

MYB-related transcription factors control chloroplast biogenesis

Eftychios Frangedakis^{1,4,*}, Nataliya E. Yelina^{1,2,4}, Kumari Billakurthi¹, Lei Hua¹, Tina Schreier^{1,3}, Patrick J. Dickinson¹, Marta Tomaselli¹, Jim Haseloff¹, Julian M. Hibberd^{1,5,*}

¹Department of Plant Sciences, University of Cambridge, Cambridge, CB2 3EA, UK

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

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

⁴These authors contributed equally

⁵Lead contact

* Corresponding authors: jmh65@cam.ac.uk (J.M.H.), ef391@cam.ac.uk (E.F.)

Summary

Chloroplast biogenesis is dependent on master regulators from the GOLDEN2-LIKE (GLK) family of transcription factors, but *glk* mutants contain residual chlorophyll and therefore 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 whilst those allowing CO₂ 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 *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 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 one 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*¹⁰, rice¹⁴ 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 rationalised 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*, *Mpglk* knockouts contain low but detectable levels of chlorophyll.¹⁶ We therefore hypothesised that an unknown transcription factor conserved between *M. polymorpha* and *A. thaliana* acts alongside *GLK* to control chloroplast biogenesis during photomorphogenesis.

After re-examination of publicly available RNA sequencing data, gene editing of transcription factors and phenotypic analysis we identify two RR-MYB 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 Photosystem I and II, the RR-MYBs have a broader set of targets that extends to genes allowing CO₂ 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 sulphur assimilation, the biosynthesis of amino acids, fatty acids and carotenoids.

Results

MpRR-MYB5 regulates chloroplast development synergistically with its paralog MpRR-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* (**Figure 1A and 1B**).²⁰ This identified 108 and 144 transcription factors upregulated after exposure to light in *M. polymorpha* and *A. thaliana* respectively (**Table 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 (**Figure 1A and 1B**). Fourteen candidates from *M. polymorpha* were identified (**Figure 1B and Table S2**). Two of these (MpGLK and MpGATA4) are homologs of known photosynthesis regulators in *A. thaliana*^{10,21} and MpGLK 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 HY5²², a homeobox-leucine zipper (HD-ZIP) protein whose *ATHB17* ortholog in *A. thaliana* 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 MpGLK, which has previously been reported to lead to a pale phenotype when mutated¹⁶ only one other candidate had low chlorophyll. This was Mp5g11830, annotated as MpRR-MYB5 in the *M. polymorpha* genome database, but previously also referred to as a CIRCADIAN CLOCK ASSOCIATED1-like RR-MYB-Related transcription factor.^{26,27} MpRR-MYB5 has a single paralog (MpRR-MYB2 - **Figure 1B and C**) showing high similarity to MpRR-MYB5 at the amino acid level with for example the CCA1-like/RR-Myb domain being 93% identical (**Figure 1C**). Mutant alleles of MpRR-MYB5 but not MpRR-MYB2 appeared

pale (**Figure 1D, E, F, Figure S1A-B**) and analysis of chlorophyll content confirmed this (**Figure 1H**). All lines in which insertions or deletions introduced premature stop codons in *MpRR-MYB5* (**Figure S1A**) had 40-50% less chlorophyll than controls (**Figure 1E and H**). The *Mprr-myb5* mutant was complemented when *MpRR-MYB5* was expressed from its own promoter (**Figure S1C**) confirming the pale phenotype was unlikely associated with off-target Cas9 editing. Mutating *MpRR-MYB5* and *MpRR-MYB2* simultaneously (**Figure S1D**) led to extremely pale plants with chlorophyll content reduced by ~95% compared with controls (**Figure 1G and H**). Analysis of a timecourse associated with the first hours after exposure to light¹⁹ indicated that transcripts derived from *MpRR-MYB5*, *MpRR-MYB2* and *MpGLK* were induced by 24 hours after light was perceived but that *MpRR-MYB5* and *MpRR-MYB2* expression was maintained for longer (**Figure 1I**). To test whether the photosynthetic apparatus was functional in the single *Mprr-myb5* and double *Mprr-myb5,2* mutants we applied the inhibitor Di-Chlorophenyl Di-Methyl Urea (DCMU) that blocks photosynthetic electron transport²⁸ 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 centres. There was a small but significant reduction in F_v/F_m in the double *Mprr-myb5,2* mutant (**Figure 1J**) indicating compromised function in this genotype. Consistent with low chlorophyll levels in *Mprr-myb5* and *Mprr-myb5,2* mutants, chloroplasts were smaller and thylakoids underdeveloped compared with controls and *Mprr-myb2* (**Figure 1K-S and Figure S1E**). Poorly developed chloroplasts from *Mprr-myb5,2* mutants contained more starch (**Figure 1T**).

To determine whether *MpRR-MYB5* and *MpRR-MYB2* limit greening, we generated overexpression lines driven by the strong *MpUBE2* constitutive promoter²⁹ and used eGFP to mark the plasma membrane (**Figure S2A-E**). Although quantitative polymerase chain reactions confirmed each transgene was over-expressed (**Figure S2F-N**) plants appeared similar to controls and there were no evident perturbations to chlorophyll content, chloroplast size or morphology (**Figure S2O-P**). We conclude that *MpRR-MYB5* and *MpRR-MYB2* act redundantly and are necessary for chloroplast biogenesis but in contrast with *MpGLK*¹⁶ they are not sufficient to activate this process. Moreover, in the absence of both *MpRR-MYB5* and *MpRR-MYB2* assembly of the photosynthetic apparatus was very limited.

***MpRR-MYB5* and *MpRR-MYB2* act with *MpGLK* to control chloroplast biogenesis**

As double *Mprr-myb5,2* mutants showed residual chloroplast development, we hypothesised that their limited photoautotrophic growth was associated with activity of the previously

characterised master regulator *GLK*. To test this, we attempted to generate higher order mutants which combined mutant alleles of *Mpglk*, *Mprrr-myb5* and *Mprrr-myb2*. We were able to knock-out *MpRR-MYB2* or *MpRR-MYB5* in the presence of *Mpglk* mutant alleles (**Figure S3A-C**) and these double mutants were similar or paler than the *Mpglk* mutant (**Figure 2A-E and Figure S3D**). Although previous analysis indicated that GATAs are not required for chloroplast biogenesis in *M. polymorpha*,¹⁶ we produced double *Mpgata4,rr-myb5* mutants. Chlorophyll content in these lines was similar to that of *Mprrr-myb5*, and they showed similar perturbations to thallus morphology evident in *Mpgata4* mutants (**Figure S3E-F**). This implies that *MpGATA4* is unlikely control chloroplast development with *Mprrr-MYB5*. Application of DCMU confirmed that Photosystem II was functional in *Mpglk,rr-myb5* double mutants (**Figure 2F**). Chloroplasts were smaller and contained fewer thylakoid membranes with reduced granal stacking in *Mpglk,rr-myb5* compared with each single mutant (**Figure 2G-O**). Thus, in the absence of both *MpRR-MYB5* and *MpGLK* very limited assembly of the photosynthetic apparatus took place.

We were unable to generate triple *Mpglk,rr-myb5,2* mutants implying that this allelic combination could be lethal. For example, after super-transforming *Mpglk,rr-myb5* mutants with a vector allowing expression of the same guide RNA used to generate the *Mprrr-myb2* mutants reported above, 91 lines were obtained. However, none were paler than *Mpglk,rr-myb5* mutants, and when genotyped 86 lines had no edits in *MpRR-MYB2*. Of the five lines that were edited in *MpRR-MYB2* (as well as *MpGLK* and *MpRR-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 (**Figure S3G and H**). In contrast, when the original *Mprrr-myb2* mutants were identified (**Figure 1F**) more than 60% of plants contained mutations that introduced early stop codons or disturbed reading frame (**Dataset 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 *MpRR-MYB5* and *MpRR-MYB2* we performed RNA sequencing of overexpressing lines, single and double mutants. Overexpression of *MpRR-MYB2* and *MpRR-MYB5* led to upregulation of 71 and 11 genes respectively (*padj*-value ≤ 0.01 , $\text{Log}_2\text{FC} \geq 1$ -fold) (**Figure S4A and B, Dataset S1**) and there was limited overlap between these two datasets (**Figure 3A and Figure S4C and D**). Overexpression of *MpGLK* that led to the upregulation of 493 genes (**Figure 3A and**¹⁶).

