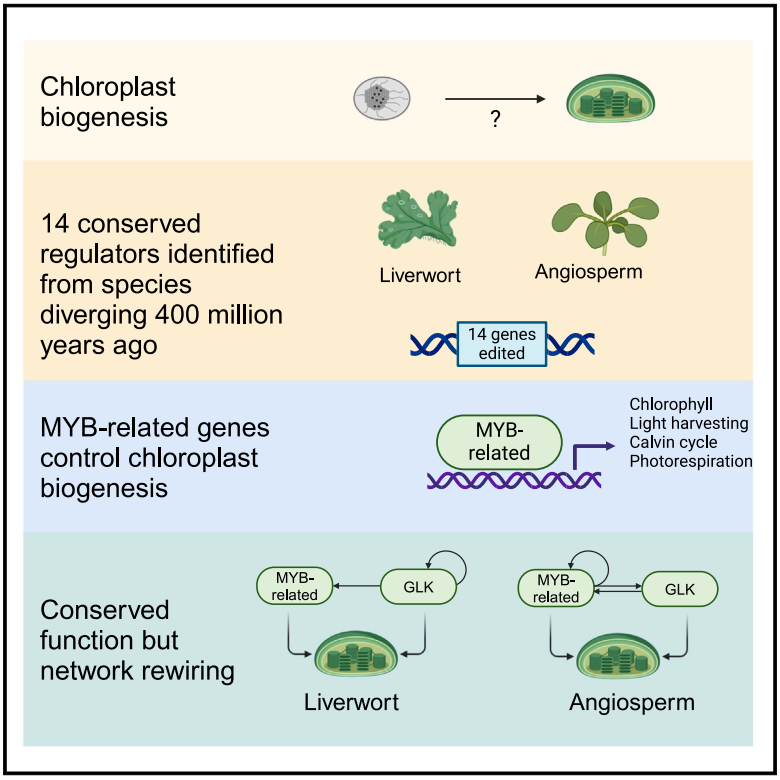
Cell

Article

MYB-related transcription factors control chloroplast biogenesis

Graphical abstract



Authors

Eftychios Frangedakis, Nataliya E. Yelina, Kumari Billakurthi, ..., Marta Tomaselli, Jim Haseloff, Julian M. Hibberd

Correspondence

ef391@cam.ac.uk, jmh65@cam.ac.uk (J.M.H.), efragedakis@gmail.com (E.F.)

In brief

Analysis of a liverwort with a streamlined genome identified MYB-related transcription factors as regulators of chloroplast development in land plants.

Highlights

- Analysis of a liverwort and an angiosperm identified regulators of chloroplast biogenesis
- Mutants in *MYB*-related genes are pale and contain small chloroplasts
- MYB-related transcription factors target genes for carbon fixation and light harvesting
- Rewiring of MYB-related and GLK targets between liverwort and angiosperm

Frangedakis et al., 2024, Cell *187*, 4859–4876
September 5, 2024
https://doi.org/10.1016/j.cell.2024.06.039

CellPress

Cell

CellPress
OPEN ACCESS

Article

MYB-related transcription factors control chloroplast biogenesis

Eftychios Frangedakis,[1,4,*] Nataliya E. Yelina,[1,2,4] Kumari Billakurthi,[1] Lei Hua,[1] Tina Schreier,[1,3] Patrick J. Dickinson,[1] Marta Tomaselli,[1] Jim Haseloff,[1] and Julian M. Hibberd[1,5,*]

[1]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK
[2]Present address: Crop Science Centre, University of Cambridge, 93 Lawrence Weaver Road, Cambridge CB3 0LE, UK
[3]Present address: Department of Biology, University of Oxford, South Parks Road, Oxford OX1 3RB, UK
[4]These authors contributed equally
[5]Lead contact
*Correspondence: ef391@cam.ac.uk or efragedakis@gmail.com (E.F.), jmh65@cam.ac.uk (J.M.H.)
https://doi.org/10.1016/j.cell.2024.06.039

SUMMARY

Chloroplast biogenesis is dependent on master regulators from the GOLDEN2-LIKE (GLK) family of transcription factors. However, *glk* mutants contain residual chlorophyll, indicating that other proteins must be involved. Here, we identify MYB-related transcription factors as regulators of chloroplast biogenesis in the liverwort *Marchantia polymorpha* and angiosperm *Arabidopsis thaliana*. In both species, double-mutant alleles in *MYB*-related genes show very limited chloroplast development, and photosynthesis gene expression is perturbed to a greater extent than in *GLK* mutants. Genes encoding enzymes of chlorophyll biosynthesis are controlled by MYB-related and GLK proteins, whereas those allowing CO_2 fixation, photorespiration, and photosystem assembly and repair require MYB-related proteins. Regulation between the MYB-related and GLK transcription factors appears more extensive in *A. thaliana* than in *M. polymorpha*. Thus, *MYB*-related and *GLK* genes have overlapping as well as distinct targets. We conclude that MYB-related and GLK transcription factors orchestrate chloroplast development in land plants.

INTRODUCTION

Photosynthesis is fundamental to life, and in eukaryotes, it takes place in organelles known as chloroplasts. It is widely accepted that chloroplasts originated from endosymbiosis between a photosynthetic prokaryote and heterotrophic eukaryote initiated more than 1 billion years ago.[1–3] Since then, significant elaborations to the control of photosynthesis gene expression have taken place. For example, in plants, the majority of genes allowing chloroplast biogenesis are encoded in the nucleus such that thousands are post-translationally imported into the chloroplast.[4,5] Improving photosynthetic efficiency is considered an important target for crop improvement and, thus, global food security,[6,7] but to predictively engineer this key process, an improved knowledge of photosynthesis gene regulatory networks is required.

Expression of photosynthesis-associated nuclear genes is responsive to light as well as processes intrinsic to the cell. For example, in angiosperms, light is required for chloroplast formation, but hormones amplify or repress this response.[8] These exogenous and endogenous inputs are integrated by key transcriptional regulators, including the bZIP transcription factor elongated hypocotyl 5 (HY5) that acts antagonistically with phytochrome interacting factors (PIFs) to activate chloroplast development in the presence of light. HY5 activates expression of master regulators of chloroplast biogenesis belonging to the GOLDEN2-LIKE (GLK) and GATA families of transcription factors (GATA nitrate-inducible carbon metabolism-involved [GNC] and cytokinin-responsive GATA factor 1 [CGA1]).[8–13] Yet, *glk* mutants in *Arabidopsis thaliana* (*A. thaliana*),[10] rice,[14] and also non-seed plants such as *Physcomitrium patens*[15] and *Marchantia polymorpha*[16] contain chlorophyll. Moreover, *A. thaliana* mutants lacking functional *GLK* and *GATA* genes are not albino.[9,17] In summary, other actors must allow assembly of the photosynthetic apparatus in the absence of these known regulators.

We sought to identify transcription factors operating alongside the master regulator GLK. As forward genetics has failed to identify such proteins, we rationalized that genetic redundancy had hindered their identification, and analysis of a species with a more compact genome would circumvent this issue. *Marchantia polymorpha* has a streamlined genome with many transcription factors represented by one or two copies, and the dominant form of lifecycle is haploid.[18] Although in *M. polymorpha GATA* orthologs are not required for greening, consistent with analysis of *A. thaliana*, Mp*glk* knockouts contain low but detectable levels of chlorophyll.[16] We therefore hypothesized that an unknown transcription factor conserved between *M. polymorpha* and

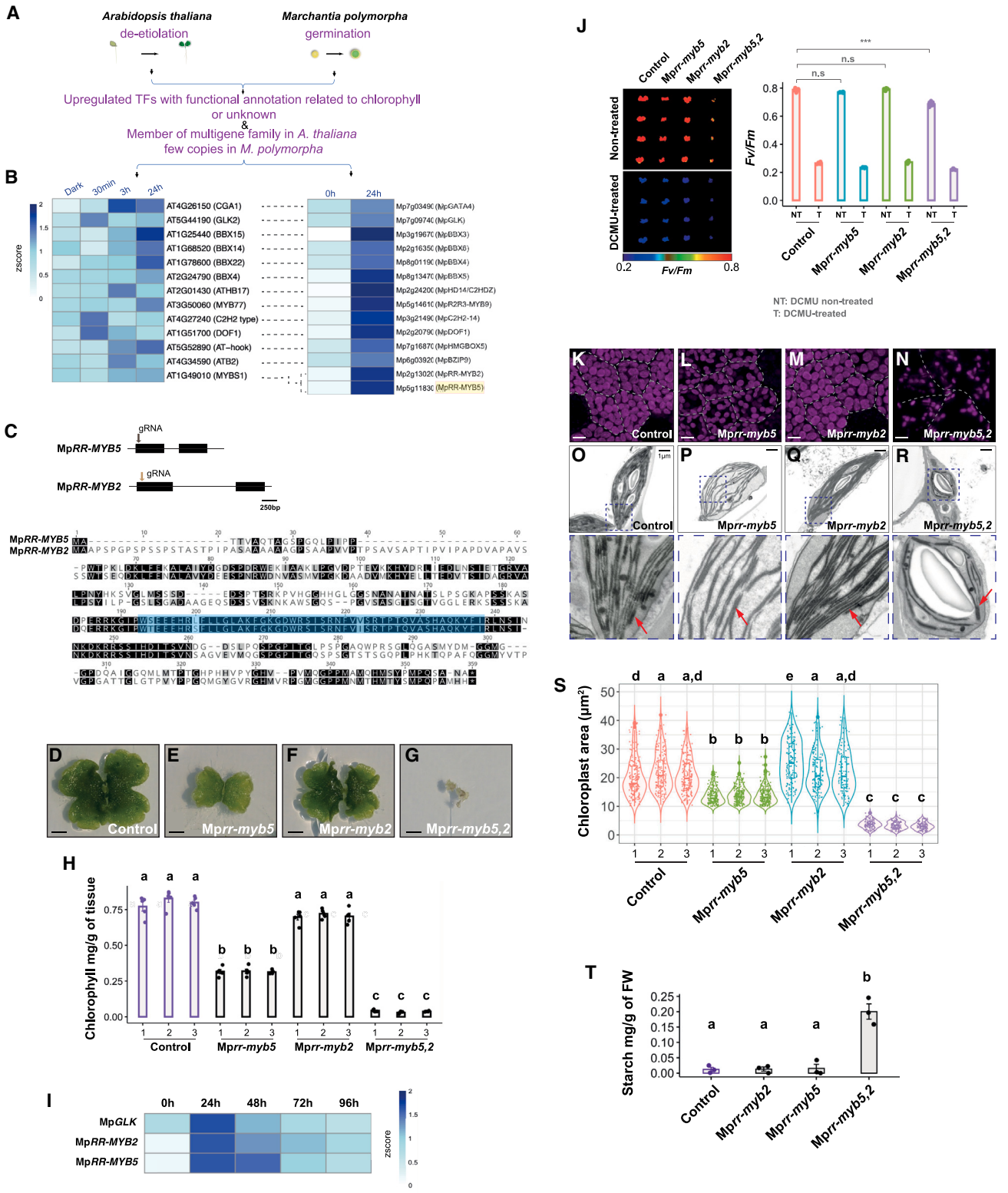


Figure 1. Mp*RR-MYB5* and Mp*RR-MYB2* act redundantly to regulate chloroplast biogenesis in *M. polymorpha*

(A) Schematic of pipeline used to identify candidate transcription factors regulating photosynthesis gene expression. RNA-seq data from *Arabidopsis thaliana* during de-etiolation[20] and *Marchantia polymorpha* during spore germination[19] were examined. Transcription factors (TFs) upregulated in response to light in both

(legend continued on next page)

A. thaliana acts alongside *GLK* to control chloroplast biogenesis during photomorphogenesis.

After re-examination of publicly available RNA sequencing (RNA-seq) data, gene editing of transcription factors, and phenotypic analysis, we identify two RR-type myoblastoma related (RR-MYB; ''R'' refers to domain repeat) transcription factors as regulators of chloroplast biogenesis and photosynthesis gene expression in *M. polymorpha* and *A. thaliana*. In contrast to the GLK proteins that regulate expression of genes allowing chlorophyll biosynthesis and function of photosystems I and II, the RR-MYBs have a broader set of targets that extends to genes allowing CO_2 fixation, photorespiration, photosystem assembly, and repair. We conclude that these proteins function as master regulators of chloroplast biogenesis and photosynthesis gene expression. The data have implications for understanding chloroplast biogenesis and photosynthesis as well as other processes taking place in plastids such as nitrogen and sulfur assimilation, the biosynthesis of amino acids, fatty acids, and carotenoids.

RESULTS

Mp*RR-MYB5* regulates chloroplast development synergistically with its paralog Mp*RR-MYB2*

We interrogated publicly available gene expression data sampled during the transition from non-photosynthetic to photosynthetic growth in *M. polymorpha*[19] as well as *A. thaliana* (Figures 1A and 1B).[20] This identified 108 and 144 transcription factors upregulated after exposure to light in *M. polymorpha* and *A. thaliana*, respectively (Tables S1 and S2). Orthologs upregulated in both datasets with an unknown or chlorophyll-related annotation and belonging to a multigene family in *A. thaliana* were selected (Figures 1A and 1B). Fourteen candidates from *M. polymorpha* were identified (Figure 1B; Table S2). Two of these (Mp*GLK* and Mp*GATA4*) are homologs of known photosynthesis regulators in *A. thaliana*,[10,21] and Mp*GLK* has a confirmed role in *M. polymorpha*.[16] The remainder included a number of B-box (BBX) domain proteins known to interact with the master regulator of photomorphogenesis HY5,[22] a homeobox-leucine zipper (HD-ZIP) protein whose *Arabidopsis thaliana* homeobox 17 (*ATHB17*) ortholog regulates photosynthesis-associated nuclear genes in response to abiotic stress,[23] a *C2H2* type zinc finger transcription factor with unknown function in *A. thaliana*, and a *MYB*-related gene predicted to regulate photosynthesis gene expression.[24,25]

We used *M. polymorpha* as a testbed by subjecting each of these candidates to CRISPR-Cas9-mediated editing. With the exception of Mp*GLK*, which has previously been reported to lead to a pale phenotype when mutated[16] only one other candidate had low chlorophyll. This was Mp5g11830, annotated as Mp*RR-MYB5* in the *M. polymorpha* genome database, but previously also referred to as a circadian clock associated-like (CCA1-like) RR-MYB-related transcription factor.[26,27] Mp*RR-MYB5* has a single paralog (Mp*RR-MYB2*—Figures 1B and 1C) showing high similarity to Mp*RR-MYB5* at the amino acid level, with, for example, the CCA1-like/RR-Myb domain being 93% identical (Figure 1C). Mutant alleles of Mp*RR-MYB5* but not Mp*RR-MYB2* appeared pale (Figures 1D–1F, S1A, and S1B), and analysis of chlorophyll content confirmed this (Figure 1H). All lines in which insertions or deletions introduced premature stop codons in Mp*RR-MYB5* (Figure S1A) had 40%–50% less chlorophyll than controls (Figures 1E and 1H). The Mp*rr-myb5* mutant was complemented when Mp*RR-MYB5* was expressed from its own promoter (Figure S1C), confirming the pale phenotype was unlikely associated with off-target Cas9 editing. Mutating Mp*RR-MYB5* and Mp*RR-MYB2* simultaneously (Figure S1D) led to extremely pale plants with chlorophyll content reduced by ~95% compared with controls (Figures 1G and 1H). Analysis of a time course associated with the first hours after exposure to light[19] indicated that transcripts derived from Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*GLK* were induced by 24 h after light was perceived but that Mp*RR-MYB5* and Mp*RR-MYB2* expression was maintained for longer (Figure 1I). To test whether the photosynthetic apparatus was functional in the single Mp*rr-myb5* and

species with either unknown or related to chlorophyll function were retained. As we hypothesized that functional redundancy had hindered identification of such regulators via forward genetic screens an additional criterion was that each family should be represented by multiple copies in *A. thaliana*.

(B) Heatmaps showing transcript abundance (*Z* score) of candidates that were selected to generate knockout mutants in *M.* polymorpha. Protein name shown in brackets after each gene identifier.

(C) Top: schematic of Mp*RR-MYB5* and Mp*RR-MYB2* gene structure (exons represented as black boxes) with guide (g) RNAs for CRISPR represented by arrows. Bottom: amino acid sequence alignments of MpRR-MYB5 and MpRR-MYB2 proteins with the characteristic RR-MYB/CCA1-like domain highlighted in blue.

(D–G) Representative images of control and Mp*rr-myb5*, Mp*rr-myb2*, and Mp*rr-myb5,2* mutants. Scale bars represent 2 mm.

(H) Chlorophyll content of Mp*rr-myb5*, Mp*rr-myb2*, and Mp*rr-myb5,2* mutants. Letters show ranking using a *post hoc* Tukey test with different letters indicating statistically significant differences at $p < 0.01$. Data presented as means and standard error of the mean, $n = 5$.

(I) Heatmaps showing transcript abundance (*Z* score) of Mp*GLK*, Mp*RR-MYB2*, and Mp*RR-MYB5* in germinating spores during the first 96 h after exposure to light.[19]

(J) Representative images and quantification after imaging of chlorophyll fluorescence parameter F_v/F_m after treatment with the inhibitor di-chlorophenyl dimethyl urea (DCMU). Asterisks indicate statistically significant difference using a two-tailed t test, $^{***}p \leq 0.0001$, n.s., non-significant, $n = 8$.

(K–N) Representative images after confocal laser scanning microscopy of Mp*rr-myb5*, Mp*rr-myb2*, and Mp*rr-myb5,2* mutants with chlorophyll autofluorescence shown in magenta. Cell borders are marked with dashed white lines. Scale bars represent 10 μm.

(O–R) Representative images after transmission electron microscopy of control, Mp*rr-myb5*, Mp*rr-myb2*, and Mp*rr-myb5,2* mutants. Scale bars represent 1 μm. Dashed area depicted in each chloroplast is enlarged, and granal stacks indicated with red arrows.

(S) Violin plots of chloroplast area for control, Mp*rr-myb5*, Mp*rr-myb2*, and Mp*rr-myb5,2*. Box and whiskers represent the 25 to 75 percentile and minimum-maximum distributions of the data. Letters show ranking using a *post hoc* Tukey test with different letters indicating statistically significant differences at $p < 0.01$. $n = 150$.

(T) Starch levels in control, Mp*rr-myb2*, Mp*rr-myb5*, and Mp*rr-myb5,2* mutants. FW: fresh weight. Data presented as means and standard error of the mean, $n = 3$. See also Figures S1 and S2.

double Mp*rr-myb5,2* mutants, we applied the inhibitor di-chlorophenyl di-methyl urea (DCMU) that blocks photosynthetic electron transport[28] and measured activity of photosystem II via chlorophyll fluorescence imaging (Figure 1J). Although the mutants had low levels of chlorophyll, the photosynthetic apparatus was operational, suggesting that any remaining chlorophyll was functionally associated with photosystem II reaction centers. There was a small but significant reduction in F_v/F_m in the double Mp*rr-myb5,2* mutant (Figure 1J) indicating compromised function in this genotype. Consistent with low chlorophyll levels in Mp*rr-myb5* and Mp*rr-myb5,2* mutants, chloroplasts were smaller and thylakoids underdeveloped compared with controls and Mp*rr-myb2* (Figures 1K–1S and S1E). Poorly developed chloroplasts from Mp*rr-myb5,2* mutants contained more starch (Figure 1T).

To determine whether Mp*RR-MYB5* and Mp*RR-MYB2* limit greening, we generated over-expression lines driven by the strong Mp*UBE2* constitutive promoter[29] and used eGFP to mark the plasma membrane (Figures S2A–S2E). Although quantitative polymerase chain reactions confirmed each transgene was over-expressed (Figures S2F–S2N), plants appeared similar to controls, and there were no evident perturbations to chlorophyll content, chloroplast size, or morphology (Figures S2O–S2P). We conclude that Mp*RR-MYB5* and Mp*RR-MYB2* act redundantly and are necessary for chloroplast biogenesis, but in contrast with Mp*GLK*,[16] they are not sufficient to activate this process. Moreover, in the absence of both Mp*RR-MYB5* and Mp*RR-MYB2* assembly of the photosynthetic apparatus was very limited.

Mp*RR-MYB5* and Mp*RR-MYB2* act with Mp*GLK* to control chloroplast biogenesis

As double Mp*rr-myb5,2* mutants showed residual chloroplast development, we hypothesized that their limited photoautotrophic growth was associated with activity of the previously characterized master regulator *GLK*. To test this, we attempted to generate higher-order mutants that combined mutant alleles of Mp*glk*, Mp*rr-myb5*, and Mp*rr-myb2*. We were able to knock out Mp*RR-MYB2* or Mp*RR-MYB5* in the presence of Mp*glk* mutant alleles (Figures S3A–S3C), and these double mutants were similar or paler than the Mp*glk* mutant (Figures 2A–2E and S3D). Although previous analysis indicated that GATAs are not required for chloroplast biogenesis in *M. polymorpha*,[16] we produced double Mp*gata4,rr-myb5* mutants. Chlorophyll content in these lines was similar to that of Mp*rr-myb5*, and they showed similar perturbations to thallus morphology evident in Mp*gata4* mutants (Figures S3E and S3F). This implies that Mp*GATA4* is unlikely to control chloroplast development with Mp*rr-MYB5*. Application of DCMU confirmed that photosystem II was functional in Mp*glk,rr-myb5* double mutants (Figure 2F). Chloroplasts were smaller and contained fewer thylakoid membranes with reduced granal stacking in Mp*glk,rr-myb5* compared with each single mutant (Figures 2G–2O). Thus, in the absence of both Mp*RR-MYB5* and Mp*GLK*, very limited assembly of the photosynthetic apparatus took place.

