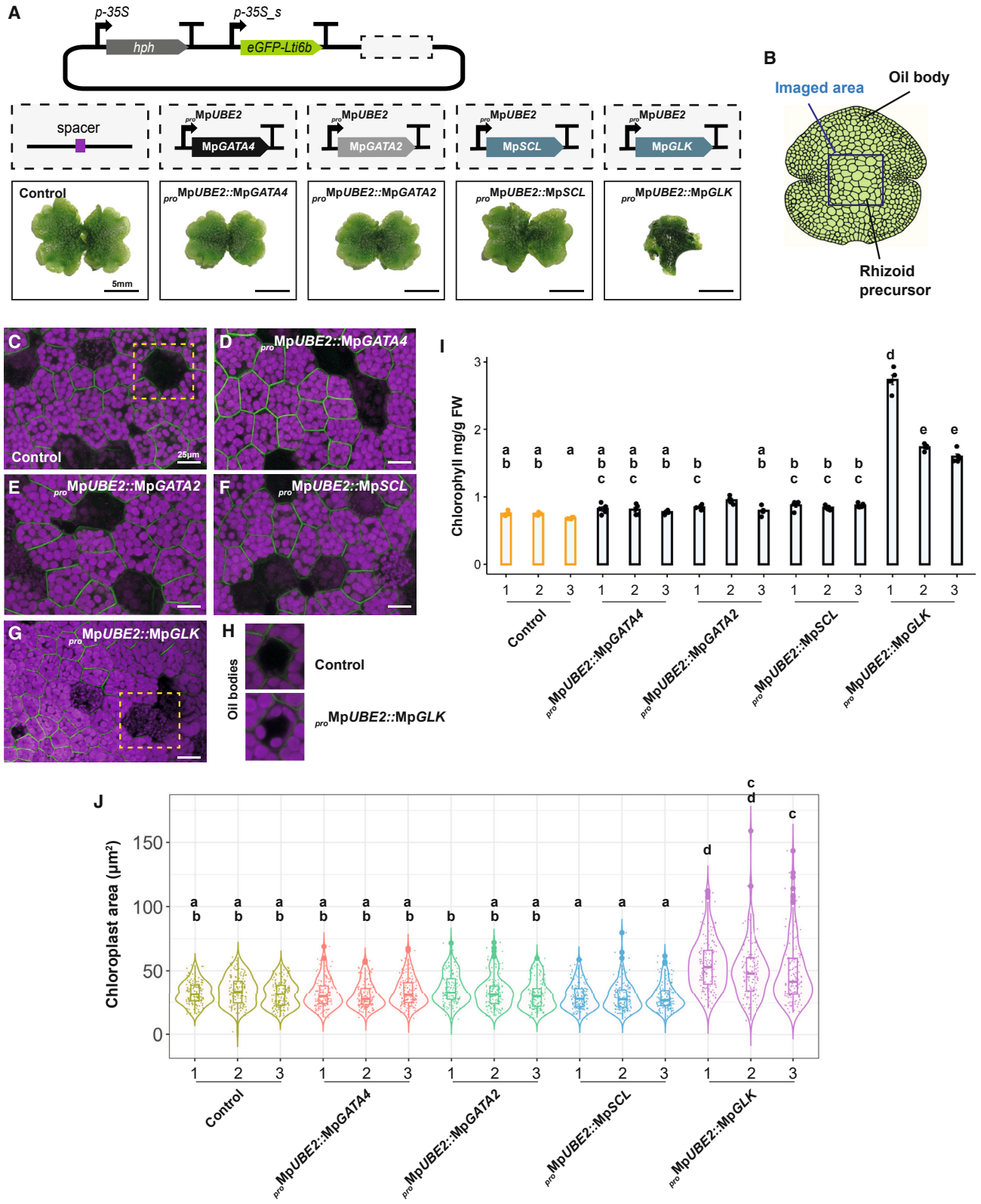
(F) Chlorophyll content in Mp*glk,* Mp*glk,gata4*, Mp*glk,gata2*, Mp*glk,scl*, and Mp*glk,hy5* double mutants. Letters show statistical ranking using a post hoc Tukey test; different letters indicate statistically significant differences at $p < 0.01$. Individual values are shown with dots. Error bars represent the SEM. 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 indicate 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*. Individual values are shown with dots. Error bars represent the SEM. Letters show statistical ranking using a post hoc Tukey test; different letters indicate 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: qPCR for analysis of Mp*GLK* and Mp*GATA4* expression in Mp*hy5* mutants in response to light. *ADENINE PHOSPHORIBOSYL TRANSFERASE 3* and *CULLIN 1* were used as housekeeping gene controls.[39] Individual values are shown with dots. Error bars represent the SEM.

membrane marker were evident (Figure S5K). qPCR confirmed that each transgene was overexpressed (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* (Figures 5B–5F, 5I, and 5J). Since in *A. thaliana, GATA2* overexpression led to constitutive photomorphogenesis,[19] we placed gemmae overexpressing Mp*GATA2* in the dark alongside controls and examined whether there were



(legend on next page)

any differences in growth. Overexpressors and controls were indistinguishable (gemmae did not germinate) (Figure S6A). However, under our standard growth conditions, plants in which Mp*GLK* was overexpressed were darker green and displayed stunted growth (Figure 5A). In fact, lines overexpressing Mp*GLK* contained up to ~4 times higher chlorophyll content (Figures 5G–5I). For Mp*GLK*, we tested two other options: Mp*GLK* fused to its own 3′ UTR and Mp*GLK re-written (*Mp*GLKrw)* (Data S6), where a synthetic nucleotide sequence encoded the wild-type amino acid sequence. Mp*GLKrw* and Mp*GLK&3′* UTR were used to test for regulatory elements in the Mp*GLK* coding sequence and 3′ UTR, respectively. We found that all three MpGLK versions led to increased chlorophyll content; however, native Mp*GLK* and Mp*GLKrw* overexpression led to increased chloroplast sizes, while overexpression of Mp*GLK&3′* UTR resulted in smaller chloroplasts compared to controls (Figures S6B–S6H). For each construct, depending on the line analyzed, there were variable but not consistent effects on chloroplast number per cell (Figures S6B–S6E). 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 able to activate chloroplast biogenesis (Figures 5J and S6B–S6H).

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

Although there were limited effects of knocking out *HY5*, *GATA4*, and *SCL* on chloroplast biogenesis in *M. polymorpha*, we undertook RNA sequencing (RNA-seq) on these lines along with those we generated for Mp*GLK*. Knocking out Mp*SCL*, Mp*GATA4* and Mp*HY5* had a moderate effect with 243, 332, and 160 genes being downregulated, respectively (p-adjusted value $\leq$ 0.01, log2 fold change [FC] $\geq$ 1-fold) (Figures S7A–S7C; Data S3). In loss-of-function mutants for Mp*GLK*, 1,065 genes showed a reduction in transcript abundance compared with controls (p-adjusted value $\leq$ 0.01, log2FC $\geq$ 1-fold) (Figure S7D; Data 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). Gene Ontology[40] analysis showed no specific enrichment in the case of the Mp*hy5* mutant. For Mp*scl*, Mp*gata4*, and Mp*glk* genotypes, oxidative stress as well as hydrogen peroxide catabolic processes were most impacted (Figure 6B). However, in Mp*glk* mutant, photosynthesis and chlorophyll biosynthesis process were also affected (Figure 6B), and comparing the effects of the loss of Mp*SCL*, Mp*GATA4*, Mp*HY5*, and Mp*GLK* function on chlorophyll biosynthesis and photosynthesis genes confirmed the greatest effect in Mp*glk*. For example, in Mp*scl*, only one chlorophyll biosynthesis gene, *CHLOROPHYLL A OXYGENASE*, was upregulated (log2FC ~0.57) (Figure 6C). In Mp*hy5* and Mp*gata4* mutants, there was a moderate reduction in transcript abundance (log2FC ~ −0.3 to 0.7) of 13 and 5 chlorophyll biosynthesis genes, respectively (Figure 6C). In contrast, Mp*GLK* mis-expression affected 17 of the 19 chlorophyll biosynthesis genes (Figure 6C).