168 In loss of function mutants for MpRR-MYB2 or MpRR-MYB5, 65 and 823 genes showed
169 reductions in transcript abundance respectively compared with controls (*p*-adj-value ≤ 0.01 ,
170 Log2FC ≥ 1 -fold) (**Figure 3B and Figure S4E and F, Dataset S1**). Knocking out MpGLK had
171 greater impact with 1065 genes being downregulated (**Figure 3B and Figure S4G**). In double
172 *Mpglk,rr-myb5* mutants (**Figure S4H**) 1161 genes had lower transcript abundance than controls,
173 and in the double *Mprrr-myb5,2* mutants this was increased to 1744 (**Figure 3B and Figure S4I**).
174 The largest overlap between genotypes (524 genes) was detected for *Mpglk,rr-myb5* and *Mprrr-*
175 *myb5,2* mutants (**Figure 3B and Figure S4J**), further supporting synergistic action of MpRR-
176 MYBs and MpGLK. Gene Ontology (GO) terms were used to provide insight into classes of genes
177 impacted by overexpression or loss of function of MpRR-MYBs and MpGLK. Consistent with the
178 lack of detectable phenotype after overexpression of MpRR-MYB5 or loss of MpRR-MYB2
179 function, no distinct GO terms were impacted in these lines. When genes downregulated in *Mprrr-*
180 *myb5* and *Mpglk* mutants were assessed, the response to oxidative stress and hydrogen
181 peroxidase catabolism terms were over-represented (**Figure 3C, Dataset S2**). *Mprrr-myb5*
182 mutants also showed changes to protein phosphorylation (**Figure 3C, Dataset S2**) whilst in *Mpglk*
183 photosynthesis, light harvesting and chlorophyll biosynthesis terms were affected (**Figure 3C**). It
184 was notable that for genes downregulated in the *Mpglk*, *Mpglk,rr-myb5* and *Mprrr-myb5,2*, mutants
185 similar terms were over-represented (**Figure 3C, Dataset S2**). For genes upregulated in *Mpglk*
186 mutants DNA replication related GO terms were over-represented, whereas response to auxin
187 and transmembrane transport were affected in *Mprrr-myb5,2* and *Mpglk,rr-myb5* mutants
188 respectively (**Figure S4K, Dataset S2**). Transcripts associated with proteins operating in similar
189 cellular locations were impacted in *Mpglk*, *Mpglk,rr-myb5* and *Mprrr-myb5,2* mutants, including
190 Photosystem I, thylakoid membranes, and also the ribosome (**Dataset S2**). In the case of *Mprrr-*
191 *myb5,2* however, a more extensive effect on chloroplast and photosystem-related GO terms was
192 evident. For upregulated genes, no GO terms were impacted in the *Mprrr-myb5,2* and *Mpglk,rr-*
193 *myb5* mutants, while for *Mpglk* the nucleosome GO term was most impacted (**Dataset S2**). Loss
194 of function of either the MpRR-MYBs or MpGLK primarily therefore caused changes in
195 overlapping GO terms, exemplified by those associated with photosynthesis, responses to
196 oxidative stress, and hydrogen peroxide catabolic processes (**Dataset S2**). However, some terms
197 were specific to each mutant, with for example iron ion transmembrane transport, translation, and
198 translational elongation being enriched in *Mprrr-myb5,2*, whilst the L-phenylalanine catabolic
199 process term was specific to *Mpglk* (**Dataset S2**). Overall, loss of function *Mpglk*, *Mpglk,rr-myb5*
200 and *Mprrr-myb5,2* mutants caused changes in GO terms primarily associated with photosynthesis

(**Dataset S2**). For the common impacted GO categories, ~40-60% of genes were co-regulated by both *MpRR-MYBs* and *MpGLK* (**Dataset S2**).

Since chlorophyll content was reduced in *Mprrr-myb5* and *Mprrr-myb5,2* mutants, we examined transcript abundance of the nineteen annotated chlorophyll biosynthesis genes (**Figure 3D**). With the exception of *HEMA* in *Mprrr-myb5* mutants, knocking out either *MpRR-MYB5* or *MpRR-MYB2* did not significantly affect transcript abundance. In contrast, in *Mpglk* mutants transcript abundance from seventeen of the genes was reduced, and in *Mprrr-myb5,2* all nineteen genes were significantly downregulated (**Figure 3D**). The effect on transcript abundance was more pronounced in *Mprrr-my5,2* than in *Mpglk,rr-myb2* (**Figure 3D**). We next examined the impact of loss of the *MpRR-MYBs* on approximately 200 other genes annotated as photosynthesis-related (**Dataset S1**). This included genes associated with CO₂ fixation, the light harvesting apparatus, and their assembly and repair. In the *Mprrr-myb5* and *Mprrr-myb2* mutant alleles there was limited effect on photosynthesis associated genes (**Figure 3E and Figure 4A-B**). For example, in *Mprrr-myb2* expression of only *petE* (Mp4g02720) was significantly perturbed (Log2FC ≥ 1-fold) (**Dataset S1**). In *Mprrr-myb5* mutants, a small number of genes were impacted including those encoding a small subunit of RuBisCO (Mp4g09890) and a CHLOROPHYLL A/B BINDING PROTEIN (Mp7g05530) (**Figure 3E and Dataset S1**). As expected, changes to photosynthesis transcripts were more evident in *Mpglk* (**Figure 3E and Figure 4A-B**), and even more severe when both *MpRR-MYB5* and *MpGLK* were absent (**Figure 3E and Figure 4A-B**). Strikingly, when *MpRR-MYB2* and *MpRR-MYB5* were simultaneously knocked out, the effect on photosynthesis associated genes was extensive and more widespread than in *Mpglk,rr-myb5*. For example, in *Mprrr-myb5,2* double mutants, the majority of genes encoding enzymes involved in the Calvin Benson Bassham cycle and photorespiration were downregulated (**Figure 3E and Figure S4L**). Moreover, genes encoding components of both photosystems and their respective light harvesting complexes as well as the Cytochrome *b₆f* complex were downregulated (**Figure 4A-B**). 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 *Mprrr-myb5,2* (**Figure S4M**). This contrasts with the *Mpglk* single mutant, where the greatest changes were associated with genes encoding components of the photosystems and their light harvesting complexes (**Figure 4A-B**). For genes eg associated with RuBisCO assembly, granal stacking, or repair of PSII that were less expressed in *Mpglk* the level of perturbation was much lower than in *Mprrr-myb5,2* (**Figure S4M, Dataset S1**). Immunoblotting demonstrated that LHCA, PsbA, PsbC and PsbS proteins were less abundant in *Mpglk* and *Mprrr-myb5* mutants and in the double *Mprrr-myb5,2* this effect was enhanced (**Figure 4C**).

MpRR-MYBs can activate photosynthesis genes

The analysis above indicated that loss of *MpRR-MYBs* had widespread impact on photosynthesis gene expression that overlapped with that seen in *Mpglk*. 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)³⁰ to identify genome-wide binding sites of *MpRR-MYB2* and *MpRR-MYB5* and DNA sequence motifs that they bind. This identified a total of 6804 and 2839 binding sites for *MpRR-MYB2* and *MpRR-MYB5* respectively (**Dataset 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 mis-regulated in the *Mprr-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, **Dataset S3**) including *MpPsbQ*, *MpPsaD* and *MpLHCB1* (**Figure 4G**). Moreover, when *MpRR-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 *MpPsaN* and *MpPetA* (**Figure S5B**). *MpRR-MYB2* activated *MpLHCB1* (**Figure 4G**). *MpRR-MYB5* also activated expression from *MpDVR* and *MpPORA* that encode enzymes allowing chlorophyll biosynthesis (**Figure S5B**). Trans-activation assays showed that *MpGLK* was sufficient to activate both *MpRR-MYB5* and *MpRR-MYB2* (**Figure 4H**). However, there was no strong evidence that *MpRR-MYB2* bound its own promoter or activated it (**Figure 4H**), and whilst 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-MYB2* to control their expression to allow co-ordination with *MpGLK*. In support of this, analysis of the RNA-seq data indicated that *MpRR-MYBs* and *MpGLK* may be linked in a gene regulatory network. For example, in *Mpglk*, transcripts derived from *MpRR-MYB5* and *MpRR-MYB2* were upregulated (**Figure 4I**) and *MpGLK* transcripts were downregulated in *Mprr-*

myb5,2 (**Figure 4I**). We note a MpRR-MYB binding site in the sixth intron of MpGLK (**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 *Mprr-myb5*, *Mpglk*, *Mpglk,rr-myb5* or *Mprr-myb5,2* mutants (**Figure 5A-G**). Quantitative polymerase chain reactions confirmed that each transgene was over-expressed (**Figure S5D**). MpRR-MYB5 and MpRR-MYB2 complemented *Mprr-myb5* mutants (**Figure 5C and D**) further arguing for functional redundancy. However, neither MpRR-MYB rescued the single *Mpglk* (**Figure 5D and E**) or double *Mpglk,rr-myb5* mutants (**Figure D and F**). MpGLK over-expression in *Mprr-myb5* increased chlorophyll content but not to the same extent as seen in *Mpglk* (**Figure 5D and E**). Finally, when MpGLK was overexpressed in *Mprr-myb5,2*, chlorophyll levels were increased by ~10% but the absolute levels were still 90% lower than wild type (**Figure 5G and H**). In summary, these data indicate that MpRR-MYB5 and MpRR-MYB2 cannot complement *Mpglk*, and MpGLK cannot complement the double *Mprr-myb5,2* mutant.