We were unable to generate triple Mp*glk,rr-myb5,2* mutants, implying that this allelic combination could be lethal. For example, after super-transforming Mp*glk,rr-myb5* mutants with a vector allowing expression of the same guide RNA used to generate the Mp*rr-myb2* mutants reported above, 91 lines were obtained. However, none were paler than Mp*glk,rr-myb5* mutants, and when genotyped 86 lines had no edits in Mp*RR-MYB2*. Of the five lines that were edited in Mp*RR-MYB2* (as well as Mp*GLK* and Mp*RR-MYB5*), mutations had limited impact on the MpRR-MYB2 protein, with reading frame maintained and amino acids being modified in a poorly conserved region of the protein (Figures S3G and S3H). By contrast, when the original Mp*rr-myb2* mutants were identified (Figure 1F), more than 60% of plants contained mutations that introduced early stop codons or disturbed reading frame (Data S1). We propose that absence of all three proteins (MpGLK, MpRR-MYB5, and MpRR-MYB2) is likely lethal, possibly due to the lack of chloroplast biogenesis.

MpRR-MYB transcription factors regulate genes allowing carbon fixation, photorespiration, and photosystem function

To provide insight into the types of genes regulated by Mp*RR-MYB5* and Mp*RR-MYB2*, we performed RNA-seq of overexpressing lines, single and double mutants. Over-expression of Mp*RR-MYB2* and Mp*RR-MYB5* led to upregulation of 71 and 11 genes, respectively (*padj* value $\leq$ 0.01, $\log_2$FC $\geq$ 1-fold) (Figures S4A and S4B; Data S1), and there was limited overlap between these two datasets (Figures 3A, S4C, and S4D). Over-expression of Mp*GLK* that led to the upregulation of 493 genes (Figure 3A and Yelina et al.[16]).

In loss-of-function mutants for Mp*RR-MYB2* or Mp*RR-MYB5*, 65 and 823 genes showed reductions in transcript abundance, respectively, compared with controls (*padj* value $\leq$ 0.01, $\log_2$FC $\geq$ 1-fold) (Figures 3B, S4E, and S4F; Data S1). Knocking out Mp*GLK* had greater impact with 1,065 genes being downregulated (Figures 3B and S4G). In double Mp*glk,rr-myb5* mutants (Figure S4H), 1,161 genes had lower transcript abundance than controls, and in the double Mp*rr-myb5,2* mutants, this was increased to 1,744 (Figures 3B and S4I). The largest overlap between genotypes (524 genes) was detected for Mp*glk,rr-myb5* and Mp*rr-myb5,2* mutants (Figures 3B and S4J), further supporting synergistic action of Mp*RR-MYBs* and Mp*GLK*. Gene Ontology (GO) terms were used to provide insight into classes of genes impacted by over-expression or loss of function of Mp*RR-MYBs* and Mp*GLK*. Consistent with the lack of detectable phenotype after over-expression of Mp*RR-MYB5* or loss of Mp*RR-MYB2* function, no distinct GO terms were impacted in these lines. When genes downregulated in Mp*rr-myb5* and Mp*glk* mutants were assessed, the response to oxidative stress and hydrogen peroxidase catabolism terms were over-represented (Figure 3C; Data S2). Mp*rr-myb5* mutants also showed changes to protein phosphorylation (Figure 3C; Data S2), while in Mp*glk* photosynthesis, light harvesting and chlorophyll biosynthesis terms were affected (Figure 3C). It was notable that for genes downregulated in the Mp*glk*, Mp*glk,rr-myb5*, and Mp*rr-myb5,2*, mutants similar terms were over-represented (Figure 3C; Data S2). For genes upregulated in Mp*glk* mutants DNA replication-related GO terms were over-represented, whereas response to auxin and transmembrane transport were affected in Mp*rr-myb5,2* and Mp*glk,rr-myb5* mutants, respectively (Figure S4K; Data S2). Transcripts associated with proteins operating in similar cellular locations were impacted in

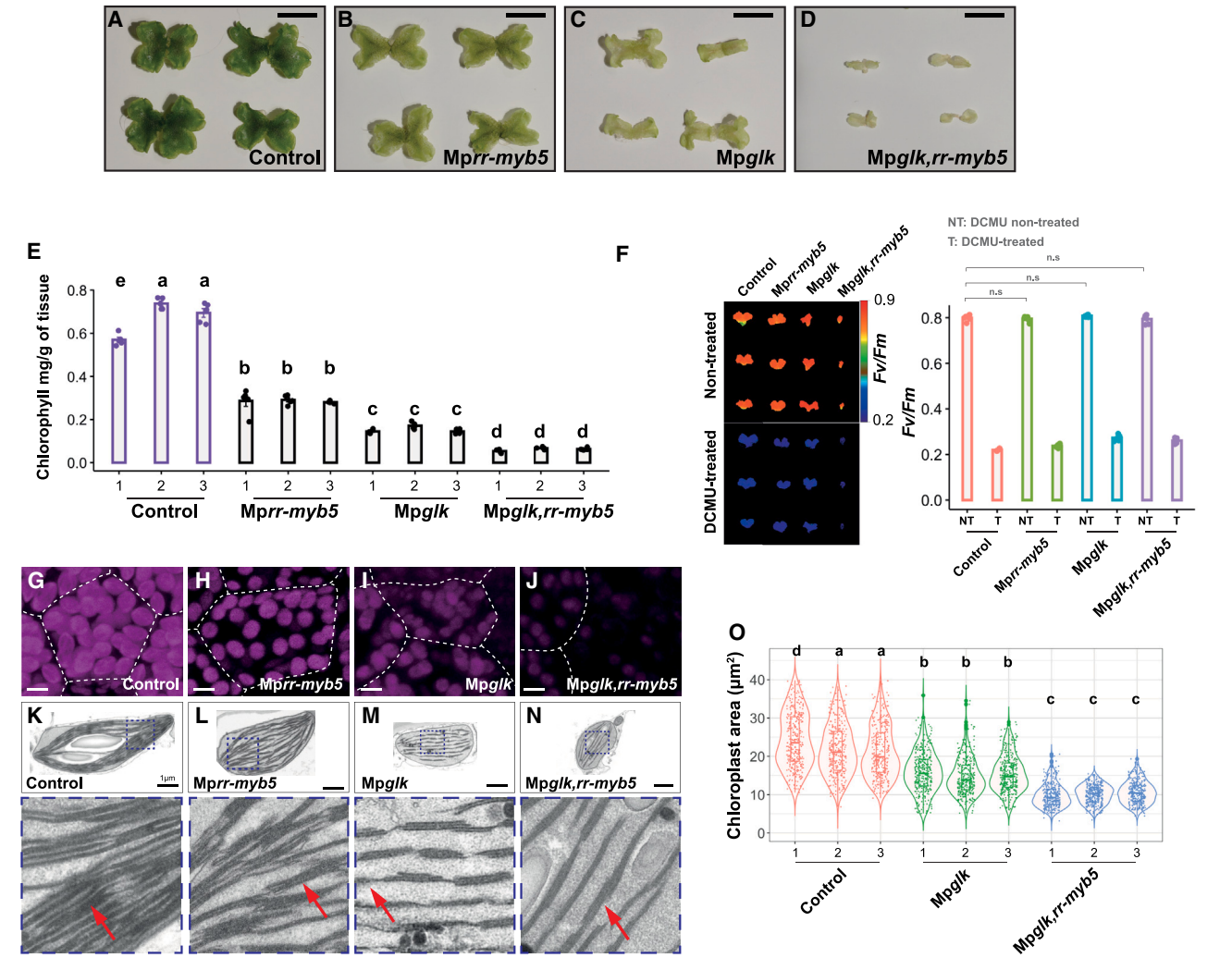


Figure 2. MpRR-MYB5 acts synergistically with MpGLK to control chloroplast biogenesis

(A–D) Representative images of control (same as Figure 1D), Mp*rr-myb5*, Mp*glk*, and Mp*glk,rr-myb5* mutants. Scale bars represent 7 mm.

(E) Chlorophyll content in Mp*rr-myb5*, Mp*glk*, and the double Mp*glk,rr-myb5* mutants compared with controls. Letters show ranking using a *post hoc* Tukey test with different letters indicating statistically significant differences at $p < 0.01$. Data presented as means and standard error of the mean, $n = 5$.

(F) Representative images and quantification after imaging of the chlorophyll fluorescence parameter F_v/F_m after treatment with the inhibitor di-chlorophenyl di-methyl urea (DCMU). n.s., non-significant statistical difference using a two-tailed t test, $n = 6$.

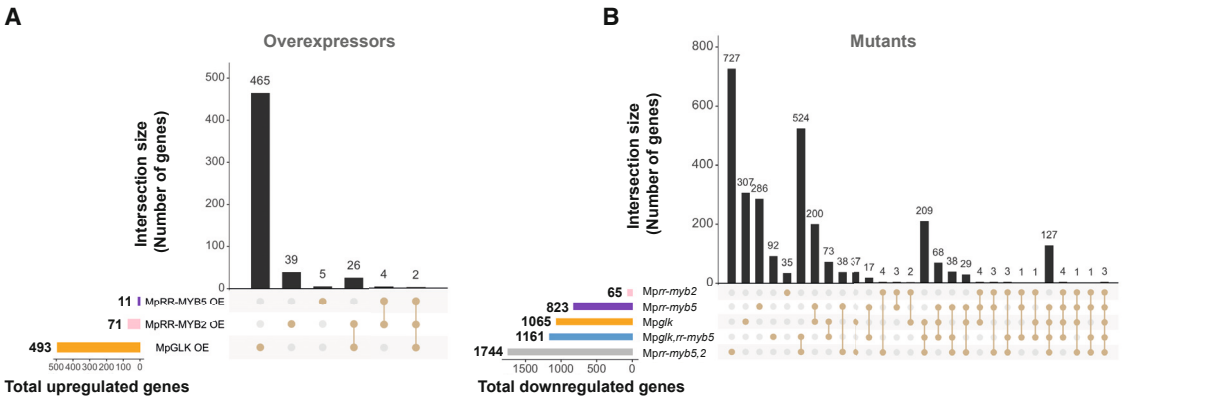
(G–J) Representative images after confocal laser scanning microscopy of control, Mp*rr-myb5*, Mp*glk*, and Mp*glk,rr-myb5* mutants. Chlorophyll autofluorescence shown in magenta, cell borders marked with dashed white lines. Scale bars represent 10 μm.

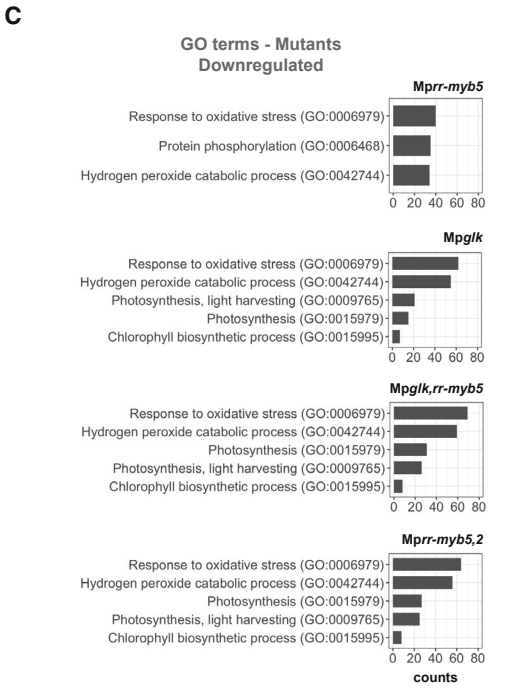
(K–N) Representative images from transmission electron microscopy images of control, Mp*rr-myb5*, Mp*glk*, and Mp*glk,rr-myb5* mutants. Scale bars represent 1 μm. Dashed area depicted in each chloroplast is enlarged, and granal stacks indicated with red arrows.

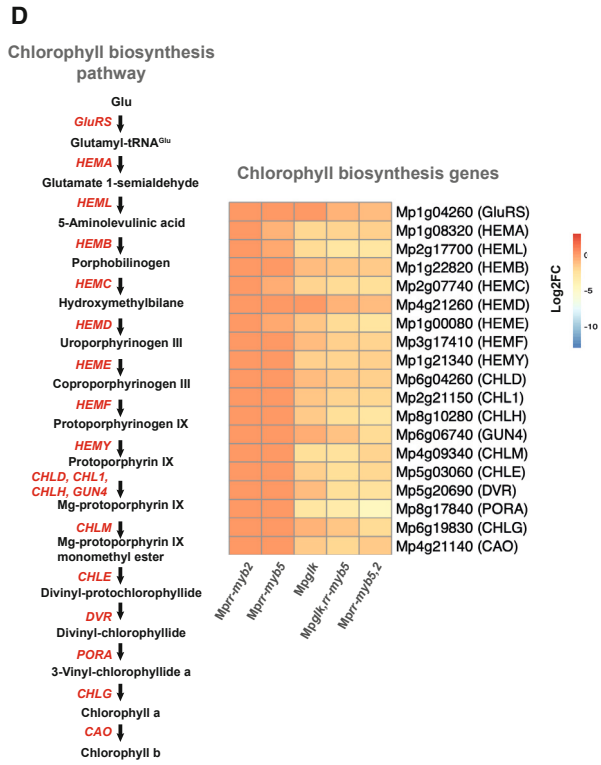
(O) Chloroplast area of Mp*glk* and Mp*glk,rr-myb5* mutants. Box and whiskers represent the 25 to 75 percentile and minimum-maximum distributions of the data. Letters show ranking using a *post hoc* Tukey test with different letters indicating statistically significant differences at $p < 0.01$. $n = 330$.

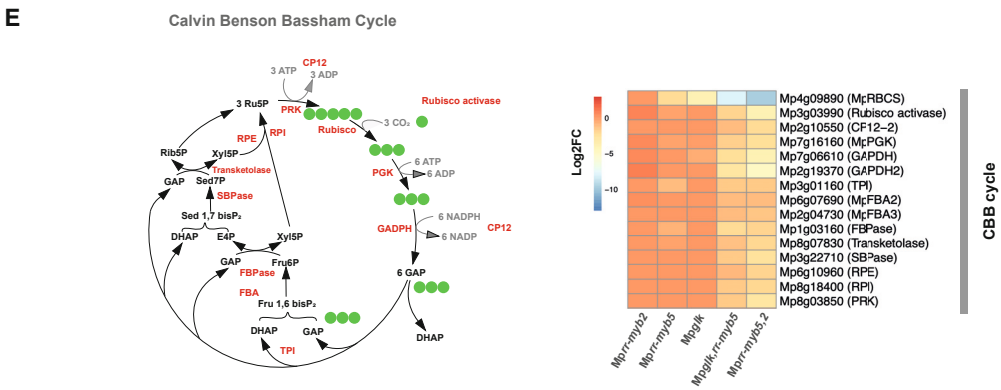
See also Figure S3.

Mp*glk*, Mp*glk,rr-myb5*, and Mp*rr-myb5,2* mutants, including photosystem I, thylakoid membranes, and also the ribosome (Data S2). In the case of Mp*rr-myb5,2*, however, a more extensive effect on chloroplast and photosystem-related GO terms was evident. For upregulated genes, no GO terms were impacted in the Mp*rr-myb5,2* and Mp*glk,rr-myb5* mutants, while for Mp*glk*, the nucleosome GO term was most impacted (Data S2). Loss of function of either the MpRR-MYBs or MpGLK primarily therefore caused changes in overlapping GO terms, exemplified by those associated with photosynthesis, responses to oxidative stress, and hydrogen peroxide catabolic processes (Data S2). However, some terms were specific to each mutant, with, for example, iron ion transmembrane transport, translation, and translational elongation being enriched in Mp*rr-myb5,2*, while the L-phenylalanine catabolic process term was specific to Mp*glk* (Data S2). Overall, loss-of-function Mp*glk*, Mp*glk,rr-myb5*,









(legend on next page)

and Mp*rr-myb5,2* mutants caused changes in GO terms primarily associated with photosynthesis (Data S2). For the common impacted GO categories, ~40%–60% of genes were co-regulated by both Mp*RR-MYBs* and Mp*GLK* (Data S2).

Since chlorophyll content was reduced in Mp*rr-myb5* and Mp*rr-myb5,2* mutants, we examined transcript abundance of the nineteen annotated chlorophyll biosynthesis genes (Figure 3D). With the exception of *HEMA* (encoding the glutamyl-tRNA reductase) in Mp*rr-myb5* mutants, knocking out either Mp*RR-MYB5* or Mp*RR-MYB2* did not significantly affect transcript abundance. By contrast, in Mp*glk* mutants, transcript abundance from seventeen of the genes was reduced, and in Mp*rr-myb5,2* all nineteen genes were significantly downregulated (Figure 3D). The effect on transcript abundance was more pronounced in Mp*rr-my5,2* than in Mp*glk,rr-myb2* (Figure 3D). We next examined the impact of loss of the Mp*RR-MYBs* on approximately 200 other genes annotated as photosynthesis-related (Data S1). This included genes associated with CO_2 fixation, the light-harvesting apparatus, and their assembly and repair. In the Mp*rr-myb5* and Mp*rr-myb2* mutant alleles, there was a limited effect on photosynthesis-associated genes (Figures 3E, 4A, and 4B). For example, in Mp*rr-myb2*, expression of only *petE* (Mp4g02720) was significantly perturbed ($\log_2$FC $\geq$ 1-fold) (Data S1). In Mp*rr-myb5* mutants, a small number of genes were impacted, including those encoding a small subunit of RuBisCO (RbcS) (Mp4g09890) and a chlorophyll a/b binding protein (Mp7g05530) (Figure 3E; Data S1). As expected, changes to photosynthesis transcripts were more evident in Mp*glk* (Figures 3E, 4A, and 4B) and even more severe when both Mp*RR-MYB5* and Mp*GLK* were absent (Figures 3E, 4A, and 4B). Strikingly, when Mp*RR-MYB2* and Mp*RR-MYB5* were simultaneously knocked out, the effect on photosynthesis-associated genes was extensive and more widespread than in *Mpglk,rr-myb5*. For example, in Mp*rr-myb5,2* double mutants, the majority of genes encoding enzymes involved in the Calvin-Benson-Bassham cycle and photorespiration were downregulated (Figures 3E and S4L). Moreover, genes encoding components of both photosystems and their respective light-harvesting complexes as well as the cytochrome b_6f complex were downregulated (Figures 4A and 4B). We also found that genes associated with assembly of RuBisCO, non-photochemical quenching, as well as granal stacking and repair of photosystem II, were impacted in Mp*rr-myb5,2* (Figure S4M). This contrasts with the Mp*glk* single mutant, where the greatest changes were associated with genes encoding components of the photosystems and their light-harvesting complexes (Figures 4A and 4B). For genes, e.g., those associated with RuBisCO assembly, granal stacking, or repair of PSII, that were less expressed in Mp*glk*, the level of perturbation was much lower than in Mp*rr-myb5,2* (Figure S4M; Data S1). Immunoblotting demonstrated that LHCA, PsbA, PsaC, and PsbS proteins were less abundant in Mp*glk* and Mp*rr-myb5* mutants, and in the double Mp*rr-myb5,2*, this effect was enhanced (Figure 4C).

MpRR-MYBs can activate photosynthesis genes

The analysis above indicated that loss of Mp*RR-MYBs* had widespread impact on photosynthesis gene expression that overlapped with that seen in Mp*glk*. However, it was not clear if these effects were associated with direct interaction between the MpRR-MYB transcription factors and these structural genes. We therefore used DNA affinity purification and sequencing (DAP-seq)[32] to identify genome-wide binding sites of MpRR-MYB2 and MpRR-MYB5 and DNA sequence motifs that they bind. This identified a total of 6,804 and 2,839 binding sites for MpRR-MYB2 and MpRR-MYB5, respectively (Data S3). 43% of MpRR-MYB2 and 35% of MpRR-MYB5 binding sites were located 3 kb upstream of predicted translational start sites (Figure 4D). For MpRR-MYB2 and MpRR-MYB5, 890 and 301 of peaks mapping to promoter regions were misregulated in the Mp*rr-myb5,2* mutant (Figure 4D). The Fisher exact and permutation testing indicated this was a statistically higher overlap than would be expected by chance (Figure S5A). Motif enrichment analysis for MpRR-MYB2 and MpRR-MYB5 binding sites showed that they shared a TTATC consensus (Figure 4E).