We assessed the effects of Mp*scl*, Mp*gata4*, Mp*hy5*, and Mp*glk* on the 68 annotated genes belonging to the light harvesting complexes A and B (LHCA and LHCB), components of PSI and PSII and the cytochrome b_6f complex. Mp*SCL* loss led to the downregulation of only two genes (log2FC ~−0.37 to 0.55). Mp*HY5* or Mp*GATA4* loss resulted in a moderate reduction in transcript levels (log2FC ~−0.4 to 0.7) of 6 and 30 genes, respectively. The only exceptions were three and four *LHCB* genes where greater reductions (log2FC $\geq$1-fold) were apparent for the Mp*gata4* and Mp*hy5* mutants, respectively (Figure 6D). Consistent with the effects on chlorophyll described above, the Mp*glk* mutant had the greatest effect on transcript abundance of photosynthesis genes (Figures 6D and 6E), with 33 genes being downregulated (log2FC $\geq$1-fold). HY5 is known to regulate cell elongation and proliferation, pigment accumulation, and nutrient uptake.[6] When we examined whether homologs of genes known to control such processes in *A. thaliana* are misregulated in Mp*hy5* mutants, we only observed upregulation of a sulfate transporter (Figure S7E), a response opposite of that observed in *A. thaliana*. In conclusion, our RNA-seq analysis indicated that MpSCL, MpHY5, and MpGATA4 have a limited ability to control the expression of photosynthesis genes, and their impact does not appear sufficiently extensive to impact on chloroplast biogenesis.

Distinct and divergent binding patterns of MpGLK compared with flowering plants

To identify direct targets of MpGLK, we performed ChIP-seq using transgenic *M. polymorpha* expressing Mp*GLK* fused to a hemagglutinin (HA) tag (Figure 7A). Applying the data analysis pipeline utilized previously for five flowering plants (*A. thaliana*, tomato, rice, maize, and tobacco),[41] we detected a total of 2,654 reproducible peaks representing binding by MpGLK (irreproducibility discovery rate [IDR] < 0.01). Motif enrichment

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.

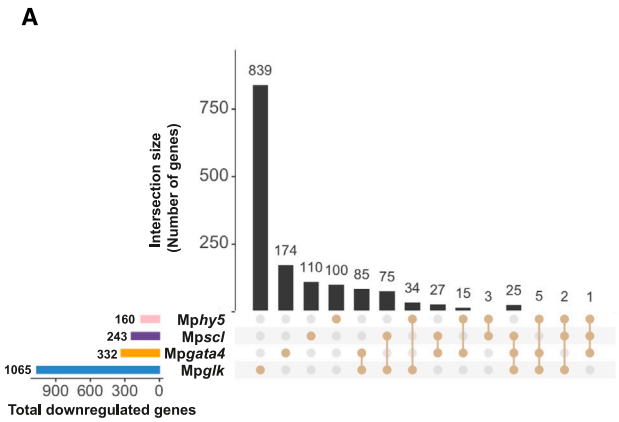
(B) Schematic of a gemmae; a dark blue square indicates imaged area; black lines indicate positions of rhizoid precursor and oil body cells.

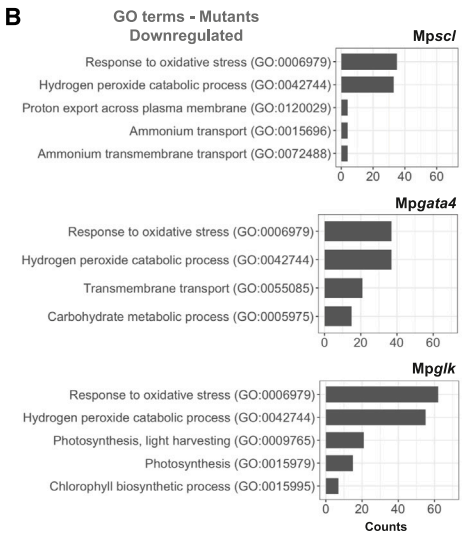
(C–G) Representative confocal microscopy images of gemmae from controls and plants overexpressing Mp*GATA4*, Mp*GATA2*, Mp*SCL*, and Mp*GLK*. Chlorophyll autofluorescence shown in magenta and plasma membrane marked with EGFP in green. Rhizoid precursors highlighted with dashed squares. Scale bars represent 25 μm.

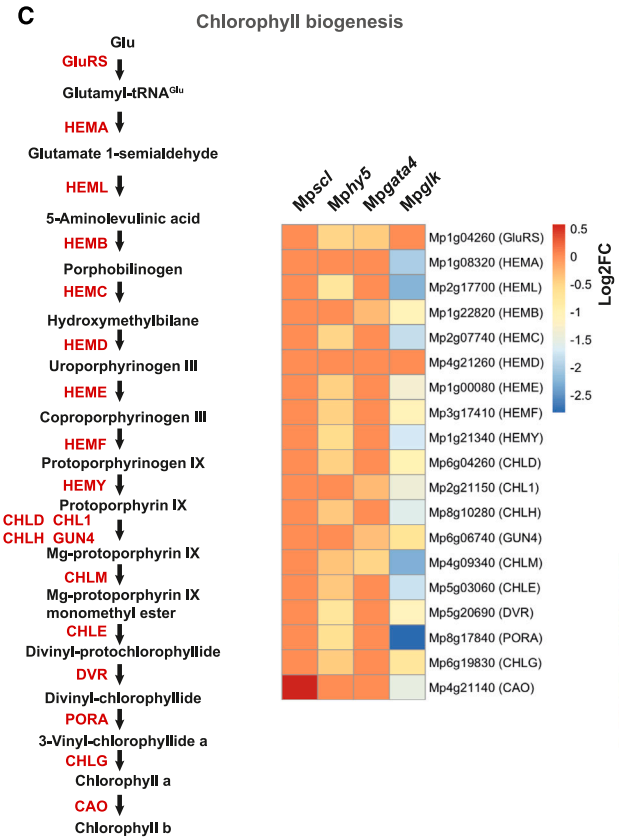
(H) Representative confocal microscopy images of oil body cells in control and Mp*GLK* overexpressing lines.

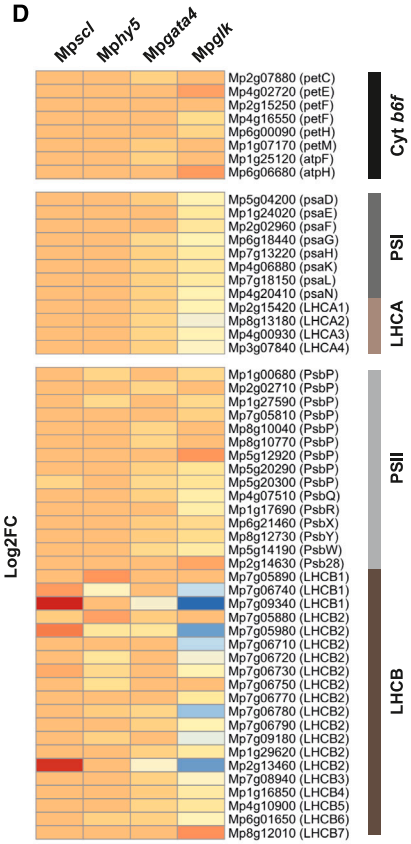
(I) Chlorophyll content in controls and overexpressing lines for Mp*GATA4*, Mp*GATA2*, Mp*SCL*, and Mp*GLK*. Individual values are shown with dots. Error bars represent the SEM. Letters show statistical ranking using a post hoc Tukey test (different letters indicate statistically significant differences at $p < 0.01$). Values indicated by the same letter are not statistically different; $n = 5$.

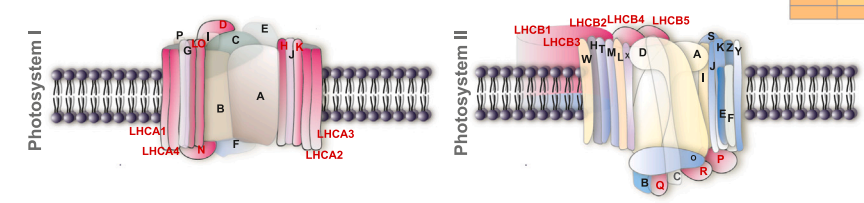
(J) Chloroplast area in controls and lines overexpressing Mp*GATA4*, Mp*GATA2*, Mp*SCL*, and Mp*GLK*. Letters show statistical ranking using a post hoc Tukey test (different letters indicate statistically significant differences at $p < 0.01$). Values indicated by the same letter are not statistically different; $n = 150$.











(legend on next page)

analysis of the summit regions of ChIP-seq peaks indicated that MpGLK binds an RGATTYY sequence motif, analogous to its orthologs in flowering plants (Figure 7B). We next associated these peaks with 1.5-kb upstream regions of annotated genes and found 1,326 targets (Data S4). We also performed an assay for transposase-accessible chromatin with sequencing (ATAC-seq) to identify accessible chromatin regions and carried out ChIP-seq for H3K4me3, a histone mark often associated with active genes near their transcription start site. MpGLK ChIP signals overlapped with that from ATAC-seq (Figures 7A and 7D), suggesting that MpGLK binds to open chromatin. Binding sites were also found upstream of the +1 nucleosome often marked by H3K4me3, a pattern consistent with the observation for *A. thaliana* GLK (Figures 7A and 7C).