RR-MYBs control chloroplast biogenesis in *Arabidopsis thaliana*

The RR-MYB/CCA1-like subfamily of MYB-related transcription factors^{27,31,32} containing MpRR-MYB5 and MpRR-MYB2 are characterised by a conserved SHAQK(Y/F)F motif (**Figure S6A**). We identified eleven members of this group in *A. thaliana* (**Figure 6A, Figure S6B-E**) of which AtMYBS1, AtMYBS2 and AT5G23650 were the closest homologs of MpRR-MYB5 and MpRR-MYB2 (**Figure S6B-E**). Re-analysis of publicly available data indicated that AT5G23650 is not expressed in photosynthetic tissues and so we focused analysis on AtMYBS1 and AtMYBS2. Due to the functional redundancy evident for MpRR-MYB5 and MpRR-MYB2 above, double *Atmybs1,mybs2* mutants were identified after CRISPR/Cas9-mediated gene editing (**Figure 6B, Figure S7A and B**) and analysed in parallel with previously generated single *Atmybs1* and *Atmybs2* mutants. There were no detectable changes to rosette phenotype in the single mutants but *Atmybs1,mybs2* mutants were pale (**Figure 6C-F**), and this was most noticeable after bolting (**Figure 6G and H**). There were no detectable changes in chloroplast size or number in mesophyll cells of single mutants, but chloroplasts were smaller in *Atmybs1,2* (**Figure 6C-F and I**). Chloroplasts of *Atmybs1,2* 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 *Atmybs1,mybs2* mutant (**Figure 6K**). Unfolding of the apical hook and greening of cotyledons was evident in both *Atglk1,2* and *Atmybs1,2* double mutants but greening appeared slower than in controls (**Figure 6L**). For the first twelve hours

after light was perceived, chlorophyll content was indistinguishable in *Atglk1,2* and *Atmybs1,2* double mutants, but by 24 hours *Atglk1,2* had less chlorophyll than *Atmybs1,2* (**Figure 6M**). F_v/F_m was reduced for the first six hours of light in both mutant backgrounds but this was more apparent in *Atmybs1,2* (**Figure 6N**). *AtGLK1* showed an increase in transcript abundance by 30 minutes of light that was maintained over 24 hours (**Figure 6O**). In contrast, transcripts derived from *AtMYBS2* had a higher abundance before exposure to light (**Figure 6O, Figure S6F**).

Consensus binding sites for *AtMYBS1&2* have previously been defined from DAP-seq³³ and along with those we defined for *M. polymorpha*, also comprise a TTATC core (**Figure 7A**). 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 *Atmybs1,s2* (**Dataset S1**) identified 2470 transcripts that were less abundant compared with controls (**Figure 7C**, *padj*-value ≤ 0.01 , LFC ≥ 0.5) and of these 122 can be bound by *AtMYBS1* or *AtMYBS2* (**Figure 7C, Dataset S1**;³³). The Fisher's exact test and permutation analysis indicated that this is greater overlap than expected by chance (**Figure S7C**). Genes downregulated in the *Atmybs1,s2* mutant compared with wild-type (**Figure S7D-E**) 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 (**Figure 7D and Figure 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 data³³ 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, Dataset S3**). We also tested whether *AtMYBS1* or *AtMYBS2* can activate themselves as well as *AtGLK1* and *AtGLK2*. Whilst the magnitude of response varied, all four transcription factors could activate *AtMYBS1* and *AtMYBS2* (**Figure 7F**). Moreover, *AtGLK2* and both *AtMYBS* proteins activated *AtGLK1*, and *AtMYBS1* and *AtMYBS2* activated *AtGLK2* (**Figure 7F**). This capacity for *AtMYBS1* and *AtMYBS2* to recognise their own promoters as well as *AtGLK1* is consistent with binding detected from DAP-seq (**Figure S7H**). We note no binding of *AtMYBS1* or *AtMYBS2* to the *AtGLK2* promoter was detected (**Figure S7H**) despite 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 *AtRR-MYBs* and *AtGLK* *in vivo*, with *AtGLK2* being upregulated in the double *Attr-mybs1,2* mutant (**Figure 7G**). Of the 98 genes downregulated in both *Atmybs1,2* and *Atglk1,2*²⁵, more were shared than would be expected by chance (**Figure 7H, Figure S7I, Dataset S1**), and of these, 27 were related to photosynthesis encoding for

example components of the light harvesting complexes and the photosystems (**Figure S7J**, **Dataset 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 sulphur assimilation, as well as the biosynthesis of amino acids, fatty acids and carotenoids and so understanding their biogenesis has long been of interest.^{34–38} 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 *SEDOPHOSPHATE BISPHOSPHATASE*,^{39,40} faster relaxation of non-photochemical quenching of photosystem II⁴¹ and rerouting of photorespiration.⁴² A complementary approach to improving photosynthesis predicted to increase yield by up to 50% would be to convert C₃ crops to use the more efficient C₄ pathway.^{43,44} Introducing C₄ photosynthesis in C₃ crops such as rice would require a remodelling of chloroplast biogenesis in mesophyll and bundle sheath cells.⁴⁵

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.^{8,10,46} Overexpression 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 C₄ pathway.³⁶ However, we have an incomplete understanding of transcription factors allowing chloroplast development. For example, *GLK* and *CGA1/GNC* loss of function mutants still possess small chloroplasts⁹ 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 *Mpglk,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*.^{47,48} 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 localisation of indispensable proteins in the chloroplast are also often associated with embryo lethality.^{47,48} 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 analysed and so at present 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 *MpRR-MYB5* overexpression failed to upregulate *MpGLK* transcripts, MYB transcription factors commonly act in multimeric complexes involving bHLH and WD40 proteins⁴⁹ or with MYCs proteins.⁵⁰ It is therefore possible that additional partners need to be over-expressed in combination with the RR-MYBs to increase expression of *MpGLK*. 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.³⁶ When *MpGLK* was mis-expressed in the double *Mpr-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 factor are needed to drive full photosynthesis gene expression, or *MpGLK* 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.³⁶ 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 *MpGLK* in the presence of RR-MYBs¹⁶ in the absence of *MpRR-MYBs* the impact of *MpGLK* 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 *AtGLK* by

AtMYBS1&2 being more evident, and members of the GATA transcription factor family also controlling photosynthesis gene expression. Moreover, inducible overexpression of AtMYBS1 in *A. thaliana* has been reported to increase expression of photosynthesis genes.²⁴ Although similar sets of genes were down-regulated in loss of function *Mprr-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 *MpRR-MYB2*, the pale phenotype of *MpRR-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 LeMYB1 has been reported to bind the promoter of the *RBCS* gene⁵¹. 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 *Atmybs1* mutants⁵². Along with transcription factors belonging to the GLK, B-Box and Nuclear Factor-Y families, random forest analysis of gene expression recently predicted that RR-MYBs regulate photosynthesis gene expression.²⁴ Moreover, an inducible AtMYBS1 over-expressor line showed upregulation of photosynthesis genes that we detected as downregulated in *Atmybs1,mybs2*.

Penetrance of the RR-MYBs on photosynthesis gene expression and chloroplast biogenesis in *M. polymorpha* was striking, with *MpRR-MYBs* activating genes allowing CO₂ fixation as well as light harvesting. Chloroplasts of *Mpmyb5,2* mutants were ~30% smaller than those of *Mpglk* mutants, and double *Mpmyb5,2* mutants were paler than those of *Mpglk*. 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.⁵³ Such cooperative binding of transcription factors is thought to allow greater variety of expression outputs.^{55,56} 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 analysis of *M. polymorpha* and *A. thaliana* whose last common ancestor 400 mya 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 plants²⁷ it seems plausible they play a conserved role in chloroplast biogenesis. While we were unable to detect MpRR-MYBs in the Zygnematophyceae algae that are sister to the land plants⁵⁴ they are in fact present in the Klebsormidiophyceae and Charophyceae^{55,56} (**Figure S6B**) representing the other two most closely related algal lineages to land plants. *GLK* homologs are present in green algae⁵⁴ 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 colonisation 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 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 signalling 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 ORE and ATAF1^{57,58} will need to be defined. It will also be important to discover whether the RR-MYBs are integrated into retrograde signalling^{10,59} 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 the mode of action of these transcription factors, but also how they themselves are regulated.

STAR METHODS

KEY RESOURCES TABLE

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 Table S1, S2, Datasets 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 author upon request.

EXPERIMENTAL MODELS AND STUDY PARTICIPANT DETAILS

Marchantia polymorpha and growth conditions

Arabidopsis thaliana and growth conditions

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

Acknowledgements

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 useful comments and support. This work was funded by BBSRC BBP0031171 and ERACAPS C4BREED grants to J.M.H., BBSRC/EPSRC BB/L014130/1 and BBSRC BB/F011458/1 to J.P.H. T.B.S. was supported by a SNSF Postdoc Mobility Fellowship (P500PB_203128) and EMBO Long-Term Fellowship (ALTF 531-2019). For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Author contributions

N.Y.E., E.F. and J.M.H. conceived and designed the work. J.M.H. guided execution of experiments and oversaw the project. N.Y.E., E.F., K.B., L.H., T.S., M.T. and P.D. did the experiments and analysed the data. N.Y.E, E.F. and J.M.H. wrote the manuscript with input from all authors.

Declaration of interests

The authors declare no competing interests.