We next asked whether the MpRR-MYB consensus binding site was more enriched in photosynthesis genes than would be expected by chance. The Fisher exact test and permutation analysis indicated that while the GLK binding site was not over-represented, there was a small but statistically significant over-representation of the RR-MYB consensus site in promoters of photosynthesis genes (Figure 4F). MpRR-MYB5 and MpRR-MYB2 were able to bind promoters of photosynthesis genes (22 genes and 10, respectively, Data S3), including Mp*PsbQ*, Mp*PsaD*, and Mp*LHCB1* (Figure 4G). Moreover, when Mp*RR-MYB5* was used in trans-activation assays as an effector, it was sufficient to activate expression from all three genes (Figure 4G) as well as Mp*PsaN* and Mp*PetA* (Figure S5B). Mp*RR-MYB2* activated Mp*LHCB1* (Figure 4G). Mp*RR-MYB5* also activated expression from Mp*DVR* (encoding divinyl chlorophyllide a 8-vinyl-reductase) and Mp*PORA* (encoding protochlorophyllide oxidoreductase A) that encode enzymes allowing chlorophyll biosynthesis (Figure S5B). Trans-activation assays showed that MpGLK was sufficient to activate both Mp*RR-MYB5* and Mp*RR-MYB2* (Figure 4H). However, there

Figure 3. Mp*RR-MYB5* and Mp*RR-MYB2* regulate expression of genes encoding chlorophyll biosynthesis and also the Calvin-Benson-Bassham cycle

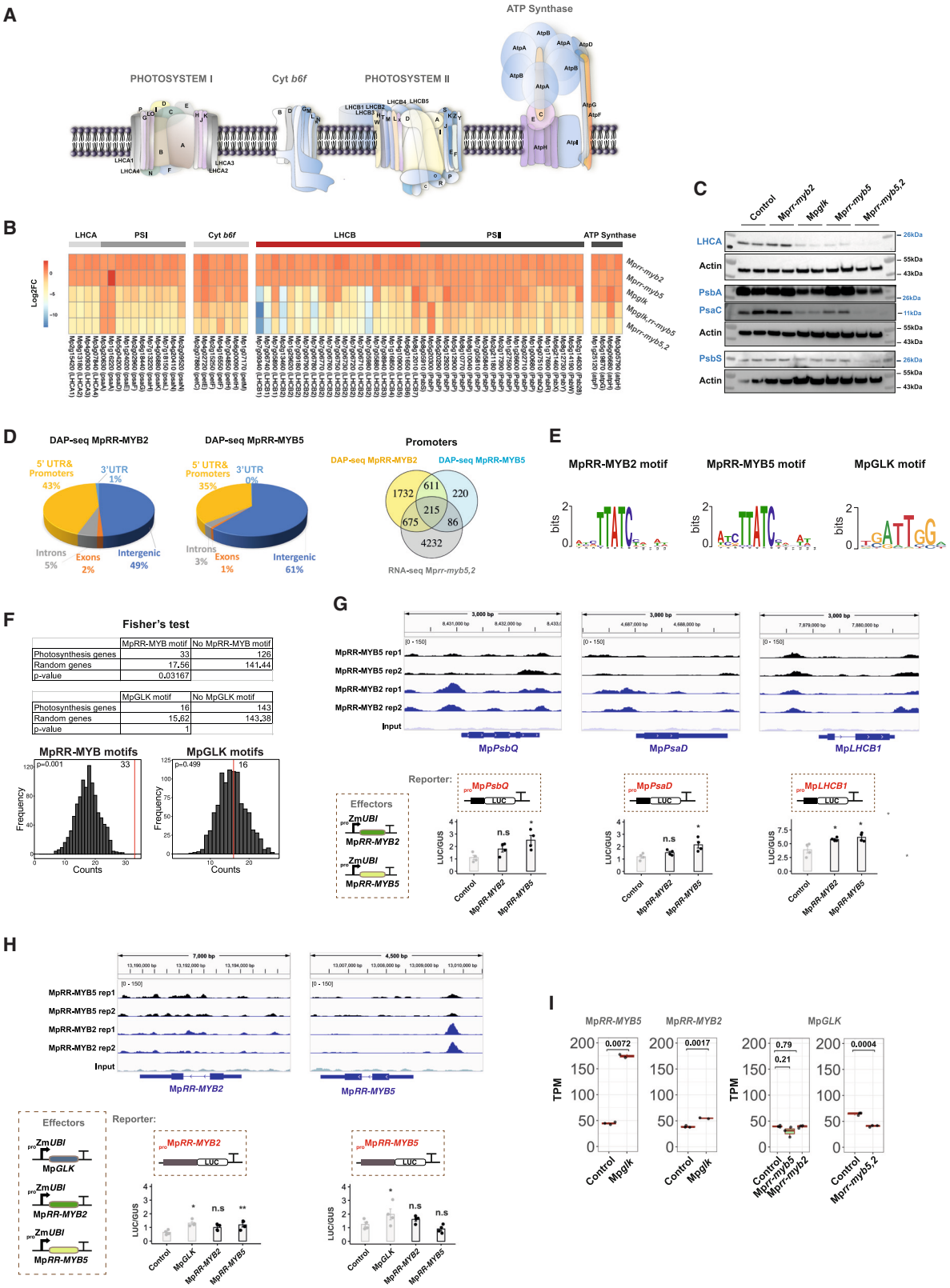
(A and B) Upset diagrams showing sets of upregulated genes in Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*GLK* over-expression lines (A) and downregulated genes in Mp*glk*, Mp*rr-myb2*, Mp*rr-myb5*, and the double Mp*glk,rr-myb5* or Mp*rr-myb5,2* mutants (B).

(C) Enriched GO terms for Mp*rr-myb5*, Mp*glk*, Mp*glk,rr-myb5*, and Mp*rr-myb5,2* mutants.

(D) Heatmap illustrating the extent of downregulation of transcripts encoding enzymes of chlorophyll biosynthesis in Mp*rr-myb2*, Mp*rr-myb5, Mpglk*, and the *Mpglk,rr-myb5*, or Mp*rr-myb5,2* mutant alleles.

(E) Heatmap indicating lower transcript abundance of genes encoding components of Calvin-Benson-Bassham cycle in Mp*rr-myb2*, Mp*rr-myb5*, Mp*glk*, as well as Mp*glk,rr-myb5*,*myb2*, and Mp*rr-myb5*,*myb2* double mutants. Schematic of Calvin-Benson-Bassham modified from Lea and Leegood.[30]

See also Figure S4.



(legend on next page)

was no strong evidence that Mp*RR-MYB2* bound its own promoter or activated it (Figure 4H), and while both MpRR-MYB2 and MpRR-MYB5 appeared able to bind the promoter of MpRR-MYB5 neither activated its expression (Figure 4H). It is possible that *in vivo* other factors interact with MpRR-MYB5 and MpRR-MYB5 to control their expression to allow co-ordination with Mp*GLK*. In support of this, analysis of the RNA-seq data indicated that Mp*RR-MYBs* and Mp*GLK* may be linked in a gene regulatory network. For example, in Mp*glk*, transcripts derived from Mp*RR-MYB5* and Mp*RR-MYB2* were upregulated (Figure 4I), and Mp*GLK* transcripts were downregulated in Mp*rr-myb5,2* (Figure 4I). We note a MpRR-MYB binding site in the sixth intron of Mp*GLK* (Figure S5C). Overall, DAP-seq and effector assays provide evidence that MpRR-MYB can bind promoters of photosynthesis genes and that they are sufficient to enhance expression from these targets.

We next tested whether MpRR-MYB2, MpRR-MYB5, or MpGLK transcription factors could rescue the pale phenotype of Mp*rr-myb5*, Mp*glk*, Mp*glk,rr-myb5*, or Mp*rr-myb5,2* mutants (Figures 5A–5G). Quantitative polymerase chain reactions confirmed that each transgene was over-expressed (Figure S5D). Mp*RR-MYB5* and Mp*RR-MYB2* complemented Mp*rr-myb5* mutants (Figures 5C and 5D), further arguing for functional redundancy. However, neither Mp*RR-MYB* rescued the single Mp*glk* (Figures 5D and 5E) or double Mp*glk,rr-myb5* mutants (Figures 5D and 5F). Mp*GLK* over-expression in Mp*rr-myb5* increased chlorophyll content but not to the same extent as seen in Mp*glk* (Figures 5D and 5E). Finally, when Mp*GLK* was over-expressed in Mp*rr-myb5,2*, chlorophyll levels were increased by ~10%, but the absolute levels were still 90% lower than wild type (Figures 5G and 5H). In summary, these data indicate that Mp*RR-MYB5* and Mp*RR-MYB2* cannot complement Mp*glk*, and Mp*GLK* cannot complement the double Mp*rr-myb5,2* mutant.

RR-MYBs control chloroplast biogenesis in *A. thaliana*

The RR-MYB/CCA1-like subfamily of MYB-related transcription factors[27,33,34] containing Mp*RR-MYB5* and Mp*RR-MYB2* is characterized by a conserved SHAQK(Y/F)F motif (Figure S6A). We identified eleven members of this group in *A. thaliana* (Figures 6A and S6B–S6E) of which At*MYBS1*, At*MYBS2*, and AT5G23650 were the closest homologs of Mp*RR-MYB5* and Mp*RR-MYB2* (Figures S6B–S6E). Re-analysis of publicly available data indicated that AT5G23650 is not expressed in photosynthetic tissues, so we focused our analysis on At*MYBS1* and At*MYBS2*. Due to the functional redundancy evident for Mp*RR-MYB5* and Mp*RR-MYB2* above, double At*mybs1,mybs2* mutants were identified after CRISPR-Cas9-mediated gene editing (Figures 6B, S7A, and S7B) and analyzed in parallel with previously generated single At*mybs1* and At*mybs2* mutants. There were no detectable changes to rosette phenotype in the single mutants, but At*mybs1,mybs2* mutants were pale (Figures 6C–6F), and this was most noticeable after bolting (Figures 6G and 6H). There were no detectable changes in chloroplast size or number in mesophyll cells of single mutants, but chloroplasts were smaller in At*mybs1,mybs2* (Figures 6C–6F and 6I). Chloroplasts of At*mybs1,mybs2* contained underdeveloped thylakoids (Figure 6J). Consistent with these findings, chlorophyll content in the single mutants was indistinguishable from wild type but was ~40% lower in the double At*mybs1,mybs2* mutant (Figure 6K). Unfolding of the apical hook and greening of cotyledons were evident in both At*glk1,glk2* and At*mybs1,mybs2* double mutants, but greening appeared slower than in controls (Figure 6L). For the first 12 h after light was perceived, chlorophyll content was indistinguishable in At*glk1,glk2* and At*mybs1,mybs2* double mutants, but by 24 h At*glk1,glk2* had less chlorophyll than At*mybs1,mybs2* (Figure 6M). F_v/F_m was reduced for the first 6 h of light in both mutant backgrounds, but this was more apparent in At*mybs1,mybs2* (Figure 6N). At*GLK1* showed an increase in transcript abundance by 30 min of light that was maintained over 24 h (Figure 6O). By contrast, transcripts derived from At*MYBS2* had a higher abundance before exposure to light (Figures 6O and S6F).

Consensus binding sites for AtMYBS1&2 have previously been defined from DAP-seq,[35] and along with those we defined for *M. polymorpha*, they also comprise a TTATC core (Figure 7A).

Figure 4. Mp*RR-MYB5&2* target photosynthesis genes

(A and B) Schematic of thylakoid membranes with photosystems (modified from Waters et al.[10] and Tu et al.[25]), and heatmap indicating lower transcript abundance of genes encoding components of photosystem II, photosystem I, and the cytochrome b_6f complex in Mp*rr-myb2*, Mp*rr-myb5*, Mp*glk*, as well as Mp*glk,rr-myb5*, and Mp*rr-myb5,myb2* double mutants.

(C) Immunoblot analysis of LHCA, PsbA, PsaC, and PsbS in Mp*rr-myb2*, Mp*glk*, Mp*rr-myb5*, and Mp*rr-myb5,myb2* double mutants.

(D) Distribution of MpRR-MYB2 and MpRR-MYB5 binding across the *M. polymorpha* genome. Venn diagram shows overlap between MpRR-MYB2 and MpRR-MYB5 binding sites and misregulated genes in Mp*rr-myb5,2* mutants.

(E) Motif sequence logos for MpRR-MYB2, MpRR-MYB5, and MpGLK[31] identified by DAP-seq and ChIP-seq, respectively.

(F) Fisher exact test and permutation testing for RR-MYB motifs in 500 bp promoters from photosynthesis versus 1,000 random gene sets. Red line indicates the frequency of the motif in *M. polymorpha* photosynthesis genes.

(G) Mp*RR-MYB5* and Mp*RR-MYB2* binding around photosynthesis genes. Gene body shown at the bottom of each panel. Schematic of effector—reporter assay based on luciferase (LUC) as an output. Bar graphs showing impact of Mp*RR-MYB2* or Mp*RR-MYB5* on expression from promoters of Mp*PsbQ*, Mp*PsaD*, and Mp*LHCB1*. Data presented as means and standard error of the mean (n = 4), asterisks indicate a statistically significant difference compared with control (*p < 0.05, **p < 0.01, n.s., non-significant, paired t test).

(H) MpRR-MYB5 and MpRR-MYB2 binding identified by DAP-seq around Mp*RR-MYB5* and Mp*RR-MYB2*. Gene body is shown at the bottom of each panel. Schematic of effector—reporter assay based on luciferase (LUC) as an output. Bar graphs showing impact of Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*GLK* on expression from promoters of Mp*RR-MYB5* and Mp*RR-MYB5*. Data presented as means and standard error of the mean (n = 4), asterisks indicate a statistically significant difference compared with control (*p < 0.05, **p < 0.01, n.s., non-significant, paired t test).

(I) Transcript abundance of Mp*GLK*, Mp*RR-MYB2*, and Mp*RR-MYB5* in control, Mp*rr-glk*, Mp*rr-myb5*, Mp*rr-myb2*, and Mp*rr-myb5,2* mutant backgrounds. Data presented as transcript per million (TPM), and p values of two-tailed t test are shown.

See also Figures S4 and S5.

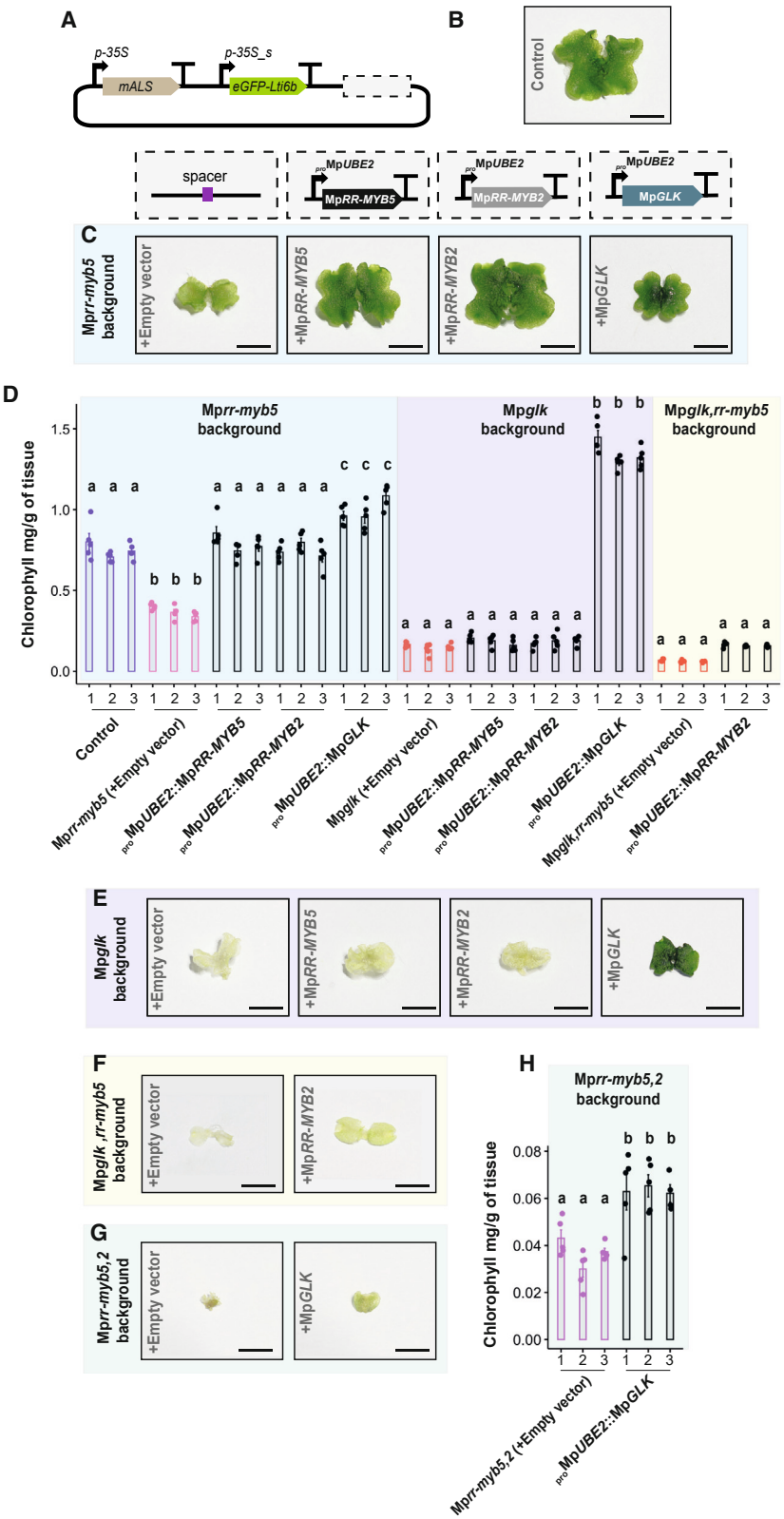


Figure 5. *M. polymorpha* RR-MYBs complement each other but there is limited redundancy between the MpRR-MYBs and MpGLK

(A) Schematic representation of design used to overexpress Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*GLK*. Each transcription factor was placed downstream of the GFP cassette (position shown with dashed box).

(B) Representative image of control plant (same as Figure 1D).

(C) Representative images of Mp*rr-myb5* complemented with Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*GLK*.

(D) Chlorophyll content in control and each of the mutant backgrounds compared with controls. Data presented as mean and standard error of the mean. Letters show ranking using a *post hoc* Tukey test (comparisons are made between groups highlighted within the same color rectangle, different letters indicating statistically significant differences at $p < 0.01$), $n = 5$.

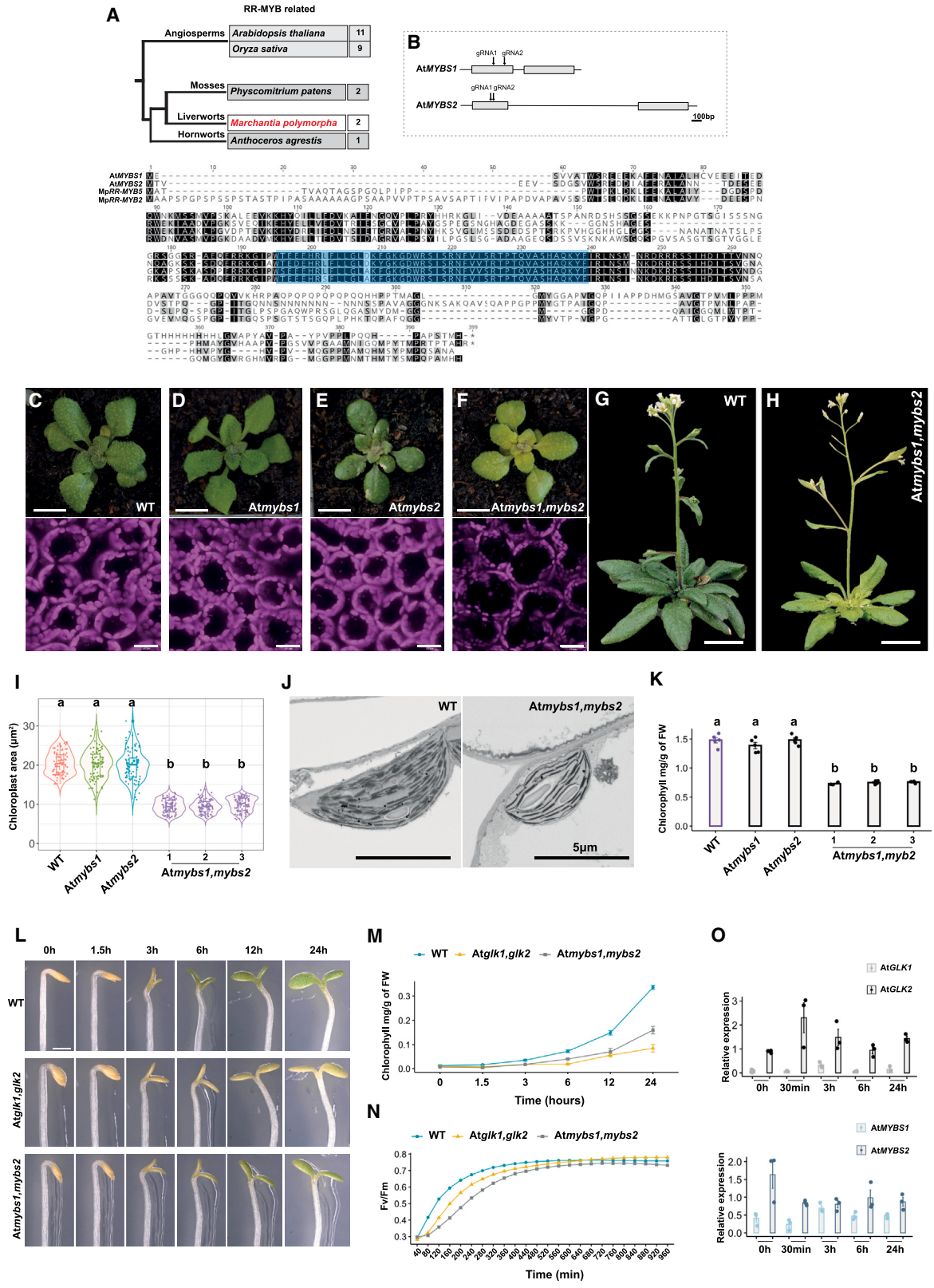
(E) Representative images of Mp*rr-glk* complemented with Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*GLK*.

(F) Representative images of Mp-*glk,rr-myb5* mutant with and without complemented by Mp*RR-MYB2*.

(G) Representative images of Mp*rr-myb5,2* mutant with and without complementation by Mp*GLK*.

(H) Chlorophyll content in Mp*rr-myb5,2* mutant with and without complementation from Mp*GLK*. Scale bars: 2 mm in (C) and (E)–(G).

See also Figure S5.



(legend on next page)

Fisher exact and permutation tests indicated that this motif is more common in promoters of photosynthesis genes than would be expected by chance (Figure 7B). RNA-seq of At*mybs1,mybs2* (Data S1) identified 2,470 transcripts that were less abundant compared with controls (Figure 7C, *padj* value $\leq$ 0.01, LFC $\geq$ 0.5), and of these, 122 can be bound by AtMYBS1 or AtMYBS2 (Figure 7C; Data S1; [35]). The Fisher's exact test and permutation analysis indicated that this is greater overlap than expected by chance (Figure S7C). Genes downregulated in the At*mybs1,mybs2* mutant compared with wild type (Figures S7D and S7E) included those encoding enzymes of chlorophyll biosynthesis pathway and components of the light-harvesting complexes as well as photorespiration and the Calvin-Benson-Bassham cycle (Figures 7D and S7F). We selected six photosynthesis genes that appeared to be under control of AtMYBS1 and AtMYBS2 for testing in effector assays. In each case, either AtMYBS1 or AtMYBS2 was able to activate expression (Figure 7E). Re-examination of publicly available DAP-seq data[35] showed that either or both AtMYBS1 or AtMYBS2 bind strongly to the promoters of these photosynthesis genes but also the promoter of 49 additional photosynthesis genes (Figure S7G; Data S3). We also tested whether AtMYBS1 or AtMYBS2 can activate themselves as well as At*GLK1* and At*GLK2*. Although the magnitude of response varied, all four transcription factors could activate At*MYBS1* and At*MYBS2* (Figure 7F). Moreover, AtGLK2 and both AtMYBS proteins activated At*GLK1*, and AtMYBS1 and AtMYBS2 activated At*GLK2* (Figure 7F). This capacity for AtMYBS1 and AtMYBS2 to recognize their own promoters as well as At*GLK1* is consistent with binding detected from DAP-seq (Figure S7H). We note no binding of AtMYBS1 or AtMYBS2 to the At*GLK2* promoter was detected (Figure S7H), despite its capacity to activate this promoter in effector assays, perhaps due to low affinity binding. As in *M. polymorpha*, transcript profiling indicated that more complex links likely operate between the At*RR-MYBs* and At*GLK in vivo*, with At*GLK2* being upregulated in the double At*mybs1,mybs2* mutant (Figure 7G). Of the 98 genes downregulated in both At*mybs1,mybs2* and At*glk1,glk2*,[25] more were shared than would be expected by chance (Figures 7H and S7I; Data S1), and of these, 27 were related to photosynthesis, encoding, for example, components of the light-harvesting complexes and the photosystems (Figure S7J; Data S1). Again, consistent with a complex regulatory network, there was no clear correlation between expression profiles of these 98 genes and the RR-MYBs and GLKs after exposure to light (Figure S7K). Based on all the data above, we therefore propose that AtMYBS1&2 and AtGLK1&2 have the capacity to control chloroplast biogenesis in *A. thaliana*, to bind and regulate each other, and that *in planta*, this is likely part of a complex regulatory framework composed of both direct and indirect links allowing photosynthesis gene expression to be tuned to the environment.