Comparing ChIP-seq data with our gene expression data for the Mp*glk* mutants, we found that 40.4% (536/1,326) of GLK target genes were downregulated in the Mp*glk* mutant (Figure 7E) and 35% (471/1,326) were upregulated in Mp*GLK* overexpression lines. Of those affected in Mp*glk* mutants and Mp*GLK* overexpression lines (Figures S7D, S7F, and S7G), 269 genes overlapped. As has been reported for *A. thaliana* and tomato, the average binding strength of MpGLK inferred from the ChIP-seq signal FC at the peak summit was higher in differentially expressed genes (DEGs) compared with non-DEGs (Figure 7F). Strikingly, more MpGLK target genes showed differential expression in Mp*glk* mutants compared with their counterparts in flowering plants. For example, only 21% (202/960) of AtGLK1/2 targets were differentially expressed in the At*glk1,glk2* double mutant. This is consistent with a more pronounced chlorophyll loss in Mp*glk* mutants described here compared with that for At*glk1,glk2* mutant alleles.[11] In summary, our findings suggest a conserved role for GLK throughout the evolution of land plants. It is possible, of course, that transcriptional networks in flowering plants have accumulated changes resulting in increased redundancy and robustness.

The dynamic nature of *cis*-regulatory elements in plant genomes can give rise to divergence in TF binding and thus evolution of gene regulatory networks. For example, when GLK binding across flowering plants was compared, only 10%–20% of target genes were conserved between *A. thaliana,* tomato, tobacco, rice, and maize. To investigate this for MpGLK, we classified protein-coding genes in *M. polymorpha* and the aforementioned five species into ortholog groups.[42] Only 53 GLK target genes in 48 ortholog groups were found to be conserved, and they were mainly involved in photosynthesis (Data S4). For example, 23 genes encoding CHLOROPHYLL A/B BINDING PROTEIN were bound by MpGLK. Additionally, GLK bound CONSTANS and B-BOX TF genes in all six species (Figure 7G), suggesting that they play an ancient regulatory role in circadian rhythms and photomorphogenesis. However, it is important to highlight that most MpGLK targets were not conserved. For instance, *1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE* and *PYRUVATE ORTHOPHOSPHATE DIKINASE* were bound by GLK in *M. polymorpha*, but their orthologs in *A. thaliana* were not bound by AtGLK1/2 (Figures 7A and 7C). Notably, even some MpGLK targets encoding chloroplast-localized proteins did not exhibit conservation between these species (Figure 7H).

To further understand the ancestral function of GLK, we performed cross-species complementation experiments. Specifically, we tested whether *GLK* genes from the moss *P. patens*, the hornwort *Anthoceros agrestis*, and the green algae *Spirogloea muscicola,* which belongs to the sister to land plants Zygnematophyceae group, could complement the Mp*glk* mutant phenotype. *GLK* genes from all three species rescued the Mp*glk* mutant phenotype, with *GLK* from *S. muscicola* having the smallest effect (Figures 7I–7J and S7I). Overall, our findings provide support for the hypothesis that while GLK has maintained an important role in chloroplast biogenesis in land plants, the transcriptional network in which it operates has diverged significantly during the ~400–500 Ma since the split between bryophyte and flowering plants.

DISCUSSION

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

The monophyletic bryophyte group diverged from vascular plants approximately 400–500 mya.[21,23,25] 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 to which regulators of chloroplast biogenesis defined in angiosperms are functionally conserved.

Our analysis indicated that chlorophyll accumulation was not detectably perturbed in mutant alleles of Mp*gata4 or* Mp*gata2*. This contrasts with an ~30%–40% decrease in chlorophyll in *A. thaliana gnc/cga1* mutants[36] and compromised photomorphogenesis when At*GATA2* expression was suppressed.[19] In *A. thaliana GNC*, overexpression increased chlorophyll 10-fold in seedlings and ~30% in the leaf,[15] and *GATA2* overexpression led to constitutive photomorphogenesis.[19] The simplest and

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* mutants.

(B) Enriched Gene Ontology terms for Mp*scl*, Mp*gata4*, and Mp*glk* mutants.

(C) Heatmap illustrating extent of downregulation (log2FC) of transcripts encoding enzymes of chlorophyll biosynthesis in Mp*scl*, Mp*hy5*, Mp*gata4*, and Mp*glk* mutant alleles.

(D) Heatmap indicating lower transcript abundance (log2FC) of genes encoding components of PSI, PSII, LHCA, LHCB, and cytochrome b_6f complex in Mp*scl*, Mp*hy5*, Mp*gata4*, and Mp*glk* mutants.

(E) Schematic representation of PSI and PSII. Subunits showing an increase (log2FC ≥1-fold) in corresponding transcript levels are highlighted in pink. Genes that are differentially expressed are highlighted in red.

Figure modified from Waters et al.[11] and Tu et al.[41]

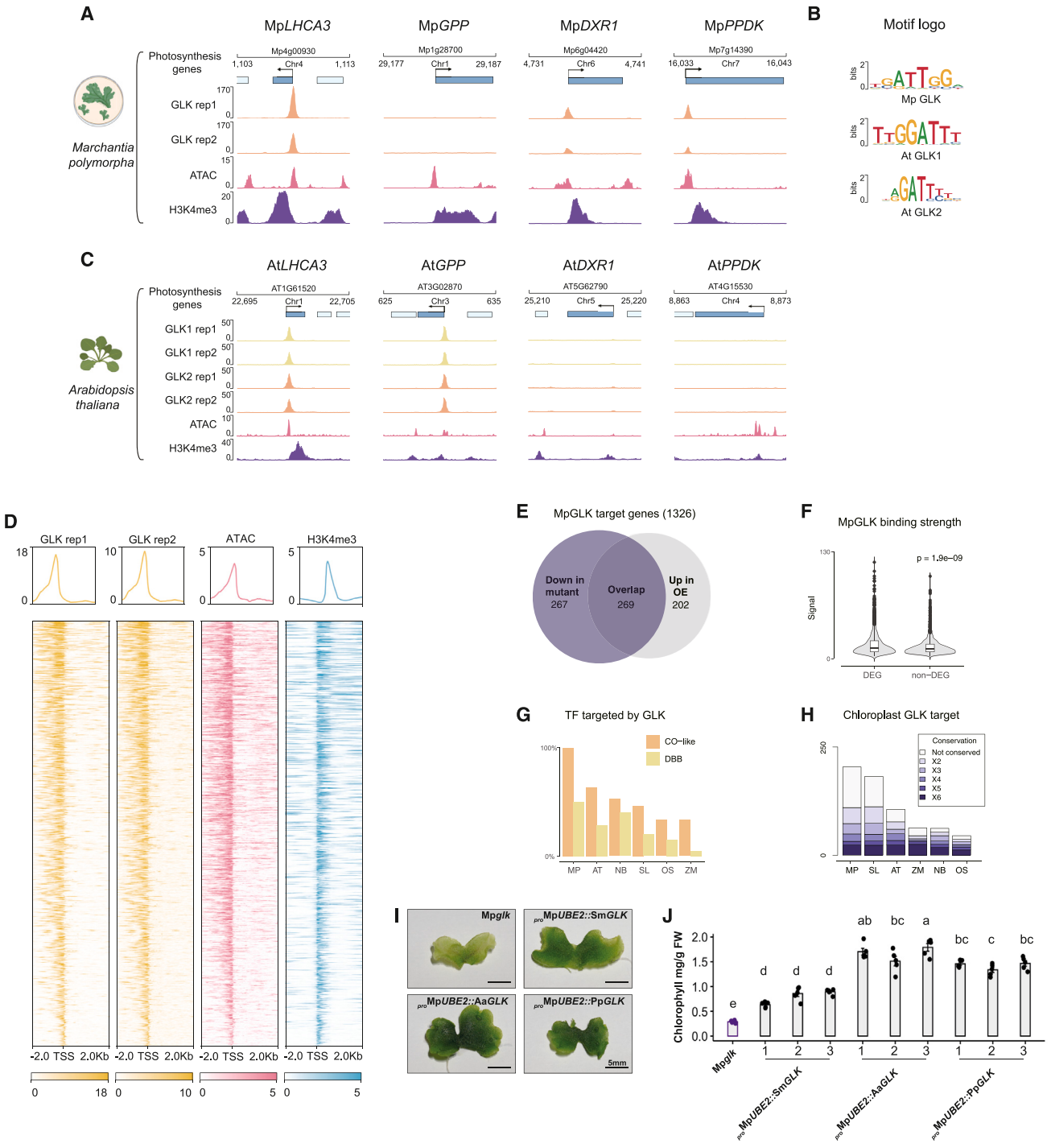


Figure 7. MpGLK ChIP-seq reveals divergence in GLK binding between *M. polymorpha* and flowering plants

(A) Genome browser tracks showing GLK ChIP-seq peaks, open chromatin regions (ATAC-seq), and H3K4me3 peaks in conserved and *M. polymorpha*-specific GLK target genes. *LHCA3: LIGTH HARVESTING COMPLEX A 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 *A. thaliana* orthologs of *M. polymorpha* genes in (A).