Main Figure titles and legends

Figure 1: MpRR-MYB5 and MpRR-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 sequencing data from *Arabidopsis thaliana* during de-etiolation²⁰ and *Marchantia polymorpha* during spore germination¹⁹ were examined. Transcription factors (TFs) upregulated in response to light in both species with either unknown or related to chlorophyll function were retained. As we hypothesised 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* but by a maximum of two in *M. polymorpha*. **B)** Heatmaps showing transcript abundance (z-score) of candidates that were selected to generate knockout mutants in *M. polymorpha*. **C)** Top: Schematic of MpRR-MYB5 and MpRR-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 *Mprr-myb5*, *Mprr-myb2* and *Mprr-myb5,2* mutants. Scale bars represent 2 mm. **H)** Chlorophyll content of *Mprr-myb5*, *Mprr-myb2* and *Mprr-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 one standard deviation, $n = 5$. **I)** Heatmaps showing transcript abundance (z-score) of MpGLK, MpRR-MYB2 and MpRR-MYB5 in germinating spores during the first 96 hours 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 Di-methyl 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 *Mprr-myb5*, *Mprr-myb2* and *Mprr-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, *Mprr-myb5*, *Mprr-myb2* and *Mprr-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, *Mprr-myb5*, *Mprr-myb2* and *Mprr-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 $p < 0.01$. Data presented as means and one standard deviation, $n = 150$. **T)** Starch levels in control, *Mprr-myb2*,

Mprr-myb5 and Mprr-myb5,2 mutants. Data presented as means and one standard deviation, n = 3.

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

A-D) Representative images of control, Mprr-myb5, Mpglk, and Mpglk,rr-myb5 mutants. Scale bars represent 7 mm. **E)** Chlorophyll content in Mprr-myb5, Mpglk and the double Mpglk,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 one standard deviation, 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 Dimethyl 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, Mprr-myb5, Mpglk and Mpglk,rr-myb5 mutants. Chlorophyll autofluorescence shown in magenta, cell borders marked with dashed white lines. Scale bars represent 10 μ m. **K-N)** Representative images from transmission electron microscopy images of control, Mprr-myb5, Mpglk and Mpglk,rr-myb5 mutants. Scale bars represent 1 μ m. Dashed area depicted in each chloroplast is enlarged and granal stacks indicated with red arrows. **O)** Chloroplast area of Mpglk and Mpglk,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. Data presented as means and one standard deviation, n=330.

Figure 3: MpRR-MYB5 and MpRR-MYB2 regulate expression of genes encoding chlorophyll biosynthesis and also the Calvin Benson Bassham cycle. A&B)

Upset diagrams showing sets of upregulated genes in MpRR-MYB5, MpRR-MYB2 and MpGLK over-expression lines (**A**), and downregulated genes in Mpglk, Mprr-myb2, Mprr-myb5, and the double Mpglk,rr-myb5 or Mprr-myb5,2 mutants (**B**). **C)** Enriched Gene Ontology terms for Mprr-myb5, Mpglk, Mpglk,rr-myb5 and Mprr-myb5,2 mutants. **D)** Heatmap illustrating the extent of down-regulation of transcripts encoding enzymes of chlorophyll biosynthesis in Mprr-myb2, Mprr-myb5, Mpglk, and the Mpglk,rr-myb5 or Mprr-myb5,2 mutant alleles. **E)** Heatmap indicating lower transcript abundance of genes encoding components of Calvin-Benson-Bassham cycle in Mprr-myb2, Mprr-myb5, Mpglk, as well as Mpglk,rr-myb5,myb2 and Mprr-myb5,myb2 double mutants. Schematic of Calvin-Benson-Bassham modified from Lea and Leegood, 1998.

Figure 4: MpRR-MYB5&2 target photosynthesis genes. A-B) Schematic of thylakoid membranes with photosystems (modified from¹⁰ and ²⁵), and heatmap indicating lower transcript abundance of genes encoding components of Photosystem II, Photosystem I and the cytochrome *b₆f* complex in *Mprr-myb2*, *Mprr-myb5*, *Mpglk* as well as *Mpglk,rr-myb5* and *Mprr-myb5,myb2* double mutants. **C)** Immunoblot analysis of LHCA, PsbA, PsbC and PsbS in *Mprr-myb2*, *Mpglk*, *Mprr-myb5* and *Mprr-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 mis-regulated genes in *Mprr-myb5,2* mutants. **E)** Motif sequence logos for MpRR-MYB2, MpRR-MYB5 and MpGLK identified by DAP-seq and ChIP-seq respectively. **F)** Fisher exact test and permutation testing for RR-MYB motifs in 500 base pair promoters from photosynthesis versus 1000 random gene sets. Red line indicates the frequency of the motif in *M. polymorpha* photosynthesis genes. **G)** MpRR-MYB5 and MpRR-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 MpRR-MYB2 or MpRR-MYB5 on expression from promoters of *MpPsbQ*, *MpPsaD* and *MpLHCB1*. Data presented as means and standard deviation (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 *MpRR-MYB5* and *MpRR-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 MpRR-MYB5, MpRR-MYB2 and MpGLK on expression from promoters of *MpRR-MYB5* and *MpRR-MYB2*. Data presented as means and one standard deviation $\pm$ SD (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 *MpGLK*, *MpRR-MYB2* and *MpRR-MYB5* in control, *Mprr-glk*, *Mprr-myb5*, *Mprr-myb2* and *Mprr-myb5,2* mutant backgrounds. Data presented as Transcript Per Million (TPM), and P-values of two-tailed *t*-test are shown.

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 *MpRR-MYB5*, *MpRR-MYB2* and *MpGLK*. Each transcription factor was placed in dashed box downstream of the GFP cassette. **B)** Representative image of wild type. **C)** Representative images of *Mprr-myb5* complemented with *MpRR-MYB5*, *MpRR-MYB2* and *MpGLK*. **D)** Chlorophyll content in control and each of the mutant backgrounds compared with controls. Letters show ranking using a *post hoc* Tukey test (comparisons are made between

groups highlighted within the same colour rectangle, different letters indicating statistically significant differences at $P < 0.01$), $n=5$. **E)** Representative images of *Mprr-glK* complemented with *MpRR-MYB5*, *MpRR-MYB2* and *MpGLK*. **F)** Representative images of *Mp-glK,rr-myb5* mutant with and without complemented by *MpRR-MYB2*. **G)** Representative images of *Mprr-myb5,2* mutant with and without complementation by *MpGLK*. **H)** Chlorophyll content in *Mprr-myb5,2* mutant with and without complementation from *MpGLK*. Scale bars represent 2 mm (**C,E,F,G**).

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 *AtMYBS1* and *AtMYBS2* gene structure showing exons as grey rectangles and guide (g) RNAs positions for gene editing. **C-F)** Images of two-week old seedlings of wild type, *Atmybs1*, *Atmybs2* and *Atmybs1,mybs2* mutants, Scale bars represent 7 mm. Representative images after confocal laser scanning microscopy also shown below, with scale bars representing 25 μ m. **G,H)** Representative images of wild type and *Atmybs1,mybs2* mutants with inflorescence. Scale bars represent 1.5 cm. **I)** Violin plots of chloroplast area for wild type, *Atmybs1*, *Atmybs2* and *Atmybs1,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 *Atmybs1,mybs2* mutants. Scale bars represent 5 μ m. **K)** Chlorophyll content of wild type, *Atmybs1*, *Atmybs2* and *Atmybs1,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 one standard error, $n=5$. **L)** Representative images of wild type, *Atglk1,glk2* and *Atmybs1,mybs2* seedlings undergoing de-etiolation. Scale bars represent 500 μ m. **M)** Chlorophyll content wild type, *Atglk1,glk2* and *Atmybs1,mybs2* seedlings during the de-etiolation timecourse. Data presented as mean and one standard deviation of the mean, $n = 3$. **N)** Chlorophyll fluorescence parameter F_v/F_m during de-etiolation of wild type, *Atglk1,glk2* and *Atmybs1,mybs2*. **O)** Transcript abundance of *AtGLK1*, *AtGLK2*, *AtMYBS1* and *AtMYBS2* genes during de-etiolation. *UBP6* (At1g51710) was used as internal control. Data presented as means and one standard deviation of the mean, $n = 3$.

647 **Figure 7: AtMYBS1&2 can bind and activate photosynthesis genes as well as transcription**
 648 **factors involved in chloroplasts biogenesis. A)** Motif sequence logo for AtRR-MYB³³. **B)** Fisher
 649 exact test and permutation testing for AtRR-MYB motifs in 500 base pair promoters from
 650 photosynthesis versus 1000 random gene sets, red line indicates frequency in *A. thaliana*
 651 photosynthesis genes. **C)** Venn diagram showing the overlap between AtMYBS1 and AtMYBS2
 652 binding sites and mis-regulated genes in *Atmybs1,mybs2* mutants. **D)** Genes relating to
 653 chlorophyll biosynthesis, the Calvin Benson Bassham cycle, photorespiration and the light
 654 harvesting complexes that were downregulated in *Atmybs1,mybs2* (highlighted with red, genes
 655 selected for the effectors assays). **E)** Effector assays to test interactions between AtMYBS1&2
 656 and photosynthesis genes. Schematic of effector-reporter (LUC) constructs used to assess
 657 binding of AtMYBS1 and AtMYBS2 to *AtLHCB4*, *AtGAPA2*, *AtLHCB5*, *AtPsaH2*, *AtLHCA2* and
 658 *AtLHCA4*. Data are presented as mean and standard deviation, n = 4. Asterisks indicate a
 659 statistically significant difference compared with control (*P<0.05, **P<0.01, t-test). **F)** Effector
 660 assays to test interactions between AtMYBS1&2 and AtGLK1&2. Schematic of effector-reporter
 661 (LUC) constructs used as well as bar graphs showing impact of each transcription factor on each
 662 other. Data presented as means and one standard deviation, n = 4). Asterisks indicate a
 663 statistically significant difference compared with control (*P<0.05, **P<0.01, n.s = non significant
 664 t-test). **G)** Transcript abundance in Transcripts Per Million (TPM) of *AtGLK1* and *AtGLK2* in wild-
 665 type and *Atmybs1,mybs2* mutant background. P-values of two-tailed t-test are shown. **H)** Venn
 666 diagram showing overlap between mis-regulated genes in *Atglk,glk2* and *Atmybs1,mybs2*
 667 mutants (*padj*-value ≤0.01, Log2FC ≥ 1-fold).