DISCUSSION

Chloroplasts allow photosynthesis, nitrogen and sulfur assimilation, as well as the biosynthesis of amino acids, fatty acids, and carotenoids and so understanding their biogenesis has long been of interest.[36–40] Moreover, a targeted reengineering of the process could contribute to crop development.[6,7] Indeed, improvements in yield have been reported after over-expression of *Sedoheptulose bisphosphatase*,[41,42] faster relaxation of non-photochemical quenching of photosystem II[43] and rerouting of photorespiration.[44] A complementary approach to improving photosynthesis predicted to increase yield by up to 50% would be to convert C_3 crops to use the more efficient C_4 pathway.[45,46] Introducing C_4 photosynthesis in C_3 crops such as rice would require a remodeling of chloroplast biogenesis in mesophyll and bundle sheath cells.[47]

In land plants, the GLK family of transcription factors are master regulators of chloroplast biogenesis, and CGA1 and GNC from the GATA family are considered ancillary players.[8,10,48] Over-expression of *GLK* in rice is sufficient to increase chloroplast occupancy of cells such as the bundle sheath and thus to partially phenocopy traits associated with the efficient C_4 pathway.[38] However, we have an incomplete understanding of transcription factors allowing chloroplast development. For

Figure 6. RR-MYBs control chlorophyll biogenesis in *A. thaliana*

(A) RR-MYB-related transcription factors in *A. thaliana*, rice, and bryophytes. Amino acid alignment of MpRR-MYB5&2 from *M. polymorpha* and AtMYBS1 and AtMYBS2 from *A. thaliana*. Characteristic SHAQK(Y/F)F DNA binding motif containing domains highlighted with blue shading.

(B) Schematic representation of At*MYBS1* and At*MYBS2* gene structure showing exons as gray rectangles and guide (g) RNAs positions for gene editing.

(C–F) Images of 2-week old seedlings of wild-type, At*mybs1*, At*mybs2*, and At*mybs1,mybs2* mutants, Scale bars represent 7 mm. Representative images after confocal laser scanning microscopy also shown below, with scale bars represent 25 μm.

(G and H) Representative images of wild-type and At*mybs1,mybs2* mutants with inflorescence. Scale bars represent 1.5 cm.

(I) Violin plots of chloroplast area for wild-type, At*mybs1*, At*mybs2*, and At*mybs1,mybs2* mutants. Box and whiskers represent the 25 to 75 percentile and minimum-maximum distributions of the data. Letters show ranking using a *post hoc* Tukey test (with different letters indicating statistically significant differences at $p < 0.01$), $n = 100$.

(J) Representative transmission electron microscopy images of wild-type and At*mybs1,mybs2* mutants. Scale bars represent 5 μm.

(K) Chlorophyll content of wild-type, At*mybs1*, At*mybs2*, and At*mybs1,mybs2* mutants (three independent lines were used for measurements). Letters show ranking using a *post hoc* Tukey test (with different letters indicating statistically significant differences at $p < 0.01$). Data presented as means and standard error of the mean, $n = 5$.

(L) Representative images of wild-type, At*glk1,glk2*, and At*mybs1,mybs2* seedlings undergoing de-etiolation. Scale bar represents 500 μm.

(M) Chlorophyll content wild-type, At*glk1,glk2*, and At*mybs1,mybs2* seedlings during the de-etiolation time course. Data presented as mean and standard error of the mean, $n = 3$.

(N) Chlorophyll fluorescence parameter F_v/F_m during de-etiolation of wild-type, At*glk1,glk2*, and At*mybs1,mybs2*.

(O) Transcript abundance of At*GLK1*, At*GLK2*, At*MYBS1*, and At*MYBS2* genes during de-etiolation. *UBP6* (At1g51710) was used as internal control. Data presented as means and standard error of the mean, $n = 3$.

See also Figure S6.

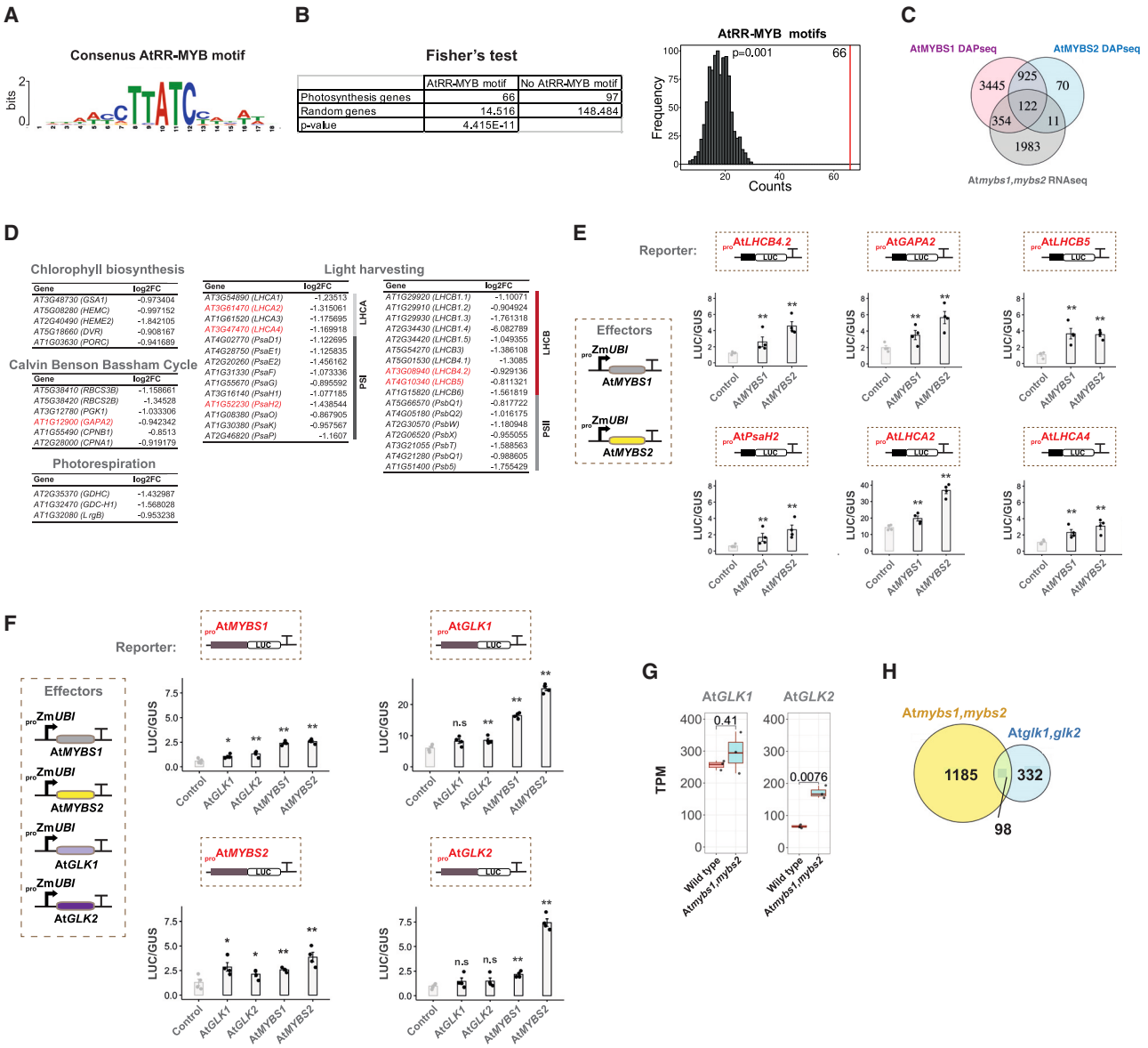


Figure 7. AtMYBS1&2 can bind and activate photosynthesis genes as well as transcription factors involved in chloroplasts biogenesis

(A) Motif sequence logo for AtRR-MYB.[35]

(B) Fisher exact test and permutation testing for AtRR-MYB motifs in 500 bp promoters from photosynthesis versus 1,000 random gene sets, red line indicates frequency in *A. thaliana* photosynthesis genes.

(C) Venn diagram showing the overlap between AtMYBS1 and AtMYBS2 binding sites and misregulated genes in At*mybs1,mybs2* mutants.

(D) Genes relating to chlorophyll biosynthesis, the Calvin-Benson-Bassham cycle, photorespiration, and the light-harvesting complexes that were downregulated in At*mybs1,mybs2* (highlighted with red, genes selected for the effector assays).

(E) Effector assays to test interactions between AtMYBS1&2 and photosynthesis genes. Schematic of effector-reporter (LUC) constructs used to assess binding of AtMYBS1 and AtMYBS2 to At*LHCB4*, At*GAPA2*, At*LHCB5*, At*PsaH2*, At*LHCA2*, and At*LHCA4* promoter regions. Data are presented as mean and standard error of the mean, $n = 4$. Asterisks indicate a statistically significant difference compared with control (*$p < 0.05$, **$p < 0.01$, t test).

(F) Effector assays to test interactions between AtMYBS1&2 and AtGLK1&2. Schematic of effector-reporter (LUC) constructs used as well as bar graphs showing impact of each transcription factor on each other. Data presented as means and standard error of the mean, $n = 4$). Asterisks indicate a statistically significant difference compared with control (*$p < 0.05$, **$p < 0.01$, n.s., non-significant t test).

(G) Transcript abundance in transcripts per million (TPM) of At*GLK1* and At*GLK2* in wild-type and At*mybs1,mybs2* mutant background. *p* values of two-tailed t test are shown.

(H) Venn diagram showing overlap between misregulated genes in At*glk1,glk2* and At*mybs1,mybs2* mutants (*padj* ≤ 0.01, $\log_2$FC $\geq$ 1-fold).

See also Figure S7.

example, *GLK* and *CGA1/GNC* loss-of-function mutants still possess small chloroplasts[9] indicating that either these mutants are hypomorphic or that additional unidentified actors control chloroplast biogenesis. Here, we report two RR-MYB-related transcription factors that act redundantly to control chlorophyll biosynthesis and photosynthesis-associated gene expression in the bryophyte *M. polymorpha*. Homologs control chloroplast biogenesis in *A. thaliana* indicating functional conservation between these species. Interestingly, we were unable to identify null mutants lacking both GLK and RR-MYB in *M. polymorpha*. Indeed, although we used super-transformation of existing mutant alleles when attempting to generate triple mutants, we did not observe white sectors. Our inability to recover a triple Mp*glk,rr-myb5,2* mutants mirrors attempts to generate loss-of-function mutations in plastidial pathways such as amino acid, vitamin, nucleotide, or fatty acid biosynthesis and those involved in chloroplast protein translation that result in an arrest of embryo development in *A. thaliana*.[49,50] This often appears to coincide with the globular-to-heart transition stage when chloroplasts start to differentiate. Mutants in genes encoding plastidial proteins required for import, modification, and localization of indispensable proteins in the chloroplast are also often associated with embryo lethality.[49,50] It is therefore possible that plants lacking both GLK and the RR-MYBs are unable to differentiate chloroplasts from proplastids. We did not detect any effect on roots or rhizoids in the mutants that we analyzed, and so at present, we have no evidence that GLKs or RR-MYBs impact plastid development in non-photosynthetic organs.

The precise architecture of the gene regulatory network involving the RR-MYBs and GLK will need to be fully elucidated. Our current data indicate that in both *M. polymorpha* and *A. thaliana*, there is overlap in the types of photosynthesis genes controlled by RR-MYBs and GLKs. However, the RR-MYBs appear to target a broader set of targets to allow the response of genes associated with carbon fixation, photorespiration, and repair of protein complexes to be regulated during photomorphogenesis. Although Mp*RR-MYB5* over-expression failed to upregulate Mp*GLK* transcripts, MYB transcription factors commonly act in multimeric complexes involving basic-helix-loop-helix (bHLH) and WD40 proteins[51] or with myelocytomatosis (MYC) proteins.[52] It is therefore possible that additional partners need to be over-expressed in combination with the RR-MYBs to increase expression of Mp*GLK*. It is also well documented that *GLK* is subject to multiple levels of regulation that can be overcome when non-native versions of the gene are mis-expressed.[38] When Mp*GLK* was mis-expressed in the double Mp*rr-myb5,2* mutant, although chlorophyll content remained low (90% of wild type), it was significantly statistically increased. We interpret these data in two ways. Either both classes of transcription factors are needed to drive full photosynthesis gene expression, or Mp*GLK* is permissive for very early stages of chloroplast biogenesis, but full assembly of the photosynthetic apparatus is strengthened by RR-MYBs. A permissive role for GLK in initiating chloroplast biogenesis is consistent with its ability to convert normally non-photosynthetic cells to contain a large chloroplast compartment.[38] Our current data therefore support a model in which *GLK* permits early stages of chloroplast biogenesis, and RR-MYBs then extend targets to activate accumulation of additional photosynthesis transcripts. Although this conditioning can be partially overcome by over-expression of Mp*GLK* in the presence of RR-MYBs,[16] in the absence of Mp*RR-MYBs*, the impact of Mp*GLK* over-expression is limited.

As would be expected from their evolutionary distance, rewiring has taken place between these transcription factors and the structural photosynthesis genes they target in *A. thaliana* and *M. polymorpha*. In *A. thaliana*, this regulatory system is more complex, with regulation of At*GLK* by AtMYBS1&2 being more evident, and members of the GATA transcription factor family also controlling photosynthesis gene expression. Moreover, inducible over-expression of At*MYBS1* in *A. thaliana* has been reported to increase expression of photosynthesis genes.[24] Although similar sets of genes were downregulated in loss-of-function Mp*rr-myb5,2* mutants, we did not detect widespread upregulation of photosynthesis genes after over-expression of *M. polymorpha* RR-MYBs. Also consistent with rewiring between these species is the fact that compared with Mp*RR-MYB2*, the pale phenotype of Mp*RR-MYB5* indicates it plays a dominant role in chloroplast biogenesis, while in *A. thaliana*, neither single mutant was pale.

The role of RR-MYBs in controlling chloroplast biogenesis is supported by previous work. For example, in tomato, LeMYBI has been reported to bind the promoter of the *RBCS* gene.[53] And, although no effect on chloroplast biogenesis was reported, a reduction in *RBCS* and the *chlorophyll a/b binding protein 1* (*CAB1*) genes expression has been reported in At*mybs1* mutants.[54] Along with transcription factors belonging to the GLK, BBX, and nuclear factor-Y families, random forest analysis of gene expression recently predicted that RR-MYBs regulate photosynthesis gene expression.[24] Moreover, an inducible At*MYBS1* over-expressor line showed upregulation of photosynthesis genes that we detected as downregulated in At*mybs1,mybs2*.

Penetrance of the RR-MYBs on photosynthesis gene expression and chloroplast biogenesis in *M. polymorpha* was striking, with Mp*RR-MYBs* activating genes allowing CO_2 fixation as well as light harvesting. Chloroplasts of Mp*myb5,2* mutants were ~30% smaller than those of Mp*glk* mutants, and double Mp*myb5,2* mutants were paler than those of Mp*glk*. This appears to be because RR-MYBs control an overlapping but broader set of photosynthesis genes than those downstream of GLK. Previous work supports the notion that these two classes of transcription factors have shared targets, as co-binding of RR-MYB and GLK to photosynthesis genes has been proposed.[55] Such cooperative binding of transcription factors is thought to allow greater variety of expression outputs. The reach of the RR-MYBs appears extensive in that they control genes encoding enzymes of the Calvin-Benson-Bassham cycle and photorespiration but also assembly and repair of RuBisCO. It seems likely that the large number of genes encoding a wide range of components underpinning photosynthesis targeted by the RR-MYBs contributes to the severe perturbation to phenotype when their function is removed. Overall, the data are consistent with overlapping as well as distinct roles for these two classes of transcription factor.

In summary, from the analysis of *M. polymorpha* and *A. thaliana*, whose last common ancestor dates to around 400 million years ago, we propose a model in which both RR-MYBs and GLKs operate as master regulators of photosynthesis gene expression. In both species, the RR-MYBs play a conserved role in controlling photosynthesis gene expression, and their targets are broader than those documented for GLK. As RR-MYBs appear ubiquitous in land plants,[27] it seems plausible they play a conserved role in chloroplast biogenesis. Although we were unable to detect Mp*RR*-MYBs in the Zygnematophyceae algae that are sister to the land plants,[56,57] they are in fact present in the Klebsormidiophyceae and Charophyceae[57,58] (Figure S6B), representing the other two most closely related algal lineages to land plants. *GLK* homologs are present in green algae[56] and both GLK and RR-MYB motifs are present in promoter regions of these genes in *K. flaccidum* and *C. braunii* (Figure S7L). These data imply that *RR-MYBs* operated alongside GLK to control chloroplast biogenesis before the colonization of land by plants.

Limitations of the study

Identification of RR-MYBs as regulators of chloroplast biogenesis and photosynthesis gene expression opens up a number of areas to be addressed in the future. These include, for example, analyses such as chromatin immunoprecipitation sequencing (ChIP-seq) to define their targets in space and time and allow downstream networks to be quantitatively defined, and an understanding of how these networks are rewired in response to developmental and environmental signals. It is also the case that currently we do not know how the RR-MYBs are regulated by the light signaling networks that act to initiate photosynthesis gene expression during de-etiolation nor the hormonal networks that tune chloroplast development in different cell types. Increased visibility of low chlorophyll in *A. thaliana* after reproductive growth implies that the RR-MYBs affect senescence, and so whether they interact with known players such as oresara 1 (ORE1) and *Arabidopsis thaliana* activating factor 1 (ATAF1)[59,60] will need to be defined. It will also be important to discover whether the RR-MYBs are integrated into retrograde signalling[61] from the chloroplast. Lastly, another area is how and what proteins the RR-MYBs interact with and if they are post-translationally regulated. Defining such complexes will help understand not only the mode of action of these transcription factors but also how they themselves are regulated.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Plant growth and transformation
  - Chlorophyll content, fluorescence imaging, microscopy and starch quantification
  - RNA extraction and sequencing
  - Immunoblotting
  - Phylogenetic analysis
  - Mapping genome-wide binding capacity of Mp RR-MYB5 and MpRR-MYB2
  - Trans-activation assays

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.06.039.

ACKNOWLEDGMENTS

We thank Karin Müller, Filomena Gallo, and Georgina Lindop from the Cambridge Advanced Imaging Centre for sample preparation as well as support during image acquisition. We also thank Facundo Romani for their useful comments and support. This work was funded by BBSRC BBP0031171 and ERA-CAPS C4BREED grants to J.M.H. and BBSRC/EPSRC, BB/L014130/1, and BBSRC BB/F011458/1 to J.H. T.S. was supported by an SNSF Postdoc Mobility Fellowship (P500PB_203128) and an EMBO Long-Term Fellowship (ALTF 531-2019). 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.

AUTHOR CONTRIBUTIONS

N.E.Y., E.F., and J.M.H. conceived and designed the work. J.M.H. guided the execution of experiments and oversaw the project. N.E.Y., E.F., K.B., L.H., T.S., M.T., and P.J.D. did the experiments and analyzed the data. 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.

Received: November 14, 2023
Revised: May 21, 2024
Accepted: June 28, 2024
Published: July 23, 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. McFadden, G.I. (2014). Origin and Evolution of Plastids and Photosynthesis in Eukaryotes. Cold Spring Harb. Perspect. Biol. *6*, a016105. https://doi.org/10.1101/cshperspect.a016105.
3. 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.
4. Abdallah, F., Salamini, F., and Leister, D. (2000). A prediction of the size and evolutionary origin of the proteome of chloroplasts of Arabidopsis. Trends Plant Sci. *5*, 141–142. https://doi.org/10.1016/s1360-1385(00)01574-0.
5. Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., and Penny, D. (2002). Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. USA *99*, 12246–12251. https://doi.org/10.1073/pnas.182432999.
6. Long, S.P., Marshall-Colon, A., and Zhu, X.-G. (2015). Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. Cell *161*, 56–66. https://doi.org/10.1016/j.cell.2015.03.019.

7. Ort, D.R., Merchant, S.S., Alric, J., Barkan, A., Blankenship, R.E., Bock, R., Croce, R., Hanson, M.R., Hibberd, J.M., Long, S.P., et al. (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. Proc. Natl. Acad. Sci. USA *112*, 8529–8536. https://doi.org/10.1073/pnas.1424031112.

8. 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.

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. Wang, P., Fouracre, J., Kelly, S., Karki, S., Gowik, U., Aubry, S., Shaw, M.K., Westhoff, P., Slamet-Loedin, I.H., Quick, W.P., et al. (2013). Evolution of GOLDEN2-LIKE gene function in C_3 and C_4 plants. Planta *237*, 481–495. https://doi.org/10.1007/s00425-012-1754-3.

15. 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.

16. Yelina, N.E., Frangedakis, E., Wang, Z., Schreier, T.B., Rever, J., Tomaselli, M., Forestier, E., Billakurthi, K., Ren, S., Bai, Y., et al. (2024). Streamlined regulation of chloroplast development in the liverwort Marchantia polymorpha. Preprint at bioRxiv. https://doi.org/10.1101/2023.01.23.525199.