(D) Heatmap and average signal plot showing the chromatin features near the MpGLK binding sites. Clustering was performed using the GLK1 ChIP-seq rep1 signal.

(E) Partial overlap of MpGLK targets genes upregulated in *GLK* overexpression (OE) lines and downregulated in *GLK* knockout lines.

(legend continued on next page)

most parsimonious explanation for the lack of impact of the loss of Mp*GATA4* or Mp*GATA2* function or overexpression on chlorophyll content in bryophytes such as *M. polymorpha* is that in the ancestral state, these proteins do not play a role in chloroplast biogenesis. Mp*glk,gata4* and Mp*glk,gata2* double mutants had chlorophyll levels similar to those of Mp*glk* single mutants, implying lack of redundancy, but it is still possible that other proteins compensate, and so redundancy in the gene regulatory networks masks loss of function. It is also possible that GATA TFs regulate chloroplast biogenesis in some but not all bryophytes. This is perhaps supported by analysis of *P. patens*, where two overexpression lines for Pp*GATA1*, one of the two *GNC/CGA1* orthologs in this species, showed an ~10%–20% increase in chlorophyll content.[43] A similar effect was reported when Pp*GATA1* was mis-expressed in *A. thaliana*,[43] but these increases in chlorophyll content are modest compared with those reported in *A. thaliana* when *GNC* overexpression increased chlorophyll 10-fold in seedlings and ~30% in the leaf.[15] Moreover, while *GNC/CGA1* overexpression in *A. thaliana* led to an increase in chloroplast number per cell, this was not reported in *P. patens*.[43] A recent study in *M. polymorpha* proposed a role in response to high-light stress[44] for MpGATA4, with mutants showing mis-regulation of the genes encoding EARLY LIGHT-INDUCED PROTEINs (ELIPs). Our transcriptome analysis of Mp*gata4* mutants revealed that a single *ELIP* (Mp*ELIP-11*, Mp3g15080) was moderately downregulated (log2FC = −0.678) alongside a small number of genes encoding components of the LHCs and PSs. We also noted that editing Mp*GATA4* affected gametophyte morphology such that the 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 the 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* nor 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, triple At*scl6,scl22,scl27* mutants and *MIR171* overexpressors lead to increased chlorophyll accumulation in *A. thaliana*.[45] As no other members of the *SCL* gene family in *M. polymorpha* contain an MIR171 recognition site, this argues against Mp*MIR171* playing a role in the regulation of chlorophyll accumulation. It is possible that the *M. polymorpha* homolog of At*SCL* does regulate chlorophyll biosynthesis but without the additional control derived from MIR171 found in *A. thaliana*. SCL6, SCL22, and SCL27 also regulate the proliferation of meristematic cells.[46–48] However, if this were also the case in *M. polymorpha*, overexpression of Mp*SCL* should decrease chlorophyll content, as seen in *A. thaliana*.[10] However, our Mp*SCL* overexpression lines appeared similar to controls. Thus, in addition to GATA2 and GATA4 discussed above, it is possible that an ancestral role of SCL relates to the development of photosynthetic tissue rather than chloroplast biogenesis per se. For SCL6, SCL22, and SCL27, this function seems to have been retained in *A. thaliana*, and a role in repressing chlorophyll synthesis was acquired.

HY5 in *M. polymorpha* appears to play a conditional role in chloroplast biogenesis, as has been reported in *A. thaliana*, but the impact on gene expression relating to chloroplast biogenesis and photosynthesis was limited. Mp*hy5* mutants accumulated more sesquiterpenes than controls indicating that MpHY5 also likely plays an additional role in secondary metabolism. In land plants, HY5 also plays a role in responding to UV-B, a major abiotic stress. For example, in *A. thaliana*, UV-B activates the UV RESISTANCE LOCUS8 photoreceptor, which then enhances HY5 activity by preventing its degradation by E3 ubiquitin ligase CONSTITUTIVELY PHOTOMORPHOGENIC1-mediated ubiquitination. Consequently, HY5 regulates the transcription of numerous genes required for the UV-B response, including those involved in DNA repair, protection of photosynthetic machinery, and production of UV-B-absorbing phenylpropanoids.[49,50] It has previously been shown that this UV-B response in *M. polymorpha* includes the production of UV-B-absorbing flavonoids, and the induction of HY5 is similar to that in *A. thaliana*.[29] Thus, it appears that HY5 in *M. polymorpha* still responds to light signals, but in flowering plants it is integrated into a more complex regulatory network linking light signals with chloroplast biogenesis.

A major and conserved role for GLK in 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*. First, Mp*glk* mutants have reduced chlorophyll accumulation and smaller chloroplasts with underdeveloped thylakoids. These perturbations to phenotype have been observed in other land plants.[12–14,26,51] Second, constitutive overexpression of Mp*GLK* resulted in increased chlorophyll accumulation, larger chloroplasts, and ectopic chloroplast development similar to reports in angiosperms.[36,52–55] Third, 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.[11,41] ChIP-seq confirmed that most are photosynthesis-related genes such as the *LHCA*, *LHCB*, *PsbQ*, and genes encoding chlorophyll biosynthesis enzymes. However, we also uncovered major divergence in

(F) Stronger ChIP-seq signals were found in differentially expressed GLK targets.
(G) Percentage of *CONSTANS-Like* (*CO-like*) and *B-BOX* (*DBB*) TF genes in each genome targeted by GLK. Note that all three CO-Like TFs in *M. polymorpha* are GLK targets.
(H) Number of predicted chloroplast localized GLK targets conserved between *M. polymorpha* and five flowering plants.
(I) Images of thallus fragments for Mp*glk* mutant as well as Sm*GLK*, Aa*GLK*, and Pp*GLK* overexpressing lines.
(J) Chlorophyll content in Mp*glk*, Sm*GLK*, Aa*GLK*, and Pp*GLK* overexpressing lines. Individual values are shown with dots. Error bars represent the SEM. Letters show statistical ranking using a post hoc Tukey test (different letters indicate statistically significant differences at $p < 0.01$). Values indicated by the same letter are not statistically different; $n = 5$.

GLK-binding patterns associated with land plant evolution, as less than 5% of MpGLK target genes were conserved with those in flowering plants.

We also examined how Mp*GLK* is regulated at transcriptional and post-transcriptional levels. We asked whether there are any regulatory motifs within the Mp*GLK* coding sequence. To do so, we "re-wrote" the coding sequence such that the nucleotide sequence was changed but predicted the amino acid sequence preserved. Re-written Mp*GLK* did not increase chlorophyll content compared with plants overexpressing the native Mp*GLK* sequence. We believe that there are two possible explanations for this. First, the Mp*GLK* coding sequence does not contain nucleotide sequence motifs responsible for post-transcriptional regulation. Second, although we codon-optimized the re-written MpGLK for *M. polymorpha*, its expression may be less efficient than that of the native sequence. Finally, we found that although Mp*GLK* is negatively regulated by its 3′ UTR, it is unlikely to be regulated by a predicted microRNA/small interfering RNA cleavage site.[32] Instead, some as-yet unknown motif located between 400 and 671 of Mp*GLK* 3′ UTR appears to be involved.

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 specialization within the thallus. Angiosperms, however, have a more complex development and morphology, allowing colonization of a broad range of environments. For example, leaves with specialized 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 specializations to preexisting pathways present in the common ancestor of bryophytes and vascular plants. Furthermore, our cross-species complementation showed that *GLK* genes from the moss *P. patens*, the hornwort *A. agrestis*, and the green algae *S. muscicola*[56] rescued the reduced chlorophyll phenotype of the Mp*glk* loss-of-function mutant. A functional GLK homolog from the green algal Zygnematophyceae is indicative of a role for GLK in chloroplast biogenesis that predates the colonization of land.