Supplemental Figure titles and legends

Figure S1: CRISPR/Cas9 mediated gene editing of MpRR-MYB5 and MpRR-MYB2. A)

Schematic representation of MpRR-MYB5 gene with exons shown as black boxes, position of gRNA shown with arrow, and sequence analysis of *Mprr-myb5* knockout lines. Wild-type *Marchantia 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) Schematic representation of MpRR-MYB2 gene with exons shown as black boxes, position of gRNA shown with arrow and sequence analysis of *Mprr-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 MpRR-MYB5 from the 3000 base pair fragment upstream of the predicted translational start codon. Representative phenotypes of control, and *Mprr-myb5* mutants complemented with the *proMpRR-MYB5::MpRR-MYB5*. Scale bars represent 5 mm. Chlorophyll content in control, and *Mprr-myb5* mutants complemented with the *proMpRR-MYB5::MpRR-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 *Mprr-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, *Mprr-myb2*, *Mprr-myb5* and *Mprr-myb5,2* thallus cross sections.

Figure S2: MpRR-MYB5, MpRR-MYB2 and MpRR-MYB2&5 overexpression analysis. A-E)

Schematics of constructs used to overexpress MpRR-MYB5, MpRR-MYB2 and MpRR-MYB2&5. F-I) Representative images of control, MpRR-MYB5, MpRR-MYB2 and MpRR-MYB2&5 over-expressors. Scale bars represent 2 mm. J-M) Representative images from confocal microscopy images of control, MpRR-MYB5, MpRR-MYB2 and MpRR-MYB2&5 over-expressors. Scale bars represent 25 μm . N) qPCR analysis of MpRR-MYB5, MpRR-MYB2 and MpRR-MYB2&5 over-expressors. *ADENINE PHOSPHORIBOSYL TRANSFERASE 3* (APT) and *ACTIN 7* (ACT) were used as housekeeping gene controls. O) Chlorophyll content for MpRR-MYB5, MpRR-MYB2 and MpRR-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 MpRR-MYB5, MpRR-MYB2 and MpRR-MYB2 & 5 over-expressors. Letters show ranking derived from a *post hoc* Tukey test (with different letters indicating statistically significant differences at $P < 0.01$).

Figure S3: CRISPR/Cas9 mediated gene editing of MpGLK, MpGATA4, MpRR-MYB5 and MpRR-MYB2. **A)** Top: Schematic of MpGLK gene with exons indicated by black boxes, position of gRNA shown with an arrow. Bottom: Sequence analysis of Mpglk 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 MpGLK and MpRR-MYB5 genes with exons indicated by black boxes. Position of gRNA shown with an arrow. Bottom: Sequence analysis of Mpglk,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 MpGLK gene as shown in B) along with sequence analysis of Mpglk mutations in Mpglk,rr-myb2 knockout mutant lines. **D)** Representative images of Mpglk and Mpglk,rr-myb2 mutants (scale bars represent 5 mm), and chlorophyll content of Mpglk and Mpglk,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 Mpgata4,rr-myb5 knockout mutants. **F)** Representative images of Mpgata4, Mprrr-myb5 and Mpgata4,rr-myb5 mutants (scale bars represent 5 mm), and chlorophyll content of Mpgata4, Mprrr-myb5 and Mpgata4,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 Mpglk,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 colouring used for each column depends on the fraction of that column made up of amino acids as follows: black 100% similar, dark-grey 80–100% similar, lighter grey 60–80% similar, white less than 60% similar.

Figure S4: *M. polymorpha* RNA sequencing. **A-B)** Volcano plots showing differentially expressed genes in MpRR-MYB5 overexpression versus controls, as well as in MpRR-MYB2 overexpression lines versus controls. Blue dots indicate genes with a padj-value ≥ 0.01 , while black have a padj-value ≤ 0.01 . Total number of differentially expressed genes (DEGs) indicated at top of each panel. **C)** Upset diagrams showing shared downregulated genes in MpRR-MYB5, MpRR-MYB2 and MpGLK over-expression lines. **D)** Enriched gene ontology terms in MpGLK and MpRR-MYB2 over-expressors. **E-I)** Volcano plots showing differentially expressed genes for Mprrr-myb5, Mprrr-myb2, Mpglk, Mpglk,rr-myb5 and Mprrr-myb5,2 mutants compared with controls.

J) Upset diagram showing sets of upregulated genes in *Mpglk*, *Mprrr-myb2*, *Mprrr-myb5*, and the double *Mpglk,rr-myb5* or *Mprrr-myb5,2* mutants. **K)** Enriched gene ontology terms for upregulated genes in *Mpglk*, *Mprrr-myb5*, *Mpglk,rr-myb5*, *Mpglk* and *Mprrr-myb5,2* mutants. **L)** Heatmap of differentially expressed photorespiration genes in *Mprrr-myb2*, *Mprrr-myb5*, *Mpglk* as well as *Mpglk,rr-myb5* and *Mprrr-myb5,2* double mutants. **M)** Heatmap of differentially expressed photosynthesis associated genes in *Mprrr-myb2*, *Mprrr-myb5*, *Mpglk* as well as *Mpglk,rr-myb5* and *Mprrr-myb5,2* double mutants.

Figure S5: *M. polymorpha* DAP-seq, effector assays and quantitative polymerase chain reactions for confirmation of transgene over-expression. **A)** Fisher exact test and permutation testing for misregulated genes in *Mprrr-myb5,2* mutants that also contained a *MpRR-MYB2* or *MpRR-MYB5* binding site. This was compared with a 976 gene set obtained by randomly sampling genes from the RNA-seq dataset 1000 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 *MpRR-MYB5* can induce activation of *proMpPsaN::LUC*, *proMpPetA::LUC*, *proMpDVR::LUC* and *proMpPORA::LUC*. Data represent means $\pm$ one standard deviation ($n = 4$), 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 *MpGLK* promoter. Schematic structures of effector and reporter constructs used in the luciferase (LUC) reporter system. Bar graphs show *MpRR-MYB5*, *MpRR-MYB2* and *MpGLK* induction or lack of activation of *proMpGLK::LUC*. Data represent means $\pm$ one standard deviation ($n = 4$), 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 *Mprrr-myb5*, *Mpglk*, *Mpglk,rr-myb5* and *Mprrr-myb5,2* mutants complemented with *MpRR-MYB5*, *MpRR-MYB2* and *MpGLK* under control of the *MpUBE2* promoter. *ADENINE PHOSPHORIBOSYL TRANSFERASE 3 (APT)* and *CULLIN 1 (CUL)* were used as housekeeping gene controls (Saint-Marcoux et al., 2015).

Figure S6: The RR-MYB/CCA1-like related transcription factor family. **A)** Schematic representation of RR-MYB/CCA1-like protein structure. *MpRR-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 categorised into three additional groups based on intron position within the R1/2 domain, called a, c and d (Du et al. 2013). 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. Rice has a total of 25 RR-MYB/CCA1-like genes, of which eight belong to the a class, six to the c class, nine to the d class and nine 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 colouring used for each column depends on the fraction of that column made up of amino acids as follows: black 100% similar, dark-grey 80–100% similar, lighter grey 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. 2010). *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. 2018). *M. polymorpha*, *A. thaliana*, rice and tomato proteins indicated with red, blue, yellow and black arrows respectively. **F)** Transcript abundance of *AtGLK1*, *AtGLK2*, *AtMYBS1* and *AtMYBS2* genes during de-etiolation. *EF1A* was used as internal control. Data presented as means and one standard deviation of the mean, n = 3.