17. 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.

18. 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.

19. Flores-Sandoval, E., Romani, F., and Bowman, J.L. (2018). Co-expression and Transcriptome Analysis of Transcription Factors Supports Class C ARFs as Independent Actors of an Ancient Auxin Regulatory Module. Front. Plant Sci. *9*, 1345. https://doi.org/10.3389/fpls.2018.01345.

20. Sullivan, A.M., Arsovski, A.A., Lempe, J., Bubb, K.L., Weirauch, M.T., Sabo, P.J., Sandstrom, R., Thurman, R.E., Neph, S., Reynolds, A.P., et al. (2014). Mapping and Dynamics of Regulatory DNA and Transcription Factor Networks in *A. thaliana*. Cell Rep. *8*, 2015–2030. https://doi.org/10.1016/j.celrep.2014.08.019.

21. Bastakis, E. (2017). The Contribution of the GATA Transcription Factors GNC and GNL in the Greening of *Arabidopsis thaliana*. PhD thesis (Technische Universität München).

22. Bursch, K., Toledo-Ortiz, G., Pireyre, M., Lohr, M., Braatz, C., and Johansson, H. (2020). Identification of BBX proteins as rate-limiting cofactors of HY5. Nat. Plants *6*, 921–928. https://doi.org/10.1038/s41477-020-0725-0.

23. Zhao, P., Cui, R., Xu, P., Wu, J., Mao, J.-L., Chen, Y., Zhou, C.-Z., Yu, L.-H., and Xiang, C.-B. (2017). ATHB17 enhances stress tolerance by coordinating photosynthesis associated nuclear gene and ATSIG5 expression in response to abiotic stress. Sci. Rep. *7*, 45492. https://doi.org/10.1038/srep45492.

24. Halpape, W., Wulf, D., Verwaaijen, B., Stasche, A.S., Zenker, S., Sielemann, J., Tschikin, S., Viehöver, P., Sommer, M., Weber, A.P.M., et al. (2023). Transcription factors mediating regulation of photosynthesis. Preprint at bioRxiv. https://doi.org/10.1101/2023.01.06.522973.

25. 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.

26. Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. Trends Plant Sci. *15*, 573–581. https://doi.org/10.1016/j.tplants.2010.06.005.

27. Du, H., Wang, Y.-B., Xie, Y., Liang, Z., Jiang, S.-J., Zhang, S.-S., Huang, Y.-B., and Tang, Y.-X. (2013). Genome-wide identification and evolutionary and expression analyses of MYB-related genes in land plants. DNA Res. *20*, 437–448. https://doi.org/10.1093/dnares/dst021.

28. 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.

29. 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.

30. P.J. Lea and R.C. Leegood, eds. (1998). Plant Biochemistry and Molecular Biology (Wiley).

31. Yelina, N.E., Frangedakis, E., Wang, Z., Schreier, T.B., Rever, J., Tomaselli, M., Forestier, E., Billakurthi, K., Ren, S., and Bai, Y. (2023). Streamlined regulation of chloroplast development in the liverwort *Marchantia polymorpha*. Preprint at bioRxiv. https://doi.org/10.1101/2023.01.23.525199.

32. Bartlett, A., O'Malley, R.C., Huang, S.C., Galli, M., Nery, J.R., Gallavotti, A., and Ecker, J.R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat. Protoc. *12*, 1659–1672. https://doi.org/10.1038/nprot.2017.055.

33. Pu, X., Yang, L., Liu, L., Dong, X., Chen, S., Chen, Z., Liu, G., Jia, Y., Yuan, W., and Liu, L. (2020). Genome-Wide Analysis of the MYB Transcription Factor Superfamily in *Physcomitrella patens*. Int. J. Mol. Sci. *21*, 975. https://doi.org/10.3390/ijms21030975.

34. Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhaofeng, G., Zhiqiang, L., Yunfei, Z., Xiaoxiao, W., Xiaoming, Q., et al. (2006). The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol. Biol. *60*, 107–124. https://doi.org/10.1007/s11103-005-2910-y.

35. O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell *165*, 1280–1292. https://doi.org/10.1016/j.cell.2016.04.038.

36. Witte, C.-P., and Herde, M. (2020). Nucleotide Metabolism in Plants. Plant Physiol. *182*, 63–78. https://doi.org/10.1104/pp.19.00955.

37. Waters, M.T., and Langdale, J.A. (2009). The making of a chloroplast. EMBO J. *28*, 2861–2873. https://doi.org/10.1038/emboj.2009.264.

38. 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_3 to C_4 Leaf Anatomy. Curr. Biol. *27*, 3278–3287.e6. https://doi.org/10.1016/j.cub.2017.09.040.

39. 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.

40. López-Juez, E. (2007). Plastid biogenesis, between light and shadows. J. Exp. Bot. *58*, 11–26. https://doi.org/10.1093/jxb/erl196.

41. Lefebvre, S., Lawson, T., Zakhleniuk, O.V., Lloyd, J.C., Raines, C.A., and Fryer, M. (2005). Increased sedoheptulose-1,7-bisphosphatase activity in transgenic tobacco plants stimulates photosynthesis and growth from an early stage in development. Plant Physiol. *138*, 451–460. https://doi.org/10.1104/pp.104.055046.

42. Driever, S.M., Simkin, A.J., Alotaibi, S., Fisk, S.J., Madgwick, P.J., Sparks, C.A., Jones, H.D., Lawson, T., Parry, M.A.J., and Raines, C.A. (2017). Increased SBPase activity improves photosynthesis and grain yield in wheat grown in greenhouse conditions. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*, 20160384. https://doi.org/10.1098/rstb.2016.0384.

43. Kromdijk, J., Głowacka, K., Leonelli, L., Gabilly, S.T., Iwai, M., Niyogi, K.K., and Long, S.P. (2016). Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. Science *354*, 857–861. https://doi.org/10.1126/science.aai8878.

44. South, P.F., Cavanagh, A.P., Liu, H.W., and Ort, D.R. (2019). Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. Science *363*, eaat9077. https://doi.org/10.1126/science.aat9077.

45. Hibberd, J.M., Sheehy, J.E., and Langdale, J.A. (2008). Using C_4 photosynthesis to increase the yield of rice-rationale and feasibility. Curr. Opin. Plant Biol. *11*, 228–231. https://doi.org/10.1016/j.pbi.2007.11.002.

46. von Caemmerer, S., Quick, W.P., and Furbank, R.T. (2012). The development of C_4 rice: current progress and future challenges. Science *336*, 1671–1672. https://doi.org/10.1126/science.1220177.

47. Langdale, J.A. (2011). C_4 cycles: past, present, and future research on C_4 photosynthesis. Plant Cell *23*, 3879–3892. https://doi.org/10.1105/tpc.111.092098.

48. Bastakis, E., Hedtke, B., Klermund, C., Grimm, B., and Schwechheimer, C. (2018). LLM-Domain B-GATA Transcription Factors Play Multifaceted Roles in Controlling Greening in Arabidopsis. Plant Cell *30*, 582–599. https://doi.org/10.1105/tpc.17.00947.

49. Budziszewski, G.J., Lewis, S.P., Glover, L.W., Reineke, J., Jones, G., Ziemnik, L.S., Lonowski, J., Nyfeler, B., Aux, G., Zhou, Q., et al. (2001). Arabidopsis genes essential for seedling viability: isolation of insertional mutants and molecular cloning. Genetics *159*, 1765–1778. https://doi.org/10.1093/genetics/159.4.1765.

50. Bryant, N., Lloyd, J., Sweeney, C., Myouga, F., and Meinke, D. (2011). Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in Arabidopsis. Plant Physiol. *155*, 1678–1689. https://doi.org/10.1104/pp.110.168120.

51. Ramsay, N.A., and Glover, B.J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci. *10*, 63–70. https://doi.org/10.1016/j.tplants.2004.12.011.

52. Dickinson, P.J., Kneřová, J., Szecówka, M., Stevenson, S.R., Burgess, S.J., Mulvey, H., Bågman, A.-M., Gaudinier, A., Brady, S.M., and Hibberd, J.M. (2020). A bipartite transcription factor module controlling expression in the bundle sheath of Arabidopsis thaliana. Nat. Plants *6*, 1468–1479. https://doi.org/10.1038/s41477-020-00805-w.

53. Rose, A., Meier, I., and Wienand, U. (1999). The tomato I-box binding factor LeMYBI is a member of a novel class of myb-like proteins. Plant J. *20*, 641–652. https://doi.org/10.1046/j.1365-313x.1999.00638.x.

54. Chen, Y.-S., Chao, Y.-C., Tseng, T.-W., Huang, C.-K., Lo, P.-C., and Lu, C.-A. (2017). Two MYB-related transcription factors play opposite roles in sugar signaling in Arabidopsis. Plant Mol. Biol. *93*, 299–311. https://doi.org/10.1007/s11103-016-0562-8.

55. 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.

56. Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T., Mori, H., Tajima, N., et al. (2014). Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. Nat. Commun. *5*, 3978. https://doi.org/10.1038/ncomms4978.

57. Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K.K., Haas, F.B., Vanderstraeten, L., Becker, D., Lang, D., et al. (2018). The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization. Cell *174*, 448–464.e24. https://doi.org/10.1016/j.cell.2018.06.033.

58. 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.

59. Rauf, M., Arif, M., Dortay, H., Matallana-Ramírez, L.P., Waters, M.T., Gil Nam, H., Lim, P.-O., Mueller-Roeber, B., and Balazadeh, S. (2013). ORE1 balances leaf senescence against maintenance by antagonizing G2-like-mediated transcription. EMBO Rep. *14*, 382–388. https://doi.org/10.1038/embor.2013.24.

60. Garapati, P., Xue, G.-P., Munné-Bosch, S., and Balazadeh, S. (2015). Transcription Factor ATAF1 in Arabidopsis Promotes Senescence by Direct Regulation of Key Chloroplast Maintenance and Senescence Transcriptional Cascades. Plant Physiol. *168*, 1122–1139. https://doi.org/10.1104/pp.15.00567.

61. Chan, K.X., Phua, S.Y., Crisp, P., McQuinn, R., and Pogson, B.J. (2016). Learning the Languages of the Chloroplast: Retrograde Signaling and Beyond. Annu. Rev. Plant Biol. *67*, 25–53. https://doi.org/10.1146/annurev-arplant-043015-111854.

62. 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.

63. 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.

64. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202–W208. https://doi.org/10.1093/nar/gkp335.

65. Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D., et al. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *48*, D87–D92. https://doi.org/10.1093/nar/gkz1001.

66. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018. https://doi.org/10.1093/bioinformatics/btr064.

67. 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.

68. Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. *47*, W256–W259. https://doi.org/10.1093/nar/gkz239.

69. 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.

70. 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.

71. 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.

72. 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.

73. 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.

74. 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.

75. 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.

76. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

77. 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.

78. Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382–2383. https://doi.org/10.1093/bioinformatics/btv145.

79. Frangedakis, E., Waller, M., Nishiyama, T., Tsukaya, H., Xu, X., Yue, Y., Tjahjadi, M., Gunadi, A., Van Eck, J., Li, F.-W., et al. (2021). An Agrobacterium-mediated stable transformation technique for the hornwort model *Anthoceros agrestis*. New Phytol. *232*, 1488–1505. https://doi.org/10.1111/nph.17524.

80. 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.

81. 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 (Bethesda) *12*, jkac105. https://doi.org/10.1093/g3journal/jkac105.

82. Romani, F., Sauret-Güeto, S., Rebmann, M., Annese, D., Bonter, I., Tomaselli, M., Dierschke, T., Delmans, M., Frangedakis, E., Silvestri, L., et al. (2024). The landscape of transcription factor promoter activity during vegetative development in Marchantia. Plant Cell *36*, 2140–2159. https://doi.org/10.1093/plcell/koae053.

83. Fauser, F., Schiml, S., and Puchta, H. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. Plant J. *79*, 348–359. https://doi.org/10.1111/tpj.12554.

84. Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc. Natl. Acad. Sci. USA *112*, 3570–3575. https://doi.org/10.1073/pnas.1420294112.

85. Ursache, R., Fujita, S., Dénervaud Tendon, V., and Geldner, N. (2021). Combined fluorescent seed selection and multiplex CRISPR/Cas9 assembly for fast generation of multiple Arabidopsis mutants. Plant Methods *17*, 111. https://doi.org/10.1186/s13007-021-00811-9.

86. Logemann, E., Birkenbihl, R.P., Ülker, B., and Somssich, I.E. (2006). An improved method for preparing Agrobacterium cells that simplifies the Arabidopsis transformation protocol. Plant Methods *2*, 16. https://doi.org/10.1186/1746-4811-2-16.

87. Edwards, K., Johnstone, C., and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res. *19*, 1349. https://doi.org/10.1093/nar/19.6.1349.

88. Schreier, T.B., Müller, K.H., Eicke, S., Faulkner, C., Zeeman, S.C., and Hibberd, J.M. (2024). Plasmodesmal connectivity in C_4 *Gynandropsis gynandra* is induced by light and dependent on photosynthesis. New Phytol. *241*, 298–313. https://doi.org/10.1111/nph.19343.

89. 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.

90. 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.

91. 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.

92. 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.

93. 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.

94. 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.

95. Szövényi, P., Frangedakis, E., Ricca, M., Quandt, D., Wicke, S., and Langdale, J.A. (2015). Establishment of *Anthoceros agrestis* as a model species for studying the biology of hornworts. BMC Plant Biol. *15*, 98. https://doi.org/10.1186/s12870-015-0481-x.

96. Page, M.T., Parry, M.A.J., and Carmo-Silva, E. (2019). A high-throughput transient expression system for rice. Plant Cell Environ. *42*, 2057–2064. https://doi.org/10.1111/pce.13542.

97. Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. *6*, 3901–3907. https://doi.org/10.1002/j.1460-2075.1987.tb02730.x.

98. 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.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-actin (plant)	Sigma	Cat#A0480; RRID:AB_476670
Antimouse-HRP	Sigma	Cat#A9044; RRID:AB_258431
LhcA1	Agrisera	Cat#AS01 005; RRID:AB_2135333
PsbS	Agrisera	Cat#AS09 533; RRID:AB_2155455
PsbA	Agrisera	Cat#AS05 084; RRID:AB_2172611
PsaC	Agrisera	Cat#AS10 939; RRID:AB_10754201
Goat anti-Rabbit IgG (H&L), HRP conjugated	Agrisera	Cat#AS09 602; RRID:AB_1966902
Bacterial and virus strains		
E. coli DH5a 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
Perchloric acid	Fisher	Cat #12993564
α-Amylase	Roche	Cat #10102814001
Amyloglucosidase	Roche	Cat#10102814001
Hexokinase	Roche	Cat#11426362001
Glucose-6-phosphate dehydrogenase	Roche	Cat#10165875001
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
Critical commercial assays		
The iBind Western Starter Kit	ThermoFisher	Cat#SLF1000S
QIAprep spin miniprep kit	QIAGEN	Cat#27104
NucleoSpin RNA/Protein kit	MACHEREY-NAGEL	Cat#740933.250
ECL Select Western Blotting Detection Reagent	GE	Cat#GERPN2235
NuPAGE gel	Invitrogen	Cat#NP0322BOX
LUC assay reagent	Promega	Cat#E1483
Passive lysis buffer	Promega	Cat#E1941
ZymoPURE™ II Plasmid Midiprep Kit	Zymo research	Cat#D4200
Turbo DNA-free kit	Invitrogen	Cat#AM1907
RNeasy Plant kit	QIAGEN	Cat#74903

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw sequencing data:*Marchantia polymorpha*	This paper	NCBI Sequence Read Archive (SRA): BioProject ID: PRJNA1039314
Raw sequencing data: *Arabidopsis thaliana*	This paper	NCBI Sequence Read Archive (SRA): BioProject ID: PRJNA1039314
Raw DAP-sequencing data: *Marchantia polymorpha*	This paper	NCBI Sequence Read Archive (SRA): BioProject ID: PRJNA1039314
Raw data from Figures and plasmid maps were deposited on Mendeley	This paper	Mendeley Data: https://doi.org/10.17632/3thmvgtxzp.1
Experimental models: Organisms/strains		
Arabidopsis thaliana; Col-0 (wild-type)	Widely distributed	N/A
Marchantia polymorpha, Cam-1 and Cam-2	Delmans et al.[62]	N/A
Arabidopsis T-DNA insertion mutant At*mybs1*	Nottingham Arabidopsis Stock Centre (NASC)	Stock number: SAIL_1184_D04
Arabidopsis T-DNA insertion mutants At*mybs2*	Nottingham Arabidopsis Stock Centre (NASC)	Stock number: SALK_150774
Oligonucleotides		
See Table S3	This paper	N/A
Software and algorithms		
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.[63]	RRID:SCR_015687; https://bioconductor.org/packages/release/ bioc/html/DESeq2.html
MEME suite	Bailey et al.[64]	RRID:SCR_001783; https://meme-suite.org/meme
JASPAR	Fornes et al.[65]	RRID:SCR_003030; https://jaspar.genereg.net
FIMO tool	Grant et al.[66]	RRID:SCR_001783; https://meme-suite.org/meme/tools/fimo
SH-aLRT test	Guindon et al.[67]	N/A
iTOL	Letunic and Bork[68]	RRID:SCR_018174; https://itol.embl.de
MAFFT	Katoh and Standley[69]	RRID:SCR_011811; https://mafft.cbrc.jp/alignment/software/
TrimAl	Capella-Gutiérrez et al.[70]	RRID:SCR_017334; http://trimal.cgenomics.org
iQTree	Nguyen et al.[71]	RRID:SCR_017254; http://www.iqtree.org
ModelFinder	Kalyaanamoorthy et al.[72]	http://www.iqtree.org/ModelFinder/
Kallisto	Bray et al.[73]	RRID:SCR_016582; https://pachterlab.github.io/kallisto/
Bowtie2-2.5.3-macos-x86_64	Langmead and Salzberg[74]	RRID:SCR_016368; https://github.com/BenLangmead/bowtie2/releases
Samtools-1.19.2	Danecek et al.[75]	RRID:SCR_002105; https://github.com/samtools/samtools/releases/
Bedtools2	Quinlan and Hall[76]	RRID:SCR_006646; https://github.com/arq5x/bedtools2
MACS3-3.0.0.dist-info	Zhang et al.[77]	RRID:SCR_013291; https://github.com/macs3-project/MACS
IGV_2.17.2	N/A	https://data.broadinstitute.org/igv/projects/downloads/2.17/
ChIPseeker	Yu et al.[78]	RRID:SCR_021322; https://www.bioconductor.org/packages/release/bioc/html/ChIPseeker.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
Plasmid: OP-023 CDS12-eGFP	Sauret-Gueto et al.[29]	N/A
Plasmid: OP-020 CDS_hph	Sauret-Gueto et al.[29]	N/A
Plasmid: OP-037 CTAG_Lti6b	Sauret-Gueto et al.[29]	N/A
Plasmid: OP-054 3TER, _Nos-35S	Sauret-Gueto et al.[29]	N/A
Plasmid: OP-049 PROM_35S	Sauret-Gueto et al.[29]	N/A
Plasmid: OP-47 PROM_MpUBE2	Sauret-Gueto et al.[29]	N/A
Plasmid: OP-48 5UTR_MpUBE2	Sauret-Gueto et al.[29]	N/A
Plasmid: OP-063, L1_HyR-Ck1	Sauret-Gueto et al.[29]	N/A
Plasmid: OP073 L1_Cas9-Ck4,	Sauret-Gueto et al.[29]	N/A
Plasmid: OP-076 L2_lacZgRNA-Cas9-CsA	Sauret-Gueto et al.[29]	N/A
Plasmid: L1_35S_s::eGFP-Lti6b	Frangedakis et al.[79]	N/A
Plasmid: pMpGE013	Sugano et al.[80]	Addgene, Plasmid #108681
Plasmid: pZmUBIpro::GUS-Tnos	This paper	N/A
Plasmid: pZmUBIpro::MpGLK-T35S	This paper	N/A
Plasmid: pZmUBIpro::MpRR-MYB2-T35S	This paper	N/A
Plasmid: pZmUBIpro::MpRR-MYB5-T35S	This paper	N/A
Plasmid: pZmUBIpro::AtGLK1-T35S	This paper	N/A
Plasmid: pZmUBIpro::AtGLK2-T35S	This paper	N/A
Plasmid: pZmUBIpro::AtMYBS1-T35S	This paper	N/A
Plasmid: pZmUBIpro::AtMYBS2-T35S	This paper	N/A
Plasmid: pMpGLK::LUC-T35S - promoter fragment size: 1666bp	This paper	N/A
Plasmid: pMpRR-MYB5::LUC-T35S - promoter fragment size: 1792bp	This paper	N/A
Plasmid: pMpRR-MYB2::LUC-T35S - promoter fragment size: 1792bp	This paper	N/A
Plasmid: pAtGLK1::LUC-T35S - promoter fragment size: 2096bp	This paper	N/A
Plasmid: pAtGLK2::LUC-T35S - promoter fragment size: 1473bp	This paper	N/A
Plasmid: pAtMYBS1::LUC-T35S - promoter fragment size: 1581bp	This paper	N/A
Plasmid: pAtMYBS2::LUC-T35S - promoter fragment size: 2621bp	This paper	N/A
Plasmid: pAtLHCB4.2::LUC-T35S (AT3G08940) - promoter fragment size: 535bp	This paper	N/A
Plasmid: pAtPsaH2::LUC-T35S (AT1G52230) - promoter fragment size: 511bp	This paper	N/A
Plasmid: pAtLHCB5::LUC-T35S (AT4G10340) - promoter fragment size: 541bp	This paper	N/A
Plasmid: pAtGAPA2::LUC-T35S (AT1G12900) - promoter fragment size: 552bp	This paper	N/A
Plasmid: pAtLHCA2::LUC-T35S (AT3G61470) - promoter fragment size: 543bp	This paper	N/A
Plasmid: pAtLHCA4::LUC-T35S (AT3G47470) - promoter fragment size: 509bp	This paper	N/A
Plasmid: pMpPsaN::LUC-T35S (Mp4g20410) - promoter fragment size: 573bp	This paper	N/A
Plasmid: pMpPetA::LUC-T35S (Mp1g27760) - promoter fragment size: 572bp	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: pMpPORA::LUC-T35S – (Mp8g17840) - promoter fragment size: 790bp	This paper	N/A
Plasmid: pMpDVR::LUC-T35S – (Mp5g20690) - promoter fragment size: 894bp	This paper	N/A
Plasmid: pMpPsad::LUC-T35S – (Mp5g04200) - promoter fragment size: 727bp	This paper	N/A
Plasmid: pMpLHCB1::LUC-T35S – (Mp7g06740) - promoter fragment size: 893bp	This paper	N/A
Plasmid: pMpPsbQ::LUC-T35S – (Mp4g07510) - promoter fragment size: 712bp	This paper	N/A
Plasmid: p35S::mALS-T35S– 35S::eGFP-Lti6b	This paper	N/A
Plasmid: p35S::mALS-T35S– 35S::eGFP-Lti6b - pMpUBE2::MpRR-MYB2(gRNA resistant)-T35SNOS	This paper	N/A
Plasmid: p35S::mALS-T35S– 35S::eGFP-Lti6b - pMpUBE2::MpRR-MYB5(gRNA resistant)-T35SNOS	This paper	N/A
Plasmid: p35S::mALS-T35S– 35S::eGFP-Lti6b - pMpUBE2::MpGLK(gRNA resistant)-T35SNOS	This paper	N/A
Plasmid: p35S::Hyg-T35S– 35S::eGFP-Lti6b	This paper	N/A
Plasmid: p35S::Hyg-T35S– 35S::eGFP-Lti6b - pMpUBE2::MpRR-MYB2-T35SNOS	This paper	N/A
Plasmid: p35S::Hyg-T35S– 35S::eGFP-Lti6b - pMpUBE2::MpRR-MYB5-T35SNOS	This paper	N/A
Plasmid: p35S::Hyg-T35S– 35S::eGFP-Lti6b - pMpUBE2::MpRR-MYB5-T35SNOS - pMpUBE2::MpRR-MYB2-T35SNOS	This paper	N/A

RESOURCE AVAILABILITY

Lead contact

Julian M Hibberd (jmh65@cam.ac.uk) is the lead contact for data and plant material.