Limitations of the study

We cannot exclude the possibility that genetic redundancy is masking an effect of GATAs on chloroplast biogenesis. The generation of higher-order mutants, such as Mp*gata4* combined with Mp*gata2*, would allow this hypothesis to be tested. Approaches such as ChIP-seq could also help define direct targets of GATAs in *M. polymorpha*, aiding our understanding of how these networks have been rewired during land plant evolution. A similar approach could be used to determine direct targets of Mp*HY5* to further delineate its role in *M. polymorpha*. For GLK, our ChIP-seq analysis revealed extensive divergence between targets in *M. polymorpha* and *A. thaliana*. To further assess the degree of GLK specialization and conservation during the evolution of vascular plants, this analysis could be expanded to include GLK homologs from vascular plants. The simplified genetic network for chloroplast biogenesis in *M. polymorpha* might represent either the ancestral state of land plants or a streamlined version resulting from the loss of certain genes specifically in liverworts. Such secondary loss of traits has been reported for *M. polymorpha* stomata and the genetic network underpinning their development.[57,58] Since the three bryophyte lineages diverged from each other about 400 mya, conducting further genetic studies in the model moss *P. patens* and the emerging hornwort model *A. agrestis* would help to determine more precisely the ancestral architecture of the chloroplast biogenesis gene network. Similarly, once tools like CRISPR-Cas9 technology are available for green algae, genetic studies could offer insights into the ancestral function of the TFs controlling chloroplast biogenesis, including *GLK* genes. Lastly, it would be interesting to determine whether GLKs from green algae and bryophytes can rescue the *glk* mutant phenotype in *A. thaliana*.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Julian M. Hibberd, jmh65@cam.ac.uk.

Materials availability

Materials generated in this study are available from the lead contact.

Data and code availability

- Transcriptome and ChIP-seq data are available in the NCBI Sequence Read Archive: PRJNA1039314 and PRJNA1043823.
- This study did not generate any new original codes.
- Any additional information required to reanalyze the data reported in this paper is available from the lead author upon request.

ACKNOWLEDGMENTS

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). Z.W. was supported by the Zhujiang postdoc fellowship. J.S.-W. was supported by the BBSRC DTP (BBSRC BB/X010899/1). For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license 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.

AUTHOR CONTRIBUTIONS

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

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY SUBJECT DETAILS
- METHOD DETAILS
  - Phylogenetic analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.114696.

Received: February 2, 2024
Revised: July 23, 2024
Accepted: August 13, 2024
Published: September 5, 2024

REFERENCES

1. Archibald, J.M. (2009). The Puzzle of Plastid Evolution. Curr. Biol. *19*, R81–R88. https://doi.org/10.1016/j.cub.2008.11.067.
2. Gould, S.B., Waller, R.F., and McFadden, G.I. (2008). Plastid evolution. Annu. Rev. Plant Biol. *59*, 491–517. https://doi.org/10.1146/annurev.arplant.59.032607.092915.
3. Cackett, L., Luginbuehl, L.H., Schreier, T.B., Lopez-Juez, E., and Hibberd, J.M. (2022). Chloroplast development in green plant tissues: the interplay between light, hormone, and transcriptional regulation. New Phytol. *233*, 2000–2016. https://doi.org/10.1111/nph.17839.
4. Jarvis, P., and López-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. Nat. Rev. Mol. Cell Biol. *14*, 787–802. https://doi.org/10.1038/nrm3702.
5. de Vries, J., and Gould, S.B. (2018). The monoplastidic bottleneck in algae and plant evolution. J. Cell Sci. *131*, jcs203414. https://doi.org/10.1242/jcs.203414.
6. Xiao, Y., Chu, L., Zhang, Y., Bian, Y., Xiao, J., and Xu, D. (2021). HY5: A Pivotal Regulator of Light-Dependent Development in Higher Plants. Front. Plant Sci. *12*, 800989. https://doi.org/10.3389/fpls.2021.800989.
7. Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., Tongprasit, W., Zhao, H., Lee, I., and Deng, X.W. (2007). Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. Plant Cell *19*, 731–749. https://doi.org/10.1105/tpc.106.047688.
8. Burko, Y., Gaillochet, C., Seluzicki, A., Chory, J., and Busch, W. (2020). Local HY5 Activity Mediates Hypocotyl Growth and Shoot-to-Root Communication. Plant Commun. *1*, 100078. https://doi.org/10.1016/j.xplc.2020.100078.
9. Oyama, T., Shimura, Y., and Okada, K. (1997). The Arabidopsis HY5 gene encodes a bZIP protein that regulates stimulus-induced development of root and hypocotyl. Genes Dev. *11*, 2983–2995. https://doi.org/10.1101/gad.11.22.2983.
10. Ma, Z., Hu, X., Cai, W., Huang, W., Zhou, X., Luo, Q., Yang, H., Wang, J., and Huang, J. (2014). Arabidopsis miR171-Targeted Scarecrow-Like Proteins Bind to GT cis-Elements and Mediate Gibberellin-Regulated Chlorophyll Biosynthesis under Light Conditions. PLoS Genet. *10*, e1004519. https://doi.org/10.1371/journal.pgen.1004519.
11. Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J., and Langdale, J.A. (2009). GLK transcription factors coordinate expression of the photosynthetic apparatus in Arabidopsis. Plant Cell *21*, 1109–1128. https://doi.org/10.1105/tpc.108.065250.
12. Waters, M.T., Moylan, E.C., and Langdale, J.A. (2008). GLK transcription factors regulate chloroplast development in a cell-autonomous manner. Plant J. *56*, 432–444. https://doi.org/10.1111/j.1365-313X.2008.03616.x.
13. Fitter, D.W., Martin, D.J., Copley, M.J., Scotland, R.W., and Langdale, J.A. (2002). *GLK*gene pairs regulate chloroplast development in diverse plant species. Plant J. *31*, 713–727. https://doi.org/10.1046/j.1365-313x.2002.01390.x.
14. Bravo-Garcia, A., Yasumura, Y., and Langdale, J.A. (2009). Specialization of the Golden2-like regulatory pathway during land plant evolution. New Phytol. *183*, 133–141. https://doi.org/10.1111/j.1469-8137.2009.02829.x.
15. Chiang, Y.-H., Zubo, Y.O., Tapken, W., Kim, H.J., Lavanway, A.M., Howard, L., Pilon, M., Kieber, J.J., and 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 Physiol. *160*, 332–348. https://doi.org/10.1104/pp.112.198705.
16. Hudson, D., Guevara, D., Yaish, M.W., Hannam, C., Long, N., Clarke, J.D., Bi, Y.-M., and Rothstein, S.J. (2011). GNC and CGA1 modulate chlorophyll biosynthesis and glutamate synthase (GLU1/Fd-GOGAT) expression in Arabidopsis. PLoS One *6*, e26765. https://doi.org/10.1371/journal.pone.0026765.
17. Naito, T., Kiba, T., Koizumi, N., Yamashino, T., and Mizuno, T. (2007). Characterization of a unique GATA family gene that responds to both light and cytokinin in Arabidopsis thaliana. Biosci. Biotechnol. Biochem. *71*, 1557–1560. https://doi.org/10.1271/bbb.60692.
18. Bi, Y.-M., Zhang, Y., Signorelli, T., Zhao, R., Zhu, T., and Rothstein, S. (2005). Genetic analysis of Arabidopsis GATA transcription factor gene family reveals a nitrate-inducible member important for chlorophyll synthesis and glucose sensitivity. Plant J. *44*, 680–692. https://doi.org/10.1111/j.1365-313x.2005.02568.x.
19. Luo, X.-M., Lin, W.-H., Zhu, S., Zhu, J.-Y., Sun, Y., Fan, X.-Y., Cheng, M., Hao, Y., Oh, E., Tian, M., et al. (2010). Integration of light- and brassinosteroid-signaling pathways by a GATA transcription factor in Arabidopsis. Dev. Cell *19*, 872–883. https://doi.org/10.1016/j.devcel.2010.10.023.
20. 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., and Wang, P. (2024). GLK transcription factors accompany ELONGATED HYPOCOTYL5 to orchestrate light-induced seedling development in Arabidopsis. Plant Physiol. *194*, 2400–2421. https://doi.org/10.1093/plphys/kiae002.
21. Li, F.-W., Nishiyama, T., Waller, M., Frangedakis, E., Keller, J., Li, Z., Fernandez-Pozo, N., Barker, M.S., Bennett, T., Blázquez, M.A., et al. (2020). Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. Nat. Plants *6*, 259–272. https://doi.org/10.1038/s41477-020-0618-2.
22. de Vries, J., and Archibald, J.M. (2018). Plant evolution: landmarks on the path to terrestrial life. New Phytol. *217*, 1428–1434. https://doi.org/10.1111/nph.14975.
23. Harris, B.J., Clark, J.W., Schrempf, D., Szöllősi, G.J., Donoghue, P.C.J., Hetherington, A.M., and Williams, T.A. (2022). Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. Nat. Ecol. Evol. *6*, 1634–1643. https://doi.org/10.1038/s41559-022-01885-x.
24. Qiu, Y.-L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., et al. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Natl. Acad. Sci. USA *103*, 15511–15516. https://doi.org/10.1073/pnas.0603335103.
25. One Thousand Plant Transcriptomes Initiative (2019). One thousand plant transcriptomes and the phylogenomics of green plants. Nature *574*, 679–685. https://doi.org/10.1038/s41586-019-1693-2.
26. Yasumura, Y., Moylan, E.C., and Langdale, J.A. (2005). A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. Plant Cell *17*, 1894–1907. https://doi.org/10.1105/tpc.105.033191.