Figure S7: CRISPR/Cas9 mediated gene editing for *AtMYBS1* and *AtMYBS2* as well as DAP-seq and effector assays. A-B) Top: Schematic of the *AtMYBS1* and *AtMYBS2* genes with exons shown as grey boxes. Positions of gRNAs shown with a grey/red arrow. Bottom: Sequence analysis of *Atmybs1* and *Atmybs2* 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 a *AtMYBS1* or *AtMYBS2* binding site and that were mis-regulated in *Atmybs1,myb2* mutants. A set of 487 genes were randomly sampled 1000 times

from the RNA-seq dataset. **D)** Volcano plot showing differentially expressed genes in *Atmybs1,mybs2* mutants. Blue dots indicate genes with a padj-value ≥ 0.01 , while black ones have a padj-value ≤ 0.01 . Total number of DEGs indicated at top of each panel. **E)** Top five GO terms in *Atmybs1,mybs2* mutants estimated using PANTHER (GO-Slim Biological Processes). **F)** Boxplots for representative downregulated photosynthesis associated genes in *Atmybs1,mybs2* mutants. Data presented as Transcripts per Million (TPM) values. P-values of two-tailed *t*-test shown. **G-H)** AtMYBS1 and AtMYBS2 binding sites in photosynthesis or AtMYBS1&S2 genes as identified by DAP-seq ([O'Malley et al 2016](#)). Gene body shown at bottom of each panel. **I)** Top GO terms in *Atglk1,glk2* mutants estimated using PANTHER (GO-Slim Biological Processes). **J)** Fisher exact test and permutation testing for overlap between downregulated genes in *Atglk1,glk2* and *Atmybs1,myb2* mutants, compared with 1000 randomly sampled sets of genes (size = 98) from the RNA-seq dataset. **K)** Heatmap (z-score) showing temporal response of the 98 genes found to be regulated by AtGLK1, AtGLK2, AtMYBS1 and AtMYBS2 gene after exposure to light (REF). **L)** Scatter plots showing position and predicted binding affinity of GLK, RR-MYB binding and other motifs 3000 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 *Atmybs1* (SAIL_1184_D04) and *Atmybs2* (SALK_150774) mutant lines. Top: Schematic of AtMYBS1 and AtMYBS2 genes with exons shown as grey rectangles. The 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 *Atmybs1* and *Atmybs2* mutants respectively. Below: PCR analysis for confirmation of T-DNA insert in mutants.

Supplemental files and one sentence descriptive title

Figure S1: CRISPR/Cas9 mediated gene editing of MpRR-MYB5 and MpRR-MYB2, related to Figure 1.

Figure S2: MpRR-MYB5, MpRR-MYB2 and MpRR-MYB2&5 overexpression analysis, related to Figure 1.

Figure S3: CRISPR/Cas9 mediated gene editing of MpGLK, MpGATA4, MpRR-MYB5 and MpRR-MYB2, related to Figure 2.

Figure S4: *M. polymorpha* RNA sequencing, related to Figures 3 and 4.

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.

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

Figure S7: CRISPR/Cas9 mediated gene editing for AtMYBS1 and AtMYBS2 as well as DAP-seq and effector assays, related to Figure 7.

Table S1. Transcription factors upregulated during *A. thaliana* de-etiolation, data related to Figure 1 and S7.

Table S2. Transcription factors upregulated during *M. polymorpha* spore germination, data related to Figure 1.

Table S3. List of oligonucleotides used in this study, related to STAR Methods.

Data S1. RNA-sequencing analysis data, related to Figures 3, 4, 7, S4 and S7. Mpr-r-myb2 genotyping data, related to Figure 1.

Data S2. Gene Ontology analysis and RR-MYB & GLK binding motif analysis, related to Figures 3, 4, S4 and S7.

Data S3. DAP-sequencing data, related to Figures 4, 7, S4 and S7.

Data S4. Phylogenetic analysis data, related to Figures 6 and S6.

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 $\mu\text{mol m}^{-2}\text{s}^{-1}$. *Arabidopsis thaliana* Col-0 was used throughout exception of *Atmybs2* 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 $\mu\text{mol m}^{-2}\text{s}^{-1}$ light. *Arabidopsis* T-DNA insertion mutants *Atmybs1* (SAIL_1184_D04) and *Atmybs2* (SALK_150774) were obtained with T-DNA and zygosity confirmed by PCR (**Figure S7M and Table S3**). For gene editing gRNAs were predicted using CasFinder. Several gRNAs (**Table S3**) per target were tested.⁶⁰ Single gRNAs were cloned²⁹ into pMpGE013⁶¹. To complement *Mprrr-myb5* mutants the *MpRR-MYB5* promoter⁶² was used. For overexpression *MpRR-MYB2* and *MpRR-MYB5* coding sequences were synthesised and cloned into the pUAP4 vector.²⁹ For complementation, gRNA resistant *MpGLK*, *MpRR-MYB2* and *MpRR-MYB5* coding sequences were synthesised and cloned into the pUAP4 vector. OpenPlant parts used listed in Star Methods. To generate *Atmybs1,s2* mutants two gRNAs per gene were cloned into pEn-Chimera⁶³ after⁶⁴. 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.⁶⁵ *A. thaliana* was transformed by floral dipping⁶⁶ and genotyping performed as reported previously.⁶⁷ T2 plants with confirmed edits were analysed.

For *M. polymorpha* transformation, 5 mL LB media were inoculated with 3-4 *Agrobacterium* colonies (GV3101: 50 $\mu\text{g/mL}$ rifampicin, 25 $\mu\text{g/mL}$ 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 , 25g/L KCl, 25g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 100g/L $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$, 12.5mg $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 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 and 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). 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 μ L 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 μ L sterile water added to each tube. 5 μ L 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 ($<1 \mu\text{mol m}^{-2} \text{s}^{-1}$) applied to evaluate dark-adapted minimum fluorescence (F_o), and a subsequent saturating pulse of $3000 \mu\text{mol m}^{-2} \text{s}^{-1}$ used to evaluate dark-adapted maximum fluorescence (F_m). A total of three plants were measured per genotype and treatment. 20 μ M 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 $100 \mu\text{mol m}^{-2} \text{s}^{-1}$. 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 μ L 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 μ L 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 $\times$ /0.40 CS2) or 20X air objective (HC PL APO 20 $\times$ /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 mm² sections of 5-6 individual 3-week-

old thalli were harvested, fixed, embedded and imaged as previously described.⁶⁸ Chloroplast area was measured using ImageJ and the Macro in **Table S4**.

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 $\mu\text{mol m}^{-2}\text{s}^{-1}$ 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 μ M. 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'-
GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGATGACTATCTCGTATGCCGTCTTCTG		
CTTG-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 Kallisto⁶⁹ to the *M. polymorpha* genome version 5 (transcripts only, obtained from MarpolBase)⁷⁰. Mapping statistics for each library are provided in **Dataset S1**. DGE analysis was performed with DESeq2⁷¹, 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 Kallisto⁶⁹ to the *A. thaliana* TAIR10 genome (transcripts only). Mapping statistics for each library are provided in **Dataset S1**. DGE analysis was performed with DESeq2⁷¹, 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 μ mol 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 gemmae 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 $\mu\text{mol m}^{-2} \text{s}^{-1}$. 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 iTAK⁷² and PlantTFDB⁷³, Phytozome, Fernbase⁷⁴, Phycozome and PhytoPlaza databases. Sequences for each species were aligned with the MpRR-MYB5 and *A. thaliana* MYBS1 and MYBS2 amino acid sequences using MAFFT⁷⁵. 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)⁷⁶, green algae genomes in PhycoCosm (/phycocosm.jgi.doe.gov), and 1KP using the MpRR-MYB5 and *A. thaliana* MYBS1 and MYBS2 amino acid sequence. Identified RR-Myb-related/CCA1-like protein sequences were aligned using MAFFT and trimmed using TrimAl⁷⁷. A maximum likelihood phylogenetic tree was inferred using iQTree⁷⁸, ModelFinder⁷⁹ and ultrafast

approximation for phylogenetic bootstrap⁸⁰ and SH-aLRT test⁸¹. The tree was visualised using iTOL.⁸² Full list of sequences in **Dataset S4**.

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

M. polymorpha genomic DNA was extracted as described previously⁸³. 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 Germ Extract 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'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC 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 bowtie2⁸⁴ version 2.5.3, sorted and converted to bam format with samtools⁸⁵ version 1.19.2. Mapped reads were visualised using IGV_2.17.2, and peaks called

with MACS3⁸⁶ version 3.0.0, with bedtools⁸⁷ used to retrieve genomic coordinates of peaks. Nearest genes around the peaks were retrieved using the ChIPseeker R package.⁸⁸

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.⁸⁹ Motifs were merged to create RR-MYB combined motifs and visualised using the Ceqlogo tool from MEME.⁹⁰ 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 FIMO⁹¹. 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 **Dataset S2** the FIMO tool⁹¹ 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.⁸⁹ To account for input sequence composition, a background model was generated using the fasta-get-markov tool from the MEME suite.⁹⁰ 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.⁹² 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

1104 module with a maize ubiquitin promoter and CAMV35S terminator module. Details of promoter
1105 fragments cloned can be found in STAR methods. For each transformation, 2 µg of transformation
1106 control plasmid, 5 µg of reporter plasmid, and 5 µg of effector plasmid were combined and mixed
1107 with 180 µL protoplasts. After 20 hours incubation in the light with intensity of 40 µmol m⁻² s⁻¹, at
1108 22°C, proteins were extracted using a passive lysis buffer (Promega). Quantification of GUS
1109 activity was performed using a fluorometric MUG assay .⁹³ MUG assay buffer (50mM phosphate
1110 buffer (pH 7.0), 10 mM EDTA-Na₂, 0.1% [v/v] Triton X-100, 0.1% [w/v] N-lauroylsarcosine sodium,
1111 10 mM DTT, 2 mM 4-methylumbelliferyl-β-D-glucuronide (MUG)), with 20 µL of protein sample in
1112 a 200 µL total volume. The reaction was conducted at 37 °C in GREINER 96 F-BOTTOM microtiter
1113 plate using a CLARIOstar plate reader. 4-Methylumbelliferone (4-MU) fluorescence was recorded
1114 every 2 minutes for 20 cycles with excitation at 360 nm and emission detected at 450 nm. LUC
1115 activity was measured with 20 µL of protein sample and 100 µL of LUC assay reagent (Promega)
1116 using a Clariostar plate reader. Transcriptional activity from the promoter was calculated as the
1117 ratio of LUC luminescence to the rate of MU accumulation per second.