Materials availability

Plasmids and plant material generated in this study will be made available upon request from the lead contact.

Data and code availability

- Transcriptome and DAP-seq data are available in the NCBI Sequence Read Archive: PRJNA1039314. This study analysed existing, publicly available transcriptome and DAP-seq data, the accession numbers for these datasets are listed in the Tables S1 and S2; Data S1 and S3 and the key resources table. Raw data from Figures, including original Western blots images, were deposited on Mendeley and are publicly available as of the date of publication. The DOI is listed in the key resources table.
- All original codes were deposited on Mendeley and are publicly available as of the date of publication. The DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Plant growth and transformation

M. polymorpha Cam-1 (male) and Cam-2 (female) were grown at 22 °C on half-strength Gamborg B5 medium plus vitamins (pH 5.8) and 1.2% (w/v) agar under continuous light at 100 μmol m^{-2} s^{-1}. *Arabidopsis thaliana Col-0* was used throughout exception of At*mybs2* which is *Col-3*. Plants were grown on F2 soil (Levington, F20117800) under 16 hours light, 8 hours dark at 20 °C, 60% humidity, and 150 μmol m^{-2} s^{-1} light. Arabidopsis T-DNA insertion mutants At*mybs1* (SAIL_1184_D04) and At*mybs2* (SALK_150774) were obtained with T-DNA and zygosity confirmed by PCR (Figure S7M; Table S3). For gene editing gRNAs were predicted using CasFinder. Several gRNAs (Table S3) per target were tested.[81] Single gRNAs were cloned into L1_lacZgRNA-Ck3, L2_lacZgRNA-Cas9-CsA[29] or pMpGE013.[80] To complement Mp*rr-myb5* mutants the Mp*RR-MYB5* promoter[82] was used. For overexpression

Mp*RR-MYB2* and Mp*RR-MYB5* coding sequences were synthesised and cloned into the pUAP4 vector.[29] For complementation, gRNA resistant Mp*GLK*, Mp*RR-MYB2* and Mp*RR-MYB5* coding sequences were synthesised and cloned into the pUAP4 vector. OpenPlant parts used listed in STAR Methods. To generate At*mybs1,mybs2* mutants two gRNAs per gene were cloned into pEn-Chimera[83] after.[84] This placed guides into the pRU294 vector that has a codon optimised and intron-containing version of *Cas9* driven by the egg-cell specific *pEC1.2* promoter.[85] *A. thaliana* was transformed by floral dipping[86] and genotyping performed as reported previously.[87] T2 plants with confirmed edits were analysed.

For *M. polymorpha* transformation, 5 mL LB media were inoculated with 3-4 *Agrobacterium* colonies (*GV3101*: 50 μg/mL rifampicin, 25 μg/mL gentamicin) and the plasmid-specific selection antibiotic. The preculture was incubated at 28°C for 2 days at 110 rpm. 5 mL of 2 d *Agrobacterium* culture were centrifuged for 7 min at 2000 x *g*. The supernatant was removed and pellet re-suspended in 5 mL liquid KNOP (0.25g/L KH_2PO_4, 0.25g/L KCl, 0.25g/L $MgSO_4$ $7H_2O$, 1g/L $Ca(NO_3)_2$ $4H_2O$, 12.5mg/L $FeSO_4 7H_2O$, 30mM MES and pH5.5) plus 1% (w/v) sucrose and 100 μM acetosyringone. The culture was then incubated with shaking (120 rpm) at 28°C for 3-4 hours. Around 20 thallus fragments were transferred into a 6-well plate with 5 mL liquid KNOP medium supplemented with 1% (w/v) sucrose, 30 mM MES, pH 5.5, 80 μL of *Agrobacterium* culture and acetosyringone at a final concentration of 100 μM. Tissue was co-cultivated with *Agrobacterium* for 3 days on a shaker at 110 rpm, at 22°C with ambient light. Using a sterile plastic pipette, liquid was removed from each well and the thallus fragments were transferred onto plates with growth media containing the appropriate antibiotic (Chlorsulfuron 0.5 μM and Cefotaxime 100μg/mL). To facilitate spreading of thallus fragments 1-2 mL sterile water was added to the petri dish. To genotype *M. polymorpha* 3 × 3 mm pieces of thalli from individual plants were placed in 1.5 mL tubes and crushed in 100 μL genotyping buffer (100 mM Tris-HCl, 1 M KCl, 1 M KCl, and 10 mM EDTA, pH 9.5). Tubes were then placed at 70 °C for 15-20 mins and 380 μL sterile water added to each tube. 5 μL aliquots of the extract were used as a template for polymerase chain reactions.

Chlorophyll content, fluorescence imaging, microscopy and starch quantification

For chlorophyll quantification in *M. polymorpha* ∼30-50mg of 10-14 days old gemmalings were used with five biological replicates per genotype. Tissue was blotted dry before weighing and then transferred into a 1.5mL microfuge tube containing 1 mL of dimethyl sulfoxide (DMSO) (D8418, Sigma Aldrich) and incubated in the dark at 65 °C for 45 minutes. Samples were allowed to cool to room temperature for approximately one hour. Chlorophyll content was then measured using a NanoDrop™One/One C Microvolume UV-Vis Spectrophotometer (ThermoFisher) following the manufacturer's instructions. Chlorophyll fluorescence measurements were carried out using a CF imager (Technologica Ltd, UK). *M. polymorpha* plants were placed in the dark for 20 mins and a minimum weak measuring light beam (<1 μmol m^{-2} s^{-1}) applied to evaluate dark-adapted minimum fluorescence (F_o), and a subsequent saturating pulse of 3000 μmol m^{-2} s^{-1} used to evaluate dark-adapted maximum fluorescence (F_m). A total of three plants were measured per genotype and treatment. 20 μM DCMU (#45463, Sigma Aldrich) was added to half-strength MS media, and thalli placed in DCMU for 24 h before chlorophyll fluorescence measurements were obtained.

To assess de-etiolation in *A. thaliana*, seeds were sown on plates, stratified at 4°C in darkness for 3 days and then incubated at 22°C for 3 days before transfer to continuous white light with an intensity of 100 μmol m^{-2} s^{-1}. Chlorophyll was measured at 0, 1.5, 3, 6, 12 and 24 hours after exposure to light by pooling ∼50 cotyledon pairs into a 1.5 mL microfuge tube containing 100 μL of dimethyl sulfoxide (DMSO) (D8418, Sigma Aldrich), with three biological replicates per time point. Tubes were incubated in the dark at 65 °C for 30 minutes, samples were allowed to cool to room temperature for approximately one hour and chlorophyll was quantified using a nanodrop following the same procedure as described for *M. polymorpha*.

For confocal laser scanning microscopy of *M. polymorpha*, five to seven gemma were placed within a medium-filled gene frame together with 30 μL water prior to being sealed with a cover slip. Plants were imaged immediately using a Leica SP8X spectral fluorescent confocal microscope with either a 10X air objective (HC PL APO 10×/0.40 CS2) or 20X air objective (HC PL APO 20×/0.75 CS2). Excitation laser wavelength and captured emitted fluorescence wavelength windows were 488 nm, 498−516 nm for GFP, and 488 or 515nm, 670−700 nm for chlorophyll autofluorescence. For electron microscopy ∼2 mm^2 sections of 5-6 individual 3-week-old thalli or 2-week-old leaves were harvested, fixed, embedded and imaged as previously described.[88] Chloroplast area was measured using ImageJ and the Macro is deposited on Mendeley.

To quantify starch, *M. polymorpha* thallus tissue (300-400 mg) that had been grown for two weeks under continuous light with a light intensity of 100 μmol $m^{-2}s^{-1}$ was harvested into tubes containing 2 metal beads, flash-frozen in liquid nitrogen, and ground using a tissue lyser (settings: 2:30 min; 28Hz). 1000 μL 0.7 M perchloric acid was added into the tubes prior to vigorous vortexing, grinding again in a tissue lyser (settings: 1:30 min; 28Hz) and then being centrifuged for 5 minutes. The supernatant was transferred into a fresh tube, and 400 μL neutralisation buffer [2 M KOH, 400 mM MES] added to achieve pH 6-7. The precipitate was spun down for 4 minutes at maximum speed using a benchtop centrifuge. 1 mL of dH_2O was added to wash the pellet, vortexed to resuspend the pellet and spun at 3,000 *g* for 3 minutes. The pellet was washed by adding 1 mL 80% (v/v) EtOH and resuspended by vortexing. A total of three ethanol washes were performed, or until the final wash was clear. Excess EtOH was evaporated from the pellet by leaving the tube open in a fume hood for 20 minutes. The pellet was resuspended in 400 μL of dH_2O, and two aliquots of 200 μL were placed into 1.5 mL tubes and incubated at 100°C for 15 minutes. Controls comprised 190 μL NaAc buffer and 10 μL H_2O. Sample digests were set up by adding 190 μL 0.22M Na-Acetate Buffer (pH 4.8) and 10 μL of 9:1 mix of amyloglucosidase:α-amylase (Roche #10102857001) (Roche #10102814001). Reactions were incubated at 37°C for 2 hours. Insoluble material was spun out (3 min, max speed). A master mix was prepared consisting of 50 mM HEPES-NaOH (pH 7.4-7.6), 1mM $MgCl_2$, 1mM ATP, 1mM NAD,

and 1.4 Units of Hexokinase (Roche #11426362001). Reactions consisted of 50 μL Master mix, 50 μL sample, 3 μL of glucose-6-phosphate (Glucose-6-phosphate dehydrogenase (Roche #10165875001) diluted 4X in water) and dH_2O to a final volume of 200 μL. The reaction was conducted in a GREINER 96 F-BOTTOM microtiter plate using a CLARIOstar plate reader. The initial absorbance at 340nm was monitored. Using a multichannel pipette, 2 μL of diluted G-6-P dehydrogenase was dispensed. Immediately after, the absorbance was monitored at 340nm for minutes.

RNA extraction and sequencing

For *M. polymorpha*, RNA was extracted from 3-4 two-week old gemmae using the RNeasy Plant kit (#74903, Qiagen) with RLT buffer supplemented with beta-mercaptoethanol, and residual genomic DNA removed using the Turbo DNA-free kit (# AM1907, Invitrogen). 500 ng of DNase-treated RNA was used as template for cDNA preparation (SuperScript™ IV First-Strand Synthesis System, #18091050, Invitrogen) according to manufacturer's instructions except that reverse transcription was 40 minutes and used 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 six times and oligonucleotides (Table S3) used at a final concentration of 0.5 μM. Reaction conditions comprised initial denaturation at 94°C for 2 minutes followed by 40 cycles of 94°C for 15 seconds (denaturation) and 60°C for 1 minute (annealing, extension, and fluorescence reading). Primer sequences are in Table S3. 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, first strand cDNA was synthesised using random hexamer primers. Library concentration was measured on a Qubit instrument using the manufacturer's procedure (Thermo Fisher Scientific) followed by real-time qPCR quantification. Library size distribution was analysed on a bioanalyzer (Agilent) following the manufacturer's protocol. Quantified libraries were pooled and sequenced on a NovaSeq PE150 Illumina platform and 6G raw data per sample obtained. Adapter sequences were: 5' Adapter: 5'-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-3'. 3' Adapter: 5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGATGACTATCTCGTATGCCGTCTTCTGCTTG-3'

FastQC was used to assess read quality and TrimGalore (https://doi.org/10.5281/zenodo.5127899) to remove low-quality reads and adapters. Reads were pseudo-aligned using Kallisto[73] to the *M. polymorpha* genome version 5 (transcripts only, obtained from MarpolBase).[89] Mapping statistics for each library are provided in Data S1. DGE analysis was performed with DESeq2,[63] with padj-values < 0.01.

For *A. thaliana*, RNA was extracted from two-week old seedlings following the same pipeline. Library preparation and RNA sequencing was performed by Novogene (Cambridge, UK). FastQC was used to assess read quality and TrimGalore to remove low-quality reads and adapters. Reads were pseudo-aligned using Kallisto[73] to the *A. thaliana* TAIR10 genome (transcripts only). Mapping statistics for each library are provided in Data S1. DGE analysis was performed with DESeq2,[63] with padj-values < 0.01. For qPCR analysis, seeds were sown on plates and stratified at 4°C in darkness for 3 days, prior to incubation at 22°C for 3 days and transferred to continuous white light with an intensity of 100 μmol m^{-2} s^{-1}. Approximately 70-80 cotyledon pairs were collected into a 1.5 mL microfuge tubes at 0h, 30 min, 3h, 6h, and 24h after exposure to light. RNA was extracted using the RNeasy Plant kit (#74903, Qiagen) with RLT buffer supplemented with beta-mercaptoethanol, and residual genomic DNA removed using the Turbo DNA-free kit (# AM1907, Invitrogen). Primers used for Q-PCR are listed in Table S3.

Immunoblotting

Total protein was extracted from 200 mg of gemmaelings grown for two weeks on half strength Gamborg B5 medium including vitamins with 1.2% (w/v) agar, at 21 °C in continuous white light with an intensity of 150 μmol m^{-2} s^{-1}). Samples were ground in liquid nitrogen and total protein was extracted using the NucleoSpin RNA/Protein kit (#740933.250, MACHEREY-NAGEL). Total protein was quantified using the Klarsson protocol from the NucleoSpin RNA/Protein kit. Equal amounts of protein (4.5 μg) were separated by denaturing electrophoresis in NuPAGE gel (#NP0322BOX, Invitrogen) and transferred to nitrocellulose membranes using the iBlot2 Dry Blotting System (ThermoFisher). PsbA, PsbS, PsaC, and LHCA proteins were detected with antiLHCA1 antibody (1:4000 dilution) (# AS01 005, Agrisera), PsbA (1:4000 dilution) (# AS05 084, Agrisera), PsbS (1:4000 dilution) (# AS09 533, Agrisera), PsaC (1:4000 dilution) (# AS10 939, Agrisera) and Goat anti-rabbit-HRP (1:15000 dilution) (# AS09 602, Agrisera) antibodies. Actin was immunodetected with anti-actin (plant) (1:1500 dilution) (#A0480, Sigma) and (1:15000 dilution) anti-mouse-HRP (#A9044, Sigma) antibodies, using the iBind Western Starter Kit (#SLF1000S, ThermoFisher). Immunoblots were visualized using the ECL Select Western Blotting Detection Reagent (#GERPN2235, GE) following the manufacturer's instructions. Images were acquired using a Syngene Gel Documentation system G:BOX F3.

Phylogenetic analysis

To identify RR Myb-related/CCA1-like genes three approaches were combined. Firstly, RR Myb-related/CCA1-like genes for twenty-one plant genomes were mined from iTAK[90] and PlantTFDB,[91] Phytozome, Fernbase,[92] Phycozome and PhytoPlaza databases. Sequences for each species were aligned with the MpRR-MYB5 and *A. thaliana* MYBS1 and MYBS2 amino acid sequences using MAFFT.[69] Results were filtered manually to identify RR-Myb-related/CCA1-like orthologs distinguished from other Myb-related genes based on the conserved SHAKYF motif in the R1/2 domain. Secondly, BLASTP searches were performed against plant genomes in Phytozome v13, fern genomes (fernbase.org), hornworts genome (www.hornworts.uzh.ch),[93] green algae genomes in PhycoCosm (phycocosm.jgi.doe.gov), and 1KP using the MpRR-MYB5 and *A. thaliana* MYBS1 and MYBS2 amino acid amino

acid sequence. Identified RR-Myb-related/CCA1-like protein sequences were aligned using MAFFT and trimmed using TrimAl.[70] A maximum likelihood phylogenetic tree was inferred using iQTree,[71] ModelFinder[72] and ultrafast approximation for phylogenetic bootstrap[94] and SH-aLRT test.[67] The tree was visualised using iTOL.[68] Full list of sequences in Data S4.

Mapping genome-wide binding capacity of Mp RR-MYB5 and MpRR-MYB2

M. polymorpha genomic DNA was extracted as described previously.[95] DAP-seq was performed by CD Genomics. Briefly, genomic DNA (gDNA, 5 μg) was fragmented to ~200 bp. Fragmented gDNA was purified (Cat# A63880, Beckman) and overhangs resulting from fragmentation converted into blunt ends using T4 DNA polymerase, Klenow Fragment, and T4 Polynucleotide Kinase. After adding an 'A' base to the 3' end of the blunt phosphorylated DNA fragments, adapters were ligated to the ends of the DNA fragments. The adapter-ligated gDNA fragments were purified using the Agencourt AMPureXP Kit (Cat# A63880, Beckman). For protein expression the coding sequences of MpRR-MYB2 and MpRR-MYB5 were cloned into a pFN19K HaloTag T7 SP6 Flexi expression vector. Halo-MpRR-MYB2 or 5 fusion proteins were expressed using the TNT SP6 Coupled Wheat GermExtract System (Promega) following the manufacturer's specifications for expression in a 50 μL reaction with 2 hours incubation at 37°C. Expressed proteins were directly captured using MagneHalo Tag Beads (Promega). The protein-bound beads were incubated with 50 ng of adapter-ligated gDNA fragments on a rotator for 1 hour at room temperature in 50 μL wash/bind buffer. Beads were washed three times using the same wash buffer to remove unbound DNA fragments. The HaloTag beads were resuspended in 30 μL of the elution buffer and heated to 98 °C for 10 min to denature protein and release the bound DNA fragments into solution. The supernatant was then transferred to a new well, and 25 μL used in a 50 μL PCR employing the KAPA HiFi HotStart ReadyMixPCR Kit (Roche, Basel, Switzerland) for 10 cycles. PCR primers consisted of the full-length Illumina TruSeq Universal primer (5′-AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTCCGATC T-3′) and an Illumina TruSeq Index primer (5′-CAAGCAGAAGACGGCATACGAGAT-NNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′) where NNNNNN represents the 6 bp sequence index used for sample identification. PCR products were purified and selected using Agencourt AMPure XP Kit (Cat#A63880, Beckman), and resuspended in 20 μL nuclease-free water. DNA concentration was determined using a Qubit (Life Technologies, Burlington, Ontario, Canada). Eluted DNA fragments were sequenced on an Illumina NovaSeq6000 platform. Negative control mock (Input) DAP-seq libraries were prepared without the addition of protein to the beads. Data provided were then mapped against the *M. polymorpha* v6 genome or the TAIR10 *A. thaliana* genome using bowtie2[74] version 2.5.3, sorted and converted to bam format with samtools[75] version 1.19.2. Mapped reads were visualised using IGV_2.17.2, and peaks called with MACS3[77] version 3.0.0, with bedtools[76] used to retrieve genomic coordinates of peaks. Nearest genes around the peaks were retrieved using the ChIPseeker R package.[78]

MpRR-MYB2 and MpRR-MYB5 binding sites were identified by motif enrichment analysis (using meme-chip-5.5.5) using the 250bp flanking regions of the DAPseq identified peaks. AtMYBS1 and MYBS2 binding sites (motifs MA1186.1 and MA1399.1) were obtained from JASPAR.[65] Motifs were merged to create RR-MYB combined motifs and visualised using the Ceqlogo tool from MEME.[64] The MpGLK motif was obtained from.[16] To assess occurrence of GLK and RR-MYB motifs in gene promoter sequences (500 bp and 1.5kb upstream of the TSS) were scanned using FIMO.[66] 500bp was selected in order to focus analysis on core promoters rather than long distance enhancer elements and to reduce background signal associated with the increased probability of finding any motif by chance as the search space is greater. Promoters were then scored for presence or absence of each motif and percentage of photosynthesis of genes containing each calculated. The Fisher exact and permutation testing were used to quantify over-representation of RR-MYB motifs in photosynthesis genes. For the latter, to test the background distribution of RR-MYB motifs a set of random promoters of the same size as the list of photosynthesis genes from *M. polymorpha* or *A. thaliana* was extracted and FIMO ran to determine the presence of these motifs. This process was iterated 1000 times and distributions of the frequency of these motifs plotted. The frequency of motifs in promoters of photosynthesis genes was also determined by searching each promoter with FIMO. Permutation tests were performed to test whether the frequency of motifs was significantly different to the frequency found in random selected promoters. In Data S2 the FIMO tool[66] was used to scan promoter sequences of *A. thaliana* and *M. polymorpha* for matches to transcription factor binding motifs found in the JASPAR motif database.[65] To account for input sequence composition, a background model was generated using the fasta-get-markov tool from the MEME suite.[64] FIMO was then run with default parameters and a P value cut-off of 1×10^{-4}. Matches to GLK and RR-MYB combined motifs were highlighted in each output.