27. 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., et al. (2022). The renaissance and enlightenment of Marchantia as a model system. Plant Cell *34*, 3512–3542. https://doi.org/10.1093/plcell/koac219.

28. Sauret-Güeto, S., Frangedakis, E., Silvestri, L., Rebmann, M., Tomaselli, M., Markel, K., Delmans, M., West, A., Patron, N.J., and Haseloff, J. (2020). Systematic Tools for Reprogramming Plant Gene Expression in a Simple Model. ACS Synth. Biol. *9*, 864–882. https://doi.org/10.1021/acssynbio.9b00511.

29. 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., and Davies, K.M. (2018). UVR8-mediated induction of flavonoid biosynthesis for UVB tolerance is conserved between the liverwort Marchantia polymorpha and flowering plants. Plant J. *96*, 503–517. https://doi.org/10.1111/tpj.14044.

30. Reyes, J.C., Muro-Pastor, M.I., and Florencio, F.J. (2004). The GATA family of transcription factors in Arabidopsis and rice. Plant Physiol. *134*, 1718–1732. https://doi.org/10.1104/pp.103.037788.

31. Behringer, C., and Schwechheimer, C. (2015). B-GATA transcription factors – insights into their structure, regulation, and role in plant development. Front. Plant Sci. *6*, 90. https://doi.org/10.3389/fpls.2015.00090.

32. Lin, S.-S., and Bowman, J.L. (2018). MicroRNAs in Marchantia polymorpha. New Phytol. *220*, 409–416. https://doi.org/10.1111/nph.15294.

33. Safi, A., Medici, A., Szponarski, W., Ruffel, S., Lacombe, B., and Krouk, G. (2017). The world according to GARP transcription factors. Curr. Opin. Plant Biol. *39*, 159–167. https://doi.org/10.1016/j.pbi.2017.07.006.

34. Toledo-Ortiz, G., Johansson, H., Lee, K.P., Bou-Torrent, J., Stewart, K., Steel, G., Rodríguez-Concepción, M., and Halliday, K.J. (2014). The HY5-PIF regulatory module coordinates light and temperature control of photosynthetic gene transcription. PLoS Genet. *10*, e1004416. https://doi.org/10.1371/journal.pgen.1004416.

35. Trebst, A. (2007). Inhibitors in the functional dissection of the photosynthetic electron transport system. Photosynth. Res. *92*, 217–224. https://doi.org/10.1007/s11120-007-9213-x.

36. Zubo, Y.O., Blakley, I.C., Franco-Zorrilla, J.M., Yamburenko, M.V., Solano, R., Kieber, J.J., Loraine, A.E., and Schaller, G.E. (2018). Coordination of Chloroplast Development through the Action of the GNC and GLK Transcription Factor Families. Plant Physiol. *178*, 130–147. https://doi.org/10.1104/pp.18.00414.

37. Todesco, M., Rubio-Somoza, I., Paz-Ares, J., and Weigel, D. (2010). A collection of target mimics for comprehensive analysis of microRNA function in Arabidopsis thaliana. PLoS Genet. *6*, e1001031. https://doi.org/10.1371/journal.pgen.1001031.

38. Wang, L., Mai, Y.X., Zhang, Y.C., Luo, Q., and Yang, H.Q. (2010). MicroRNA171c-targeted SCL6-II, SCL6-III, and SCL6-IV genes regulate shoot branching in Arabidopsis. Mol. Plant *3*, 794–806. https://doi.org/10.1093/mp/ssq042.

39. Saint-Marcoux, D., Proust, H., Dolan, L., and Langdale, J.A. (2015). Identification of Reference Genes for Real-Time Quantitative PCR Experiments in the Liverwort Marchantia polymorpha. PLoS One *10*, e0118678. https://doi.org/10.1371/journal.pone.0118678.

40. Kawamura, S., Romani, F., Yagura, M., Mochizuki, T., Sakamoto, M., Yamaoka, S., Nishihama, R., Nakamura, Y., Yamato, K.T., Bowman, J.L., et al. (2022). MarpolBase Expression: A Web-Based, Comprehensive Platform for Visualization and Analysis of Transcriptomes in the Liverwort Marchantia polymorpha. Plant Cell Physiol. *63*, 1745–1755. https://doi.org/10.1093/pcp/pcac129.

41. Tu, X., Ren, S., Shen, W., Li, J., Li, Y., Li, C., Li, Y., Zong, Z., Xie, W., Grierson, D., et al. (2022). Limited conservation in cross-species comparison of GLK transcription factor binding suggested wide-spread cistrome divergence. Nat. Commun. *13*, 7632. https://doi.org/10.1038/s41467-022-35438-4.

42. Emms, D.M., and Kelly, S. (2022). SHOOT: phylogenetic gene search and ortholog inference. Genome Biol. *23*, 85. https://doi.org/10.1186/s13059-022-02652-8.

43. Luan, J., Ju, J., Li, X., Wang, X., Tan, Y., and 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. https://doi.org/10.1016/j.gene.2022.147103.

44. Schröder, P., Hsu, B.-Y., Gutsche, N., Winkler, J.B., Hedtke, B., Grimm, B., and Schwechheimer, C. (2023). B-GATA factors are required to repress high-light stress responses in Marchantia polymorpha and Arabidopsis thaliana. Plant Cell Environ. *46*, 2376–2390. https://doi.org/10.1111/pce.14629.

45. Wang, L., Mai, Y.-X., Zhang, Y.-C., Luo, Q., and Yang, H.-Q. (2010). MicroRNA171c-targeted SCL6-II, SCL6-III, and SCL6-IV genes regulate shoot branching in Arabidopsis. Mol. Plant *3*, 794–806. https://doi.org/10.1093/mp/ssq042.

46. Bolle, C. (2004). The role of GRAS proteins in plant signal transduction and development. Planta *218*, 683–692. https://doi.org/10.1007/s00425-004-1203-z.

47. Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science *297*, 2053–2056. https://doi.org/10.1126/science.1076311.

48. Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Prediction of plant microRNA targets. Cell *110*, 513–520. https://doi.org/10.1016/s0092-8674(02)00863-2.

49. Ulm, R., Baumann, A., Oravecz, A., Máté, Z., Adám, E., Oakeley, E.J., Schäfer, E., and Nagy, F. (2004). Genome-wide analysis of gene expression reveals function of the bZIP transcription factor HY5 in the UV-B response of Arabidopsis. Proc. Natl. Acad. Sci. USA *101*, 1397–1402. https://doi.org/10.1073/pnas.0308044100.

50. Binkert, M., Kozma-Bognár, L., Terecskei, K., De Veylder, L., Nagy, F., and Ulm, R. (2014). UV-B-responsive association of the Arabidopsis bZIP transcription factor ELONGATED HYPOCOTYL5 with target genes, including its own promoter. Plant Cell *26*, 4200–4213. https://doi.org/10.1105/tpc.114.130716.

51. Rossini, L., Cribb, L., Martin, D.J., and Langdale, J.A. (2001). The maize golden2 gene defines a novel class of transcriptional regulators in plants. Plant Cell *13*, 1231–1244. https://doi.org/10.1105/tpc.13.5.1231.

52. Kobayashi, K., Sasaki, D., Noguchi, K., Fujinuma, D., Komatsu, H., Kobayashi, M., Sato, M., Toyooka, K., Sugimoto, K., Niyogi, K.K., et al. (2013). Photosynthesis of root chloroplasts developed in Arabidopsis lines overexpressing GOLDEN2-LIKE transcription factors. Plant Cell Physiol. *54*, 1365–1377. https://doi.org/10.1093/pcp/pct086.

53. Nakamura, H., Muramatsu, M., Hakata, M., Ueno, O., Nagamura, Y., Hirochika, H., Takano, M., and Ichikawa, H. (2009). Ectopic overexpression of the transcription factor OsGLK1 induces chloroplast development in non-green rice cells. Plant Cell Physiol. *50*, 1933–1949. https://doi.org/10.1093/pcp/pcp138.

54. Wang, P., Khoshravesh, R., Karki, S., Tapia, R., Balahadia, C.P., Bandyopadhyay, A., Quick, W.P., Furbank, R., Sage, T.L., and Langdale, J.A. (2017). Re-creation of a Key Step in the Evolutionary Switch from C to C Leaf Anatomy. Curr. Biol. *27*, 3278–3287.e6. https://doi.org/10.1016/j.cub.2017.09.040.