1118 REFERENCES

- 1119 1. Archibald, J.M. (2009). The Puzzle of Plastid Evolution. *Current Biology* 19, R81-R88.
- 1120 2. McFadden, G.I. (2014). Origin and Evolution of Plastids and Photosynthesis in Eukaryotes.
1121 Cold Spring Harb. Perspect. Biol. 6, a016105.
- 1122 3. Gould, S.B., Waller, R.F., and McFadden, G.I. (2008). Plastid Evolution. *Ann. Rev. Plant Biol.*
1123 59, 491-517.
- 1124 4. Abdallah, F., Salamini, F., and Leister, D. (2000). A prediction of the size and evolutionary
1125 origin of the proteome of chloroplasts of Arabidopsis. *Trends Plant Sci.* 5, 141–142.
- 1126 5. Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B.,
1127 Hasegawa, M., and Penny, D. (2002). Evolutionary analysis of Arabidopsis, cyanobacterial,
1128 and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes
1129 in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12246–12251.
- 1130 6. Long, S.P., Marshall-Colon, A., and Zhu, X.-G. (2015). Meeting the global food demand of
1131 the future by engineering crop photosynthesis and yield potential. *Cell* 161, 56–66.
- 1132 7. Ort, D.R., Merchant, S.S., Alric, J., Barkan, A., Blankenship, R.E., Bock, R., Croce, R.,
1133 Hanson, M.R., Hibberd, J.M., Long, S.P., et al. (2015). Redesigning photosynthesis to
1134 sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci. U. S. A.* 112,
1135 8529–8536.
- 1136 8. Cackett, L., Luginbuehl, L.H., Schreier, T.B., Lopez-Juez, E., and Hibberd, J.M. (2022).
1137 Chloroplast development in green plant tissues: the interplay between light, hormone, and
1138 transcriptional regulation. *New Phytol.* 233, 2000–2016.
- 1139 9. Zubo, Y.O., Blakley, I.C., Franco-Zorrilla, J.M., Yamburenko, M.V., Solano, R., Kieber, J.J.,
1140 Loraine, A.E., and Schaller, G.E. (2018). Coordination of Chloroplast Development through
1141 the Action of the GNC and GLK Transcription Factor Families. *Plant Physiol.* 178, 130–147.
- 1142 10. Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J., and Langdale, J.A.
1143 (2009). GLK transcription factors coordinate expression of the photosynthetic apparatus in
1144 Arabidopsis. *Plant Cell* 21, 1109–1128.
- 1145 11. Naito, T., Kiba, T., Koizumi, N., Yamashino, T., and Mizuno, T. (2007). Characterization of a

1146 unique GATA family gene that responds to both light and cytokinin in *Arabidopsis thaliana*.
1147 *Biosci. Biotechnol. Biochem.* **71**, 1557–1560.

1148 12. Fitter, D.W., Martin, D.J., Copley, M.J., Scotland, R.W., and Langdale, J.A. (2002). *GLK* gene
1149 pairs regulate chloroplast development in diverse plant species. *The Plant Journal* **31**: 713-
1150 27.

1151 13. Zhang, T., Zhang, R., Zeng, X.-Y., Lee, S., Ye, L.-H., Tian, S.-L., Zhang, Y.-J., Busch, W.,
1152 Zhou, W.-B., Zhu, X.-G., et al. (2024). GLK transcription factors accompany ELONGATED
1153 HYPOCOTYL5 to orchestrate light-induced seedling development in *Arabidopsis*. *Plant*
1154 *Physiol.* **194**, 2400–2421.

1155 14. Wang, P., Fouracre, J., Kelly, S., Karki, S., Gowik, U., Aubry, S., Shaw, M.K., Westhoff, P.,
1156 Slamet-Loedin, I.H., Paul Quick, W., et al. (2013). Evolution of GOLDEN2-LIKE gene function
1157 in C₃ and C₄ plants. *Planta* **237**, 481-495.

1158 15. Yasumura, Y., Moylan, E.C., and Langdale, J.A. (2005). A conserved transcription factor
1159 mediates nuclear control of organelle biogenesis in anciently diverged land plants. *Plant Cell*
1160 **17**, 1894–1907.

1161 16. Yelina, N.E., Frangedakis, E., Wang, Z., Schreier, T.B., Rever, J., Tomaselli, M., Forestier,
1162 E., Billakurthi, K., Ren, S., Bai, Y., et al. (2024). Streamlined regulation of chloroplast
1163 development in the liverwort *Marchantia polymorpha*. *bioRxiv*, 2023.01.23.525199.
1164 10.1101/2023.01.23.525199.

1165 17. Chiang, Y.-H., Zubo, Y.O., Tapken, W., Kim, H.J., Lavanway, A.M., Howard, L., Pilon, M.,
1166 Kieber, J.J., and Schaller, G.E. (2012). Functional characterization of the GATA transcription
1167 factors GNC and CGA1 reveals their key role in chloroplast development, growth, and
1168 division in *Arabidopsis*. *Plant Physiol.* **160**, 332–348.

1169 18. Bowman, J.L., Arteaga-Vazquez, M., Berger, F., Briginshaw, L.N., Carella, P., Aguilar-Cruz,
1170 A., Davies, K.M., Dierschke, T., Dolan, L., Dorantes-Acosta, A.E., et al. (2022). The
1171 renaissance and enlightenment of *Marchantia* as a model system. *Plant Cell* **34**, 3512–3542.

1172 19. Flores-Sandoval, E., Romani, F., and Bowman, J.L. (2018). Co-expression and
1173 Transcriptome Analysis of Transcription Factors Supports Class C ARFs as Independent
1174 Actors of an Ancient Auxin Regulatory Module. *Front. Plant Sci.* **9**, 1345.

- 1175 20. Sullivan, A.M., Arsovski, A.A., Lempe, J., Bubb, K.L., Weirauch, M.T., Sabo, P.J., Sandstrom,
1176 R., Thurman, R.E., Neph, S., Reynolds, A.P., et al. (2014). Mapping and Dynamics of
1177 Regulatory DNA and Transcription Factor Networks in *A. thaliana*. *Cell Reports* 8, 2015–2030.
- 1178 21. Bastakis, E. (2017). The Contribution of the GATA Transcription Factors GNC and GNL in
1179 the Greening of *Arabidopsis thaliana*. [PhD thesis](#). Technische Universität München.
- 1180 22. Bursch, K., Toledo-Ortiz, G., Pireyre, M., Lohr, M., Braatz, C., and Johansson, H. (2020).
1181 Identification of BBX proteins as rate-limiting cofactors of HY5. *Nat Plants* 6, 921–928.
- 1182 23. Zhao, P., Cui, R., Xu, P., Wu, J., Mao, J.-L., Chen, Y., Zhou, C.-Z., Yu, L.-H., and Xiang, C.-
1183 B. (2017). ATHB17 enhances stress tolerance by coordinating photosynthesis associated
1184 nuclear gene and ATSIG5 expression in response to abiotic stress. *Sci. Rep.* 7, 45492.
- 1185 24. Halpape, W., Wulf, D., Verwaaijen, B., Stasche, A.S., Zenker, S., Sielemann, J., Tschikin, S.,
1186 Viehöver, P., Sommer, M., Weber, A.P.M., et al. (2023). Transcription factors mediating
1187 regulation of photosynthesis. *bioRxiv*, <https://doi.org/10.1101/2023.01.06.522973>.
- 1188 25. Tu, X., Ren, S., Shen, W., Li, J., Li, Y., Li, C., Li, Y., Zong, Z., Xie, W., Grierson, D., et al.
1189 (2022). Limited conservation in cross-species comparison of GLK transcription factor binding
1190 suggested wide-spread cistrome divergence. *Nat. Commun.* 13, 7632.
- 1191 26. Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010).
1192 MYB transcription factors in *Arabidopsis*. *Trends in Plant Science* 15, 573–81.
- 1193 27. Du, H., Wang, Y.-B., Xie, Y., Liang, Z., Jiang, S.-J., Zhang, S.-S., Huang, Y.-B., and Tang,
1194 Y.-X. (2013). Genome-wide identification and evolutionary and expression analyses of MYB-
1195 related genes in land plants. *DNA Res.* 20, 437–448.
- 1196 28. Trebst, A. (2007). Inhibitors in the functional dissection of the photosynthetic electron
1197 transport system. *Photosynth. Res.* 92, 217–224.
- 1198 29. Sauret-Güeto, S., Frangedakis, E., Silvestri, L., Rebmann, M., Tomaselli, M., Markel, K.,
1199 Delmans, M., West, A., Patron, N.J., and Haseloff, J. (2020). Systematic Tools for
1200 Reprogramming Plant Gene Expression in a Simple Model. *ACS Synth. Biol.* 9, 864–882.
- 1201 30. Bartlett, A., O'Malley, R.C., Huang, S.-S.C., Galli, M., Nery, J.R., Gallavotti, A., and Ecker,
1202 J.R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat.*

1203 Protoc. 12, 1659–1672.

1204 31. Pu, X., Yang, L., Liu, L., Dong, X., Chen, S., Chen, Z., Liu, G., Jia, Y., Yuan, W., and Liu, L.
1205 (2020). Genome-Wide Analysis of the MYB Transcription Factor Superfamily in
1206 *Physcomitrella patens*. Int. J. Mol. Sci. doi:10.3390/ijms21030975

1207 32. Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhaofeng, G., Zhiqiang, L., Yunfei,
1208 Z., Xiaoxiao, W., Xiaoming, Q., et al. (2006). The MYB transcription factor superfamily of
1209 Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family.
1210 Plant Mol. Biol. 60, 107–124.