Trans-activation assays

Rice protoplasts were subjected to transformation as previously described.[96] Golden gate level 1 modules for transformation were isolated using the ZymoPURE™ II Plasmid Midiprep Kit, with ZmUBIpro::GUS-T35S (maize ubiquitin promoter and CAMV35S terminator module) serving as the transformation control. Transcription factor coding sequences were amplified using CTAB extracted *M. polymorpha* gDNA or synthesised by IDT, with *Bsa*I and *Bpi*I sites mutated, and then cloned into Golden gate SC level 0 modules. These modules were assembled into a level 1 module with a maize ubiquitin promoter and CAMV35S terminator module. Details of promoter fragments cloned can be found on Mendeley and in STAR Methods. For each transformation, 2 μg of transformation control plasmid, 5 μg of reporter plasmid, and 5 μg of effector plasmid were combined and mixed with 180 μL protoplasts. After 20 hours incubation in the light with intensity of 40 μmol m^{-2} s^{-1}, at 22°C, proteins were extracted using a passive lysis buffer (Promega). Quantification of GUS activity was performed using a fluorometric MUG assay.[97] MUG assay buffer (50mM phosphate buffer (pH 7.0), 10 mM EDTA-Na_2, 0.1% [v/v] Triton X-100, 0.1% [w/v] N-lauroylsarcosine sodium, 10 mM DTT, 2 mM

4-methylumbelliferyl-β-D-glucuronide (MUG)), with 20 μL of protein sample in a 200 μL total volume. The reaction was conducted at 37 °C in GREINER 96 F-BOTTOM microtiter plate using a CLARIOstar plate reader. 4-Methylumbelliferone (4-MU) fluorescence was recorded every 2 minutes for 20 cycles with excitation at 360 nm and emission detected at 450 nm. LUC activity was measured with 20 μL of protein sample and 100 μL of LUC assay reagent (Promega) using a Clariostar plate reader. Transcriptional activity from the promoter was calculated as the ratio of LUC luminescence to the rate of MU accumulation per second.

CellPress
OPEN ACCESS

Supplemental figures

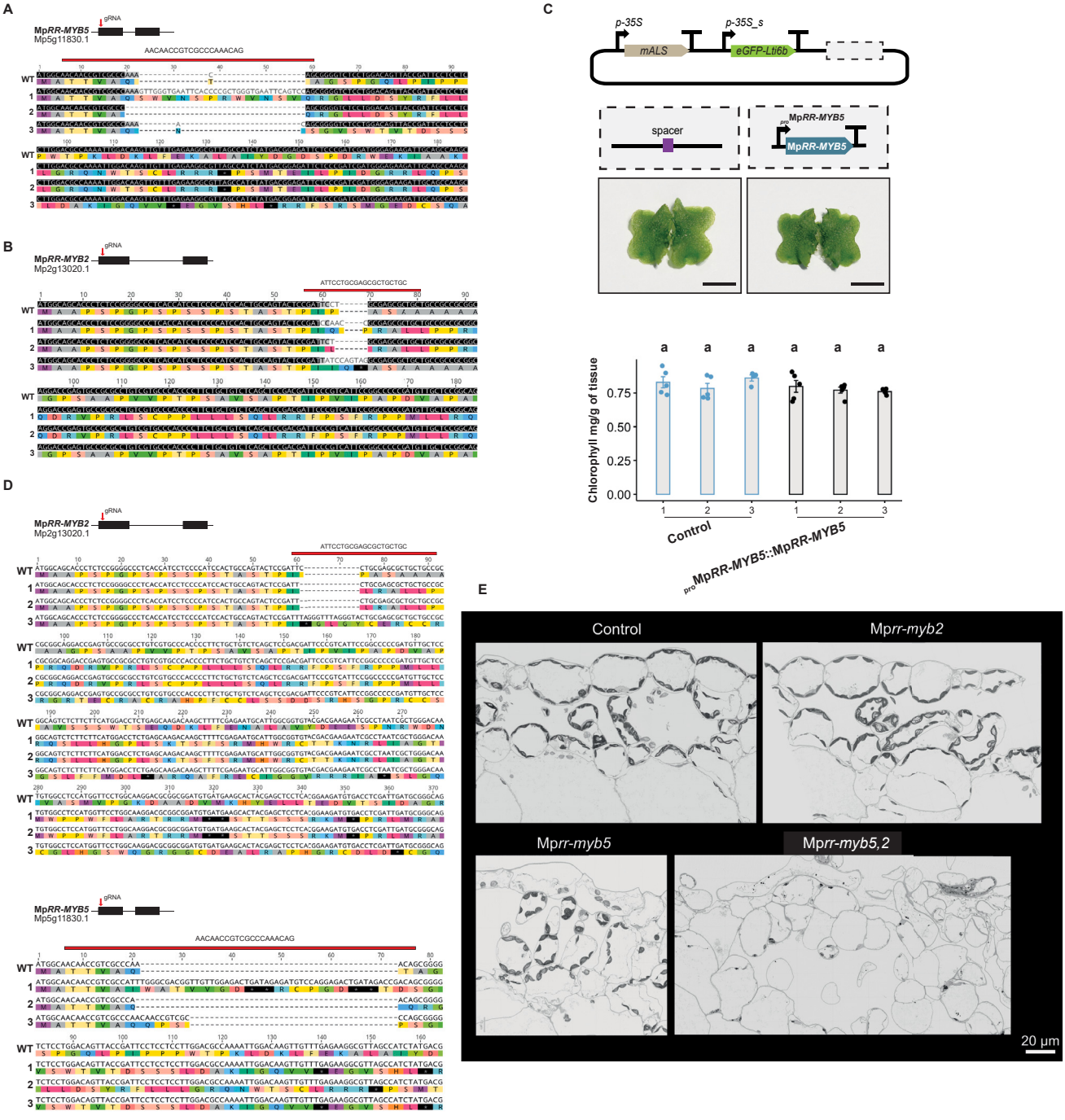


Figure S1. CRISPR-Cas9-mediated gene editing of Mp*RR-MYB5* and Mp*RR-MYB2*, related to Figure 1

(A) Schematic representation of Mp*RR-MYB5* gene with exons shown as black boxes, position of gRNA shown with arrow, and sequence analysis of Mp*rr-myb5* knockout lines. Wild-type *M. polymorpha* Cam-1 sequence shown at the top, with the 20 bp gRNA target sequence highlighted by a red line. Amino acid sequence depicted below nucleotide sequence.

(legend continued on next page)

(B) Schematic representation of Mp*RR-MYB2* gene with exons shown as black boxes, position of gRNA shown with arrow, and sequence analysis of Mp*rr-myb2* knockout lines. Wild-type *M. polymorpha* Cam-1 sequence shown at the top, with the 20 bp gRNA target sequence highlighted by a red line. Amino acid sequence depicted below nucleotide sequence.

(C) Schematic representation of control construct and that used to express Mp*RR-MYB5* from the 1,792 bp fragment upstream of the predicted translational start codon. Representative phenotypes of control and Mp*rr-myb5* mutants complemented with the $_{pro}$Mp*RR-MYB5*::Mp*RR-MYB5*. Scale bars represent 5 mm. Chlorophyll content in control, and Mp*rr-myb5* mutants complemented with the $_{pro}$Mp*RR-MYB5*::Mp*RR-MYB5*. Letters show ranking using a *post hoc* Tukey test (with different letters indicating statistically significant differences at $p < 0.01$), $n = 5$.

(D) Sequence analysis of Mp*rr-myb5,2* knockout mutant lines. Wild-type *M. polymorpha* Cam-1 sequence shown at the top, with the 20 bp gRNA target sequence highlighted by a red line. Amino acid sequence depicted below nucleotide sequence.

(E) Scanning electron micrograph maps of control, Mp*rr-myb2*, Mp*rr-myb5*, and Mp*rr-myb5,2* thallus cross sections. Scale bars represent 20 μm.

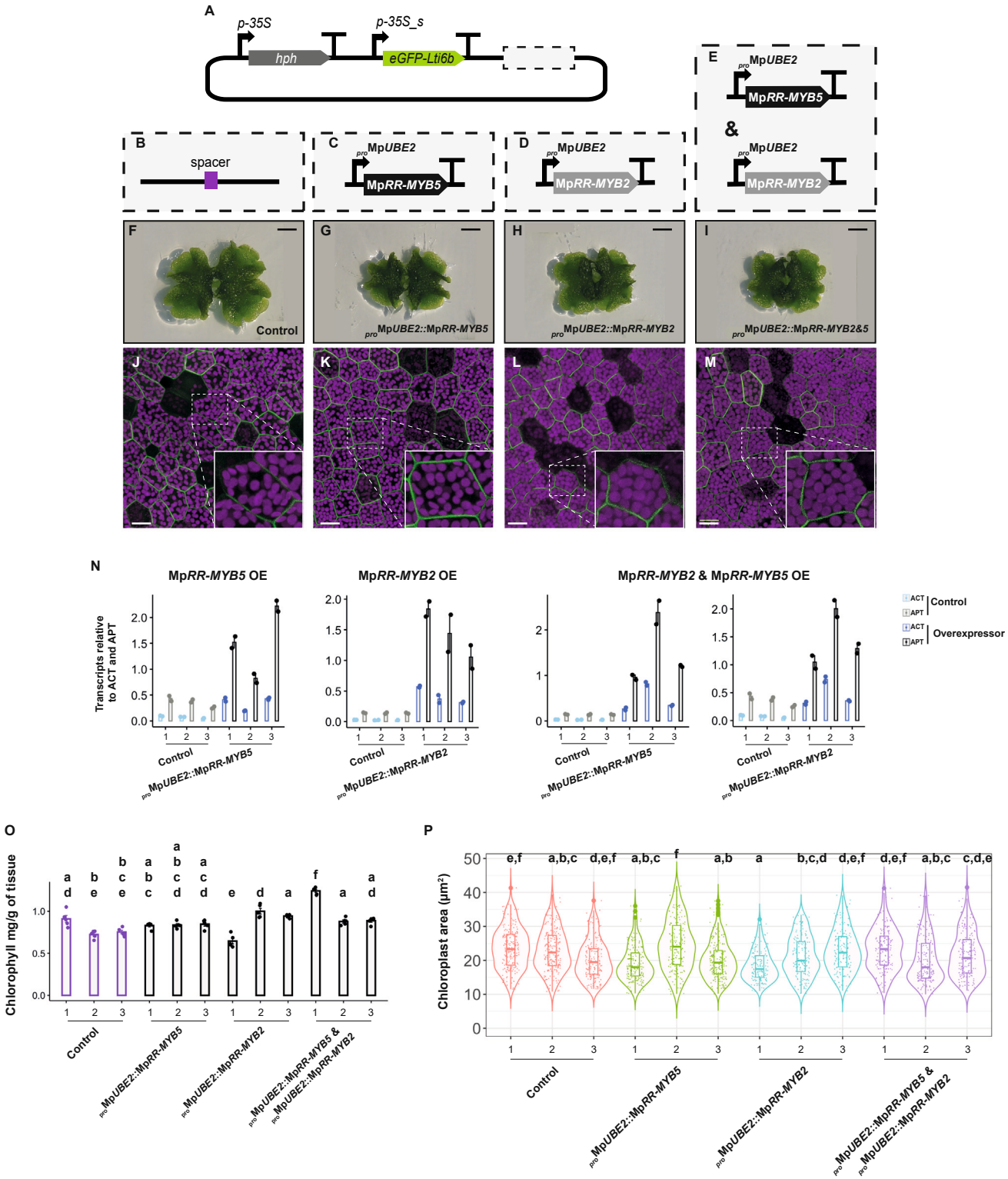


Figure S2. Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*RR-MYB2&5* over-expression analysis, related to Figure 1

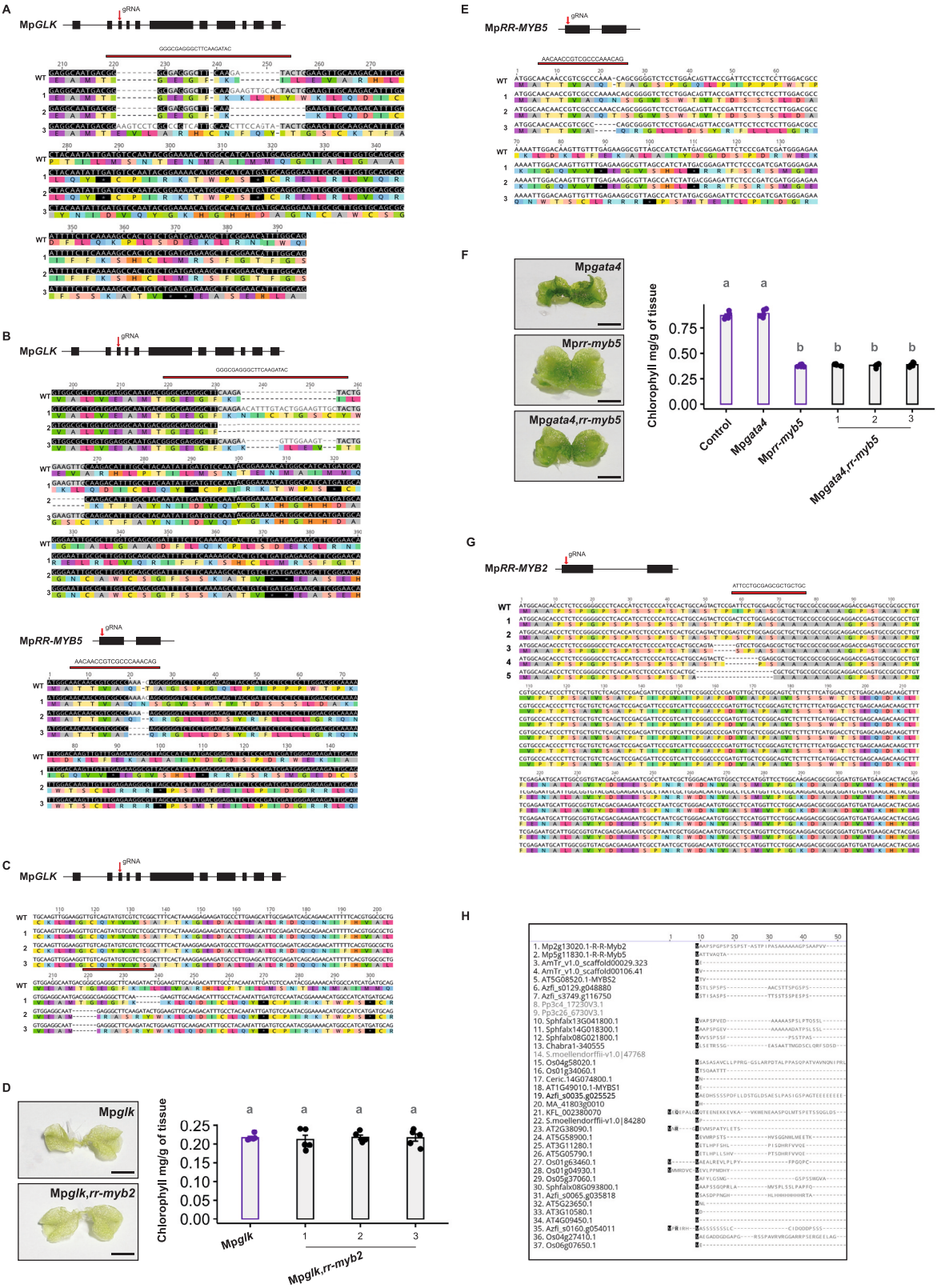
(A–E) Schematics of constructs used to overexpress Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*RR-MYB2&5*.

(F–I) Representative images of control, Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*RR-MYB2&5* over-expressors. Scale bars represent 2 mm.

(J–M) Representative images from confocal microscopy images of control, Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*RR-MYB2&5* over-expressors. Scale bars represent 25 µm.

(legend continued on next page)

(N) qPCR analysis of Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*RR-MYB2&5* over-expressors. *Adenine phosphoribosyl transferase 3* (*APT*) and *Actin 7* (*ACT*) were used as housekeeping gene controls.
(O) Chlorophyll content for Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*RR-MYB2 & 5* over-expressors. Letters show ranking using a *post hoc* Tukey test (with different letters indicating statistically significant differences at $p < 0.01$), $n = 5$.
(P) Quantification of chloroplast area of Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*RR-MYB2 & 5* over-expressors. Letters show ranking derived from a *post hoc* Tukey test (with different letters indicating statistically significant differences at $p < 0.01$).



(legend on next page)

Figure S3. CRISPR-Cas9-mediated gene editing of Mp*GLK*, Mp*GATA4*, Mp*RR-MYB5*, and Mp*RR-MYB2*, related to Figure 2

(A) Top: schematic of Mp*GLK* gene with exons indicated by black boxes, position of gRNA shown with an arrow. Bottom: sequence analysis of Mp*glk* knockout lines. Wild-type *M. polymorpha* Cam-1 sequence shown at the top, with the 20 bp gRNA target sequence highlighted by a red line. Amino acid sequence depicted below nucleotide sequence.

(B) Top: schematic representation of Mp*GLK* and Mp*RR-MYB5* genes with exons indicated by black boxes. Position of gRNA shown with an arrow. Bottom: sequence analysis of Mp*glk,rr-myb5* double-mutant lines. Wild-type *M. polymorpha* Cam-1 sequence shown at the top, with the 20 bp gRNA target sequence highlighted by a red line. Amino acid sequence depicted below nucleotide sequence.

(C) Schematic of Mp*GLK* gene as shown in (B) along with sequence analysis of Mp*glk* mutations in Mp*glk,rr-myb2* knockout mutant lines.

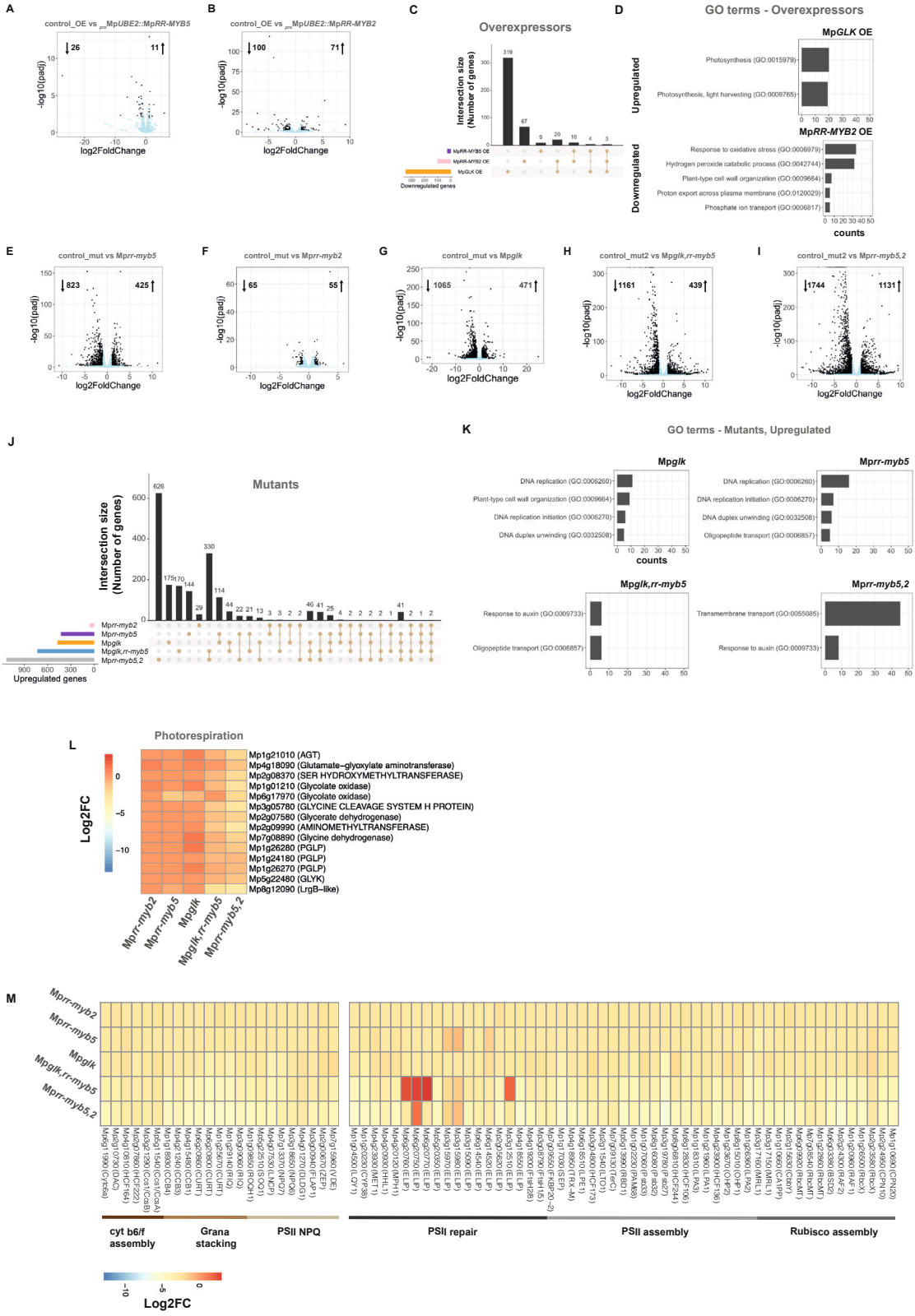
(D) Representative images of Mp*glk* and Mp*glk*,*rr-myb2* mutants (scale bars represent 5 mm), and chlorophyll content of Mp*glk* and Mp*glk*,*rr-myb2* mutants. Letters show ranking using a *post hoc* Tukey test (with different letters indicating statistically significant differences at $p < 0.01$), $n = 5$.

(E) Gene structure and sequence analysis of Mp*gata4*,*rr-myb5* knockout mutants.