55. 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., et al. (2012). Uniform ripening encodes a Golden 2-like transcription factor regulating tomato fruit chloroplast development. Science *336*, 1711–1715. https://doi.org/10.1126/science.1222218.

56. Cheng, S., Xian, W., Fu, Y., Marin, B., Keller, J., Wu, T., Sun, W., Li, X., Xu, Y., Zhang, Y., et al. (2019). Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant Evolution. Cell *179*, 1057–1067.e14. https://doi.org/10.1016/j.cell.2019.10.019.

57. Harris, B.J., Harrison, C.J., Hetherington, A.M., and Williams, T.A. (2020). Phylogenomic Evidence for the Monophyly of Bryophytes and the Reductive Evolution of Stomata. Curr. Biol. *30*, 2001–2012.e2. https://doi.org/10.1016/j.cub.2020.03.048.

58. Rich, M.K., and Delaux, P.-M. (2020). Plant Evolution: When Arabidopsis Is More Ancestral Than Marchantia. Curr. Biol. *30*, R642–R644. https://doi.org/10.1016/j.cub.2020.04.077.

59. Delmans, M., Pollak, B., and Haseloff, J. (2017). MarpoDB: An Open Registry for Marchantia Polymorpha Genetic Parts. Plant Cell Physiol. *58*, e5. https://doi.org/10.1093/pcp/pcw201.

60. Waller, M., Frangedakis, E., Marron, A.O., Sauret-Güeto, S., Rever, J., Sabbagh, C.R.R., Hibberd, J.M., Haseloff, J., Renzaglia, K.S., and Szövényi, P. (2023). An optimized transformation protocol for Anthoceros agrestis and three more hornwort species. Plant J. *114*, 699–718. https://doi.org/10.1111/tpj.16161.

61. Sugano, S.S., Nishihama, R., Shirakawa, M., Takagi, J., Matsuda, Y., Ishida, S., Shimada, T., Hara-Nishimura, I., Osakabe, K., and Kohchi, T. (2018). Efficient CRISPR/Cas9-based genome editing and its application to conditional genetic analysis in Marchantia polymorpha. PLoS One *13*, e0205117. https://doi.org/10.1371/journal.pone.0205117.

62. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

63. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321. https://doi.org/10.1093/sysbio/syq010.

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

65. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

66. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

67. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274. https://doi.org/10.1093/molbev/msu300.

68. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589. https://doi.org/10.1038/nmeth.4285.

69. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527. https://doi.org/10.1038/nbt.3519.

70. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

71. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

72. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

73. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

74. Zheng, Y., Jiao, C., Sun, H., Rosli, H.G., Pombo, M.A., Zhang, P., Banf, M., Dai, X., Martin, G.B., Giovannoni, J.J., et al. (2016). iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. Mol. Plant *9*, 1667–1670. https://doi.org/10.1016/j.molp.2016.09.014.

75. Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., and Gao, G. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. *45*, D1040–D1045. https://doi.org/10.1093/nar/gkw982.

76. 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., et al. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. Nat. Plants *4*, 460–472. https://doi.org/10.1038/s41477-018-0188-8.

77. Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings Bioinf. *20*, 1160–1166. https://doi.org/10.1093/bib/bbx108.

78. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. *35*, 518–522. https://doi.org/10.1093/molbev/msx281.

79. Yelina, N.E., Holland, D., Gonzalez-Jorge, S., Hirsz, D., Yang, Z., and 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*, jkac105. https://doi.org/10.1093/g3journal/jkac105.

80. Schreier, T.B., Müller, K.H., Eicke, S., Faulkner, C., Zeeman, S.C., and Hibberd, J.M. (2024). Plasmodesmal connectivity in C4 Gynandropsis gynandra is induced by light and dependent on photosynthesis. New Phytol. *241*, 298–313. https://doi.org/10.1111/nph.19343.

81. Kumar, S., Kempinski, C., Zhuang, X., Norris, A., Mafu, S., Zi, J., Bell, S.A., Nybo, S.E., Kinison, S.E., Jiang, Z., et al. (2016). Molecular Diversity of Terpene Synthases in the Liverwort Marchantia polymorpha. Plant Cell *28*, 2632–2650. https://doi.org/10.1105/tpc.16.00062.

82. Adams, R.P. (2007). Identification of Essential Oil Components by Gas Chromatography/mass Spectrometry (Gruver, TX USA: Texensis Publishing).

83. Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., et al. (2017). Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. Cell *171*, 287–304.e15. https://doi.org/10.1016/j.cell.2017.09.030.

84. 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., and Zhong, S. (2020). Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. Nat. Commun. *11*, 5089. https://doi.org/10.1038/s41467-020-18832-8.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-HA Tag Antibody	Invitrogen	Cat#26183; RRID: AB_10978021
Bacterial and virus strains		
E. coli DH5α chemically competent cells	Widely distributed	N/A
Agrobacterium tumefaciens GV3101	Widely distributed	N/A
Chemicals, peptides, and recombinant proteins		
Gamborg B5 medium including vitamins	DUCHEFA	Cat#G0210.0050
DMSO	Sigma-Aldrich	Cat#D8418
KOD DNA Polymerase	Sigma-Aldrich	Cat#71085
DCMU	Sigma-Aldrich	Cat#45463
Kanamycin	Melford	Cat#K22000–1.0
SuperScript™ IV First-Strand Synthesis System	Invitrogen	Cat#18091050
SYBR Green JumpStart *Taq* Ready Mix	Sigma-Aldrich	Cat#S4438
Spectinomycin	MERCK	Cat#S4014-5G
Chloramphenicol	Sigma-Aldrich	Cat#C0378
X-GlcA	Melford	Cat#MB1021
Chlorsulfuron	Sigma-Aldrich	Cat#34322
Hygromycin B	Melford	Cat#H7502
Cefotaxime sodium salt	Sigma-Aldrich	Cat#C7039
Agar capsules	Melford	Cat#A20021
Plant agar	PHYTOTECH LABS	Cat#A296
Dodecane	Sigma-Aldrich	Cat#297879
Dynabeads Protein G	ThermoFisher	Cat#10003D
Critical commercial assays		
Turbo DNA-free kit	Invitrogen	Cat#AM1907
RNeasy Plant kit	QIAGEN	Cat#74903
Qiagen MinElute	QIAGEN	Cat #28004
Deposited data		
Raw RNA sequencing data: *Marchantia polymorpha*	This paper	NCBI Sequence Read Archive (SRA): BioProject ID: PRJNA1039314
Raw ChIP sequencing data: *Marchantia polymorpha*	This paper	NCBI Sequence Read Archive (SRA): BioProject ID: PRJNA1043823
Experimental models: Organisms/strains		
Arabidopsis thaliana; Col-0 (wild-type)	Widely distributed	N/A
Marchantia polymorpha, Cam-1 and Cam-2	Delmans et al.[59]	N/A
Marchantia polymorpha, Tak-1 and Tak-2	Widely distributed	N/A
Oligonucleotides		
See Data S5	This paper	N/A
Recombinant DNA		
Plasmid: OP-023 CDS12-eGFP	Sauret-Gueto et al.[28]	N/A
Plasmid: OP-020 CDS_hph	Sauret-Gueto et al.[28]	N/A
Plasmid: OP-062 L1_CsR-Ck1	Sauret-Gueto et al.[28]	N/A
Plasmid: OP-037 CTAG_Lti6b	Sauret-Gueto et al.[28]	N/A
Plasmid: OP-054 3TER _Nos-35S	Sauret-Gueto et al.[28]	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: OP-049 PROM_35S	Sauret-Gueto et al.[28]	N/A
Plasmid: OP-47 PROM_MpUBE2	Sauret-Gueto et al.[28]	N/A
Plasmid: OP-48 5UTR_MpUBE2	Sauret-Gueto et al.[28]	N/A
Plasmid: OP-063 L1_HyR-Ck1	Sauret-Gueto et al.[28]	N/A
Plasmid: OP073 L1_Cas9-Ck4	Sauret-Gueto et al.[28]	N/A
Plasmid: OP-076 L2_lacZgRNA-Cas9-CsA	Sauret-Gueto et al.[28]	N/A
Plasmid: OP-075 L1_lacZgRNA-Ck2	Sauret-Gueto et al.[28]	N/A
Plasmid: OP-074 L1_lacZgRNA-Ck3	Sauret-Gueto et al.[28]	N/A
Plasmid: L1_35S_s:eGFP-Lti6b	Waller et al.[60]	N/A
Plasmid: pMpGE013	Sugano et al.[61]	Addgene, Plasmid #108681
Plasmid: p35S::Hyg-T35S– 35S:eGFP-Lti6b	This paper	N/A
Plasmid: p35S::Hyg-T35S– 35S:eGFP-Lti6b - pMpUBE2:MpGLKrw-T35SNOS	This paper	N/A
Plasmid: p35S::Hyg-T35S– 35S:eGFP-Lti6b - pMpUBE2:MpGLK-T35SNOS	This paper	N/A
Plasmid: p35S::Hyg-T35S– 35S:eGFP-Lti6b - pMpUBE2:MpGLK+3UTR-T35SNOS	This paper	N/A
p35S:mALS-T35S– 35S:eGFP-Lti6b - pMpUBE2:MpGATA4-T35SNOS	This paper	N/A
p35S: mALS -T35S– 35S:eGFP-Lti6b - pMpUBE2:MpGATA2 or MpSCL-T35SNOS	This paper	N/A
p35S: mALS -T35S- pMpUBE2:PpGLK-T35SNOS	This paper	N/A
p35S: mALS -T35S- pMpUBE2:AaGLK-T35SNOS	This paper	N/A
p35S: mALS -T35S- pMpUBE2:SmGLK-T35SNOS	This paper	N/A
Plasmid sequences are deposited on Mendeley	This paper	https://doi.org/10.17632/d3vk6b4krm.1
Software and algorithms		
UpSet	N/A	RRID:SCR_022731; https://upset.app/
ImageJ (Fiji), Version: 2.1.0-rc-62/1.53 s	N/A	RRID:SCR_003070; https://imagej.net/Fiji/
Rstudio, Version 4.2.1	N/A	RRID:SCR_000432; https://www.rstudio.com/
FastQC	N/A	RRID:SCR_014583; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
DESeq2	Love et al.[62]	RRID:SCR_015687; https://bioconductor.org/packages/release/bioc/html/DESeq2.html
SH-aLRT	Guindon et al.[63]	N/A
iTOL	Letunic & Bork.[64]	RRID:SCR_018174; https://itol.embl.de
MAFFT	Katoh & Standley.[65]	RRID:SCR_011811; https://mafft.cbrc.jp/alignment/software/
TrimAl	Capella-Gutiérrez et al.[66]	RRID:SCR_017334; http://trimal.cgenomics.org
iQTree	Nguyen et al.[67]	RRID:SCR_017254; http://www.iqtree.org
ModelFinder	Kalyaanamoorthy et al.[68]	http://www.iqtree.org/ModelFinder/
Kallisto	Bray et al.[69]	RRID:SCR_016582; https://pachterlab.github.io/kallisto/
bowtie2	Langmead and Salzberg.[70]	RRID:SCR_016368; https://github.com/BenLangmead/bowtie2/releases
MACS2	Zhang et al.[71]	RRID:SCR_013291; https://github.com/macs3-project/MACS
HOMER	N/A	RRID:SCR_01088; http://homer.ucsd.edu/homer/motif/
SAMtools	Danecek et al.[72]	RRID:SCR_002105; https://github.com/samtools/samtools/releases/
OrthoFinder	Emms and Kelly[73]	RRID:SCR_017118; https://github.com/davidemms/OrthoFinder

EXPERIMENTAL MODEL AND STUDY SUBJECT DETAILS

Marchantia polymorpha accessions Cam-1 (male) and Cam-2 (female) were used in this study[59] with the exception of ChIP-SEQ and ATAC- SEQ for which *Marchantia polymorpha* accessions Tak-1 (male) and Tak-2 (female) accessions were used.

METHOD DETAILS

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 (Data S1) were mined from the iTAK[74] and PlantTFDB databases,[75] Phytozome, Fernbase,[76] Phycozome and PhytoPlaza. Sequences for each individual species were aligned with the AtGNC and AtCGA1 amino acid sequences using MAFFT.[77] 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.[31] 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),[76] hornwort genome (www.hornworts.uzh.ch),[21] 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 two approaches were checked against Orthofinder searches.[73] The identified GATA protein sequences were aligned using MAFFT. Alignments were then trimmed using TrimAl.[66] A maximum likelihood phylogenetic tree was inferred using iQTree,[67] ModelFinder[68] and ultrafast approximation for phylogenetic bootstrap[78] and SH-aLRT test.[63] The tree was visualised using iTOL.[64]

To identify *GLK* genes, three different approaches were combined: Firstly, the G2-GARP protein sequences for 21 plant genomes (Data S1) 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[25] 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 over-expression construct generation

M. polymorpha 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[28] for spore transformation and[60] for thallus transformation.

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[79] using oligonucleotides in Data S5. gRNA sequences that were selected to generate Mp*glk*, Mp*gata2*, Mp*gata4*, Mp*scl,* Mp*hy5* and Mp*mir171* mutants are listed in Data S5. Single gRNA7 and gRNA3 to mutate Mp*GLK* was cloned using OpenPlant toolkit.[28] Multiple gRNAs to mutate Mp*GATA4*, Mp*SCL* and Mp*MIR171* and a single gRNA to mutate Mp*GATA2* were cloned as described in,[79]using oligonucleotides listed in Data S5 and the destination vector pMpGE013.[61] For the over-expression constructs, Mp*GLK,* Mp*GLKrw* CDS, Aa*GLK*, Pp*GLK* and Sm*GLK* were synthesised (Integrated DNA Technologies, sequence can be found in Data S6), Mp*GLK* 3′UTR was amplified from *M. polymorpha* genomic DNA and cloned into the pUAP4 vector.[28] Constructs were generated using the OpenPlant toolkit.[28] OpenPlant parts used: OP-023 CDS12-eGFP, 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, OP-062 L1_CsR-Ck1, OP073 L1_Cas9-Ck4, OP-076 L2_lacZgRNA-Cas9-CsA, OP-075 L1_lacZgRNA-Ck2, OP-074 L1_lacZgRNA-Ck3 and L1_35S_s:eGFP-Lti6b.

Chlorophyll determination, fluorescence measurements and imaging analysis

For chlorophyll measurements, ~30-50mg of 10–14 days old gemmalings or thallus tissue 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 65°C with 300 rpm shaking for 45 min. Samples were let to cool down to room temperature for approximately 1 h. Chlorophyll content was measured using a NanoDropOne/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.[80] *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 (F_m), variable

fluorescence F_v ($F_v = F_m - F_o$). 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 coverslip. 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.[80]

Extraction of terpenes and analysis on gas chromatograph - Mass spectrometer

Approximately 200 mg of frozen *M. polymorpha* plant material was extracted with 1 mL of cold methanol containing 5 mM of NaCl to quench enzymatic activities[81] and 900 ng of dodecane as internal standard (#297879, Sigma Aldrich). Mechanical disruption was achieved using a TissueLyser II (Qiagen), and the mixture was agitated on a Vibrax shaker at 2000 rpm for 2 h. The samples were centrifuged to isolate the methanolic extracts devoid of plant debris, then 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 × 0.25 mm x 0.25 μm) (ThermoFisher).

Sample injection (1 μL) was conducted in splitless mode, following GC-MS oven parameters modified from[82]: 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-min delay, with the transfer line and ion source temperatures set at 270°C and 230°C, respectively. Compounds were detected using the scan mode (scan time 0.17 s) within a mass detection range of 40–600 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.[82]

RNA extraction, cDNA preparation, qPCR and RNA sequencing

RNA was extracted from 3 to 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 min 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 Data 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 min followed by 40 cycles of 94°C for 15 s (denaturation) and 60°C for 1 min (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 analyzed 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 (10.5281/zenodo.5127899) to trim low-quality reads and remaining sequencing adapters. Reads were pseudo-aligned using Kallisto[69] to the *M. polymorpha* Genome version 5 (primary transcripts only, obtained from MarpolBase).[83] Kallisto estimates the abundance of each transcript in units of transcripts per million (TPM). Mapping statistics for each library are provided in Data S3. DGE analysis was performed with DESeq2,[62] Data S3. Plots were generated using R. All raw data has been deposited in NCBI under the accession number PRJNA1039314.

ChIP-SEQ, ATAC-SEQ and data analysis

ChIP-SEQ was performed using two Mp*GLK* over-expression 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.[41] Chromatin was then sonicated using a Bioruptor (4 × 5 min cycles of 30 s ON/OFF). Sonicated chromatin was incubated with 2μg Anti-HA Tag Antibody (#26183, Invitrogen) for 6 h at 4°C and an additional 2 h 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.[84] 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 was performed at 55°C for 1 h 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.[84] 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 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[71] was 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 fold-change >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. Orthofinder[73] (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 in Figure 7H. For example, the number of ortholog groups with GLK targets in 6 and 5 species were labeled as X6 and X5 respectively.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

Statistical analysis was performed using the R software (https://www.rstudio.com/). The statistical analysis details are described in the figure legends including exact value of n, and the statistical tests used.