(F) Representative images of Mp*gata4*, Mp*rr-myb5*, and Mp*gata4*,*rr-myb5* mutants (scale bars represent 5 mm), and chlorophyll content of Mp*gata4*, Mp*rr-myb5*, and Mp*gata4*,*rr-myb5* mutants. Letters show ranking derived from a *post hoc* Tukey test (with different letters indicating statistically significant differences at $p < 0.01$), $n = 5$.

(G) Gene structure and sequence analysis of Mp*glk,rr-myb5,2* knockout mutants. Wild-type *M. polymorpha* Cam-1 sequence shown at top. Amino acid sequence depicted below nucleotide sequence.

(H) MAFFT alignment of the N terminus of representative RR subclass RR-MYB/CCA1-like proteins. The coloring used for each column depends on the fraction of that column made up of amino acids as follows: black 100% similar, dark gray 80%–100% similar, lighter gray 60%–80% similar, white less than 60% similar.



(legend on next page)

Figure S4. *M. polymorpha* RNA-seq, related to Figures 3 and 4

(A and B) Volcano plots showing differentially expressed genes in Mp*RR-MYB5* over-expression versus controls, as well as in Mp*RR-MYB2* over-expression lines versus controls. Blue dots indicate genes with a *p*adj value $\geq$ 0.01, while black have a *p*adj $\leq$ 0.01. Total number of differentially expressed genes (DEGs) indicated at top of each panel.

(C) Upset diagrams showing shared downregulated genes in Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*GLK* over-expression lines.

(D) Enriched GO terms in Mp*GLK* and Mp*RR-MYB2* over-expressors.

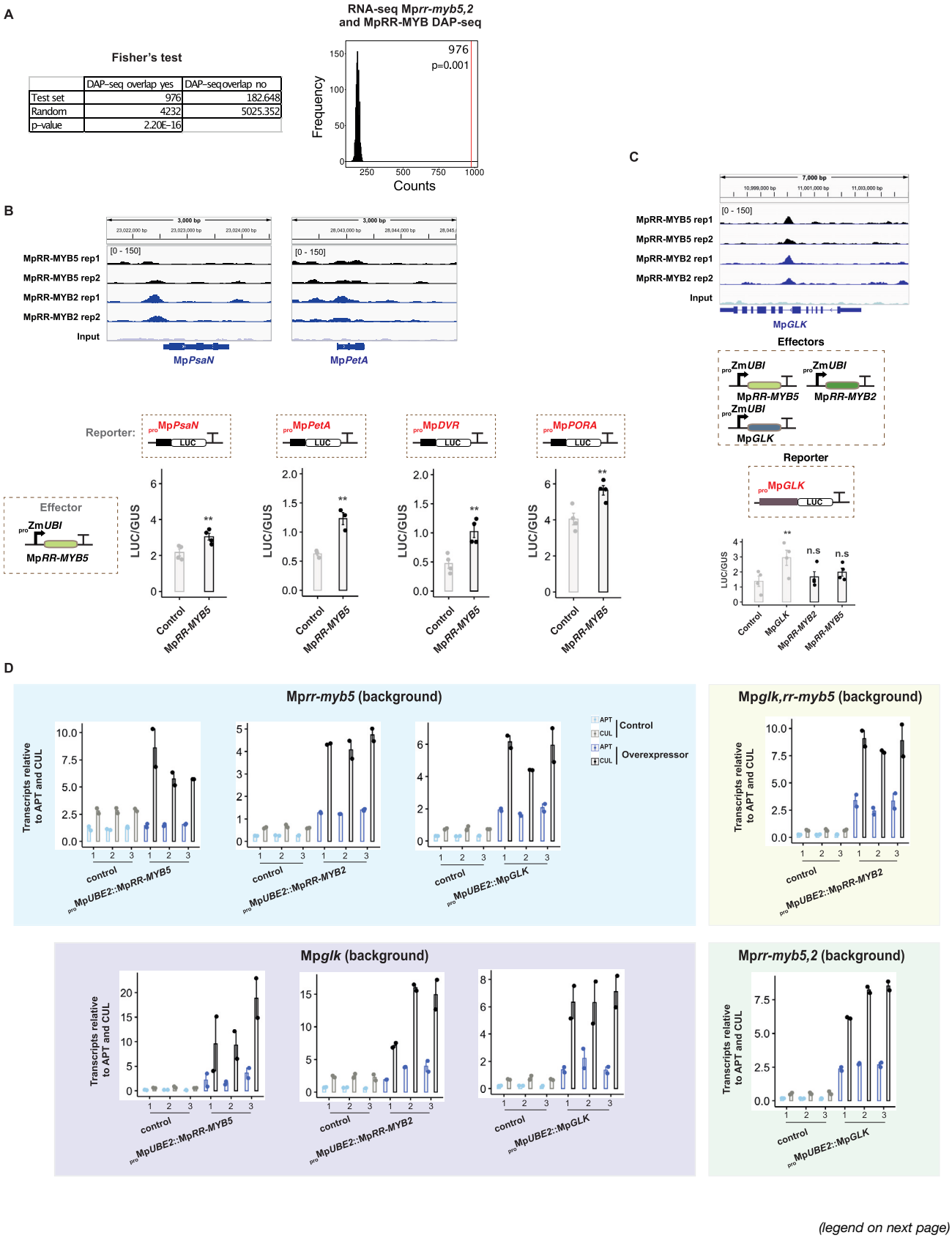
(E–I) Volcano plots showing differentially expressed genes for Mp*rr-myb5*, Mp*rr-myb2*, Mp*glk*, Mp*glk,rr-myb5*, and Mp*rr-myb5,2* mutants compared with controls.

(J) Upset diagram showing sets of upregulated genes in Mp*glk*, Mp*rr-myb2*, Mp*rr-myb5*, and the double Mp*glk,rr-myb5*, or Mp*rr-myb5,2* mutants.

(K) Enriched GO terms for upregulated genes in Mp*glk*, Mp*rr-myb5*, Mp*glk,rr-myb5*, Mp*glk*, and Mp*rr-myb5,2* mutants.

(L) Heatmap of differentially expressed photorespiration genes in Mp*rr-myb2*, Mp*rr-myb5*, Mp*glk*, as well as Mp*glk,rr-myb5* and Mp*rr-myb5*,*2* double mutants.

(M) Heatmap of differentially expressed photosynthesis-associated genes in Mp*rr-myb2*, Mp*rr-myb5*, Mp*glk*, as well as Mp*glk,rr-myb5* and Mp*rr-myb5*,*2* double mutants.



(legend on next page)

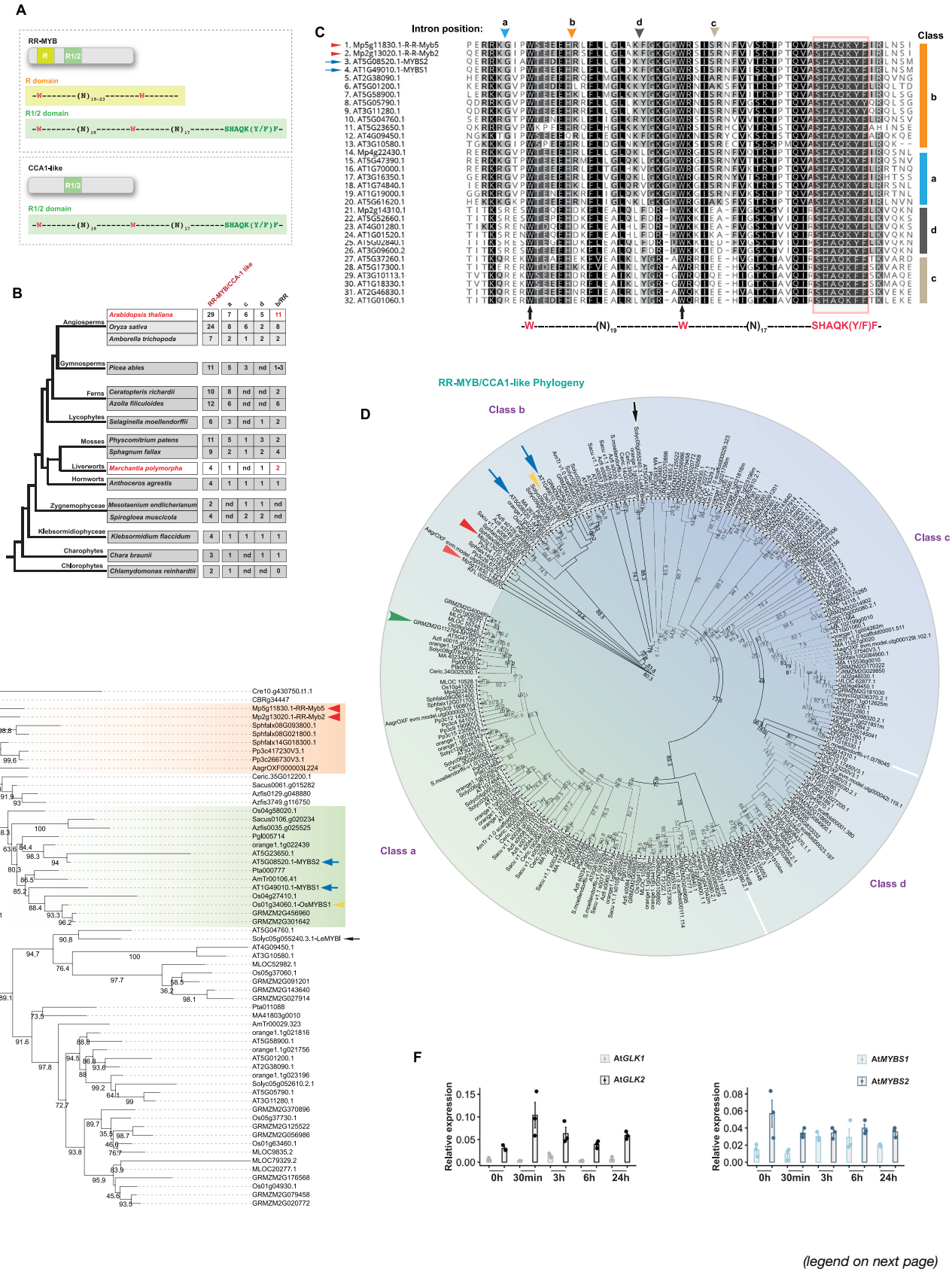
Figure S5. *M. polymorpha* DAP-seq, effector assays, and quantitative polymerase chain reactions for confirmation of transgene over-expression, related to Figures 4 and 5

(A) Fisher exact test and permutation testing for misregulated genes in Mp*rr-myb5,2* mutants that also contained a Mp*RR-MYB2* or Mp*RR-MYB5* binding site. This was compared with a 976 gene set obtained by randomly sampling genes from the RNA-seq dataset 1,000 times.

(B) MpRR-MYB5 and MpRR-MYB2 binding sites in photosynthesis genes identified by DAP-seq. Schematic of gene body shown at bottom of each panel. Schematics of effector and reporter constructs used in the luciferase (LUC) reporter system. Bar graphs show that Mp*RR-MYB5* can induce activation of $_{pro}$Mp*PsaN*::LUC, $_{pro}$Mp*PetA*::LUC, $_{pro}$Mp*DVR*::LUC, and $_{pro}$Mp*PORA*::LUC. Data represent means and standard error of the mean (n = 4), and asterisks indicate a significant difference compared with control (*p < 0.05, **p < 0.01, n.s., non-significant, paired t test).

(C) MpRR-MYB5 and MpRR-MYB2 binding sites identified by DAP-seq in the Mp*GLK* promoter. Schematic structures of effector and reporter constructs used in the luciferase (LUC) reporter system. Bar graphs show Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*GLK* induction or lack of activation of $_{pro}$Mp*GLK*::LUC. Data represent means and standard error of the mean (n = 4), and asterisks indicate a significant difference compared with control (*p < 0.05, **p < 0.01, n.s., non-significant, paired t test).

(D) qPCR analysis of Mp*rr-myb5*, Mp*glk*, Mp*glk,rr-myb5*, and Mp*rr-myb5,2* mutants complemented with Mp*RR-MYB5*, Mp*RR-MYB2*, and Mp*GLK* under control of the Mp*UBE2* promoter. *ADENINE PHOSPHORIBOSYL TRANSFERASE 3* (*APT*) and *CULLIN 1* (*CUL*) were used as housekeeping gene controls (Saint-Marcoux et al.[98]).



(legend on next page)

Figure S6. The RR-MYB/CCA1-like related transcription factor family, related to Figure 6

(A) Schematic representation of RR-MYB/CCA1-like protein structure. Mp*RR-MYB5* belongs to the RR-MYB/CIRCADIAN CLOCK ASSOCIATED1-like (CCA1-like) subfamily of the MYB-related transcription factors. The RR-MYB/CCA1-like subfamily contains two different classes of genes, the RR and the CCA1-like. While both contain a Myb domain with the characteristic R1/2 motif containing SHAQK(Y/F)F, the RR class has an additional MYB-like domain called R. Conserved "W" residues and SHAQK(Y/F)F domain indicated at bottom of the alignment. The CCA1-like class can further be categorized into three additional groups based on intron position within the R1/2 domain, called a, c, and d Du et al.[27] Intron position within the R1/2 domain of the RR class is also distinct from the three intron positions found in the CCA1-like class (b group) and can serve as an additional character for RR ortholog identification. *A. thaliana* has a total of 29 RR-MYB/CCA1-like genes, of which seven belong to the a class, six to the c class, five to the d class, and 11 to the b class/RR. *M. polymorpha* has only four RR-MYB/CCA1-like genes, of which one belongs to the a class, one to the d class, and two to the b class/RR. The c class seems to be absent.

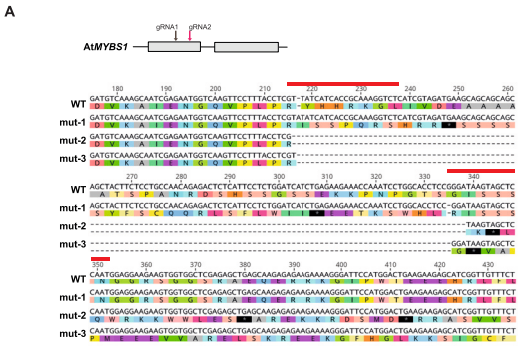
(B) Plant phylogeny with representative species for which whole genome assemblies are available. RR-MYB/CCA1-like gene numbers are shown (nd, not detected).

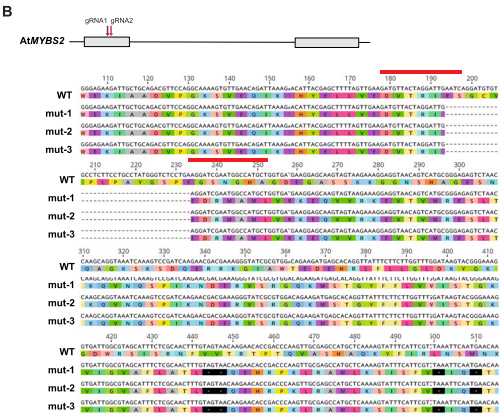
(C) MAFFT alignment of R1/2 domain of *M. polymorpha* and *A. thaliana* RR-MYB/CCA1-like transcription factor. *M. polymorpha* and *A. thaliana* proteins examined in this study indicated with red arrowheads and blue arrows, respectively. Positions of introns marked above alignment with arrowheads. The coloring used for each column depends on the fraction of that column made up of amino acids as follows: black 100% similar, dark gray 80%–100% similar, lighter gray 60%–80% similar, white less than 60% similar.

(D) RR-MYB/CCA1-like related phylogeny. Numbers on branches represent SH-aLRT test support Guindon et al.[67] *M. polymorpha*, *A. thaliana*, rice, and tomato proteins indicated with red, blue, yellow, and black arrows, respectively.

(E) RR-MYB-related phylogeny. Numbers on branches represent ultrafast bootstrap support Hoang et al.[94] *M. polymorpha*, *A. thaliana*, rice, and tomato proteins indicated with red, blue, yellow, and black arrows, respectively.

(F) Transcript abundance of At*GLK1*, At*GLK2*, At*MYBS1*, and At*MYBS2* genes during de-etiolation. Elongation factor 1A (*EF1A*) was used as internal control. Data presented as means and standard error of the mean, $n = 3$.

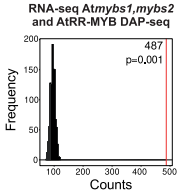




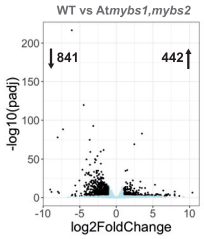
C

Fisher's test

	DAP-seq overlap yes	DAP-seq overlap no
Test set	487	1983
Random	95.593	2374.407
p-value	2.20E-16	



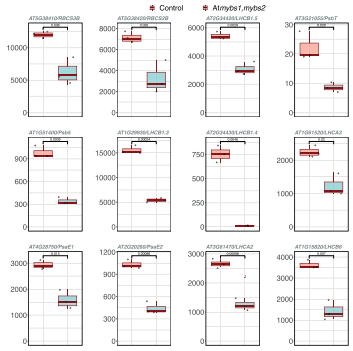
D



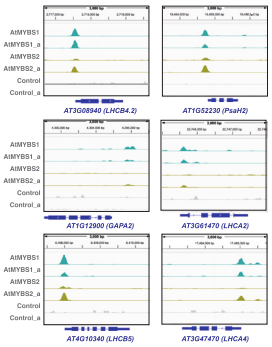
E

GO description	Counts	FDR
Photosynthesis	18	1.05E-11
Endoplasmic reticulum unfolded protein response	4	4.22E-02
Response to light stimulus	16	3.39E-06
Generation of precursor metabolites and energy	23	4.53E-06
Response to organonitrogen compound	8	4.26E-02

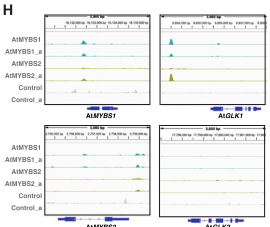
F



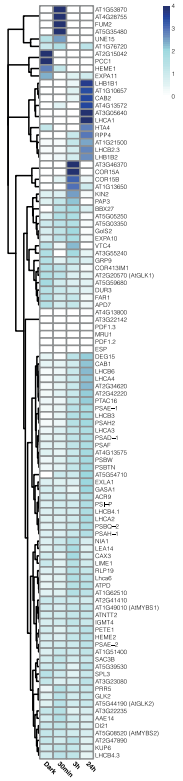
G



H



K



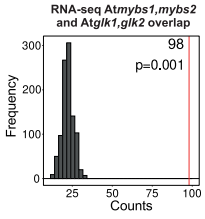
I

Go description	Counts	FDR
Photosynthesis	21	1.20E-31
Response to light stimulus	20	6.45E-22
Generation of precursor metabolites and energy	23	4.78E-21
Ribosomal large subunit biogenesis	7	2.89E-03

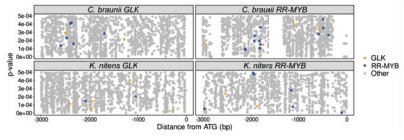
J

Fisher's test

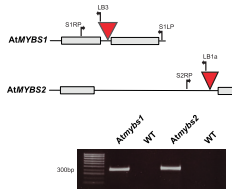
	RR-MYB yes	RR-MYBS no
Test set	98	332
Random combined	22.266	407.734
p-value	2.511E-14	



L



M



(legend on next page)

Figure S7. CRISPR-Cas9-mediated gene editing for At*MYBS1* and At*MYBS2* as well as DAP-seq and effector assays, related to Figure 7

(A and B) Top: schematic of the At*MYBS1* and At*MYBS2* genes with exons shown as gray boxes. Positions of gRNAs shown with a gray/red arrow. Bottom: sequence analysis of At*mybs1* and At*mybs2* knockout double mutants. Wild-type *A. thaliana Col-0* sequence shown at the top, with the 20 bp gRNA target sequence highlighted by a red line. Amino acid sequence depicted below nucleotide sequence.

(C) Fisher exact test and permutation testing for overlap between genes that contained an At*MYBS1* or At*MYBS2* binding site and that were misregulated in At*mybs1,myb2* mutants. A set of 487 genes were randomly sampled 1,000 times from the RNA-seq dataset.

(D) Volcano plot showing differentially expressed genes in At*mybs1,mybs2* mutants. Blue dots indicate genes with a *p*adj $\geq$ 0.01, while black ones have a *p*adj $\leq$ 0.01. Total number of DEGs indicated at top of each panel.

(E) Top five GO terms in At*mybs1,mybs2* mutants estimated using PANTHER (GO-slim biological processes).

(F) Boxplots for representative downregulated photosynthesis-associated genes in At*mybs1,mybs2* mutants. Data presented as transcripts per million (TPM) values. *p* values of two-tailed t test shown.

(G and H) AtMYBS1 and AtMYBS2 binding sites in photosynthesis or At*MYBS1&S2* genes as identified by DAP-seq O'Malley et al.[35] Gene body shown at bottom of each panel.

(I) Top GO terms in At*glk1,glk2* mutants estimated using PANTHER (GO-slim biological processes).

(J) Fisher exact test and permutation testing for overlap between downregulated genes in Atglk*1,glk2* and At*mybs1,myb2* mutants, compared with 1,000 randomly sampled sets of genes (size = 98) from the RNA-seq dataset.

(K) Heatmap (*Z* score) showing temporal response of genes found to be regulated by At*GLK1*, At*GLK2*, At*MYBS1*, and At*MYBS2* gene after exposure to light.[20]

(L) Scatterplots showing position and predicted binding affinity of GLK, RR-MYB binding, and other motifs 3,000 bp upstream of the translation start site of *Chara braunii* and *Klebsormidium flaccidum* GLK and RR-MYB putative homologs. y axis shows *p* values of matches between upstream regions and motif position weight matrices, and x axis shows position of the motif center relative to the transcriptional start site.

(M) Genotyping of At*mybs1* (SAIL_1184_D04) and At*mybs2* (SALK_150774) mutant lines. Top: schematic of At*MYBS1* and At*MYBS2* genes with exons shown as gray rectangles. The transfer DNA (T-DNA) insertion positions are indicated with red arrowheads (not in scale). Positions of primer sequences used for genotyping shown—primer pairs S1RP & LB3 and S2RP & LB1a were used to confirm presence of the T-DNA insert in At*mybs1* and At*mybs2* mutants, respectively. Bottom: PCR analysis for confirmation of T-DNA insert in mutants.