1211 33. O'Malley, R.C., Huang, S.-S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M.,
1212 Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the
1213 Regulatory DNA Landscape. Cell 165, 1280-1292.

1214 34. Witte, C.-P., and Herde, M. (2020). Nucleotide Metabolism in Plants. Plant Physiol. 182, 63–
1215 78.

1216 35. Waters, M.T., and Langdale, J.A. (2009). The making of a chloroplast. EMBO J. 28, 2861–
1217 2873.

1218 36. Wang, P., Khoshravesh, R., Karki, S., Tapia, R., Balahadia, C.P., Bandyopadhyay, A., Quick,
1219 W.P., Furbank, R., Sage, T.L., and Langdale, J.A. (2017). Re-creation of a Key Step in the
1220 Evolutionary Switch from C₃ to C₄ Leaf Anatomy. Curr. Biol. 27, 3278–3287.

1221 37. Jarvis, P., and López-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other
1222 plastids. Nat. Rev. Mol. Cell Biol. 14, 787–802.

1223 38. López-Juez, E. (2007). Plastid biogenesis, between light and shadows. J. Exp. Bot. 58, 11–
1224 26.

1225 39. Lefebvre, S., Lawson, T., Zakhleniuk, O.V., Lloyd, J.C., Raines, C.A., and Fryer, M. (2005).
1226 Increased sedoheptulose-1,7-bisphosphatase activity in transgenic tobacco plants stimulates
1227 photosynthesis and growth from an early stage in development. Plant Physiol. 138, 451–460.

1228 40. Driever, S.M., Simkin, A.J., Alotaibi, S., Fisk, S.J., Madgwick, P.J., Sparks, C.A., Jones, H.D.,
1229 Lawson, T., Parry, M.A.J., and Raines, C.A. (2017). Increased SBPase activity improves
1230 photosynthesis and grain yield in wheat grown in greenhouse conditions. Philos. Trans. R.

1231 Soc. Lond. B Biol. Sci. 372, 20160384.

1232 41. Kromdijk, J., Głowacka, K., Leonelli, L., Gabilly, S.T., Iwai, M., Niyogi, K.K., and Long, S.P.
 1233 (2016). Improving photosynthesis and crop productivity by accelerating recovery from
 1234 photoprotection. *Science* 354, 857–861.

1235 42. South, P.F., Cavanagh, A.P., Liu, H.W., and Ort, D.R. (2019). Synthetic glycolate metabolism
 1236 pathways stimulate crop growth and productivity in the field. *Science* 363, eaat9077.

1237 43. Hibberd, J.M., Sheehy, J.E., and Langdale, J.A. (2008). Using C₄ photosynthesis to increase
 1238 the yield of rice-rationale and feasibility. *Curr. Opin. Plant Biol.* 11, 228–231.

1239 44. von Caemmerer, S., Quick, W.P., and Furbank, R.T. (2012). The development of C₄ rice:
 1240 current progress and future challenges. *Science* 336, 1671–1672.

1241 45. Langdale, J.A. (2011). C₄ cycles: past, present, and future research on C₄ photosynthesis.
 1242 *Plant Cell* 23, 3879–3892.

1243 46. Bastakis, E., Hedtke, B., Klermund, C., Grimm, B., and Schwechheimer, C. (2018). LLM-
 1244 Domain B-GATA Transcription Factors Play Multifaceted Roles in Controlling Greening in
 1245 Arabidopsis. *Plant Cell* 30, 582–599.

1246 47. Budziszewski, G.J., Lewis, S.P., Glover, L.W., Reineke, J., Jones, G., Ziemnik, L.S.,
 1247 Lonowski, J., Nyfeler, B., Aux, G., Zhou, Q., et al. (2001). Arabidopsis genes essential for
 1248 seedling viability: isolation of insertional mutants and molecular cloning. *Genetics* 159, 1765–
 1249 1778.

1250 48. Bryant, N., Lloyd, J., Sweeney, C., Myouga, F., and Meinke, D. (2011). Identification of
 1251 nuclear genes encoding chloroplast-localized proteins required for embryo development in
 1252 Arabidopsis. *Plant Physiol.* 155, 1678–1689.

1253 49. Ramsay, N.A., and Glover, B.J. (2005). MYB-bHLH-WD40 protein complex and the evolution
 1254 of cellular diversity. *Trends Plant Sci.* 10, 63–70.

1255 50. Dickinson, P.J., Kneřová, J., Szecówka, M., Stevenson, S.R., Burgess, S.J., Mulvey, H.,
 1256 Bågman, A.-M., Gaudinier, A., Brady, S.M., and Hibberd, J.M. (2020). A bipartite transcription
 1257 factor module controlling expression in the bundle sheath of Arabidopsis thaliana. *Nat Plants*
 1258 6, 1468–1479.

- 1259 51. Rose, A., Meier, I., and Wienand, U. (1999). The tomato I-box binding factor LeMYBI is a
1260 member of a novel class of myb-like proteins. *Plant J.* 20, 641–652.
- 1261 52. Chen, Y.-S., Chao, Y.-C., Tseng, T.-W., Huang, C.-K., Lo, P.-C., and Lu, C.-A. (2017). Two
1262 MYB-related transcription factors play opposite roles in sugar signaling in *Arabidopsis*. *Plant*
1263 *Mol. Biol.* 93, 299–311.
- 1264 53. Tu, X., Mejía-Guerra, M.K., Valdes Franco, J.A., Tzeng, D., Chu, P.-Y., Shen, W., Wei, Y.,
1265 Dai, X., Li, P., Buckler, E.S., et al. (2020). Reconstructing the maize leaf regulatory network
1266 using ChIP-seq data of 104 transcription factors. *Nat. Commun.* 11, 5089.
- 1267 54. Cheng, S., Xian, W., Fu, Y., Marin, B., Keller, J., Wu, T., Sun, W., Li, X., Xu, Y., Zhang, Y.,
1268 et al. (2019). Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant
1269 Evolution. *Cell* 179, 1057–1067.e14.
- 1270 55. Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada,
1271 T., Mori, H., Tajima, N., et al. (2014). *Klebsormidium flaccidum* genome reveals primary
1272 factors for plant terrestrial adaptation. *Nat. Commun.* 5, 3978.
- 1273 56. Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K.K.,
1274 Haas, F.B., Vanderstraeten, L., Becker, D., Lang, D., et al. (2018). The *Chara* Genome:
1275 Secondary Complexity and Implications for Plant Terrestrialization. *Cell* 174, 448–464.e24.
- 1276 57. Rauf, M., Arif, M., Dortay, H., Matallana-Ramírez, L.P., Waters, M.T., Gil Nam, H., Lim, P.-
1277 O., Mueller-Roeber, B., and Balazadeh, S. (2013). ORE1 balances leaf senescence against
1278 maintenance by antagonizing G2-like-mediated transcription. *EMBO Rep.* 14, 382–388.
- 1279 58. Garapati, P., Xue, G.-P., Munné-Bosch, S., and Balazadeh, S. (2015). Transcription Factor
1280 ATAF1 in *Arabidopsis* Promotes Senescence by Direct Regulation of Key Chloroplast
1281 Maintenance and Senescence Transcriptional Cascades. *Plant Physiol.* 168, 1122–1139.
- 1282 59. Chan, K.X., Phua, S.Y., Crisp, P., McQuinn, R., and Pogson, B.J. (2016). Learning the
1283 Languages of the Chloroplast: Retrograde Signaling and Beyond. *Annu. Rev. Plant Biol.* 67,
1284 25–53.
- 1285 60. Yelina, N.E., Holland, D., Gonzalez-Jorge, S., Hirsz, D., Yang, Z., and Henderson, I.R.
1286 (2022). Coexpression of MEIOTIC-TOPOISOMERASE VIB-dCas9 with guide RNAs specific
1287 to a recombination hotspot is insufficient to increase crossover frequency in *Arabidopsis*. *G3*

1288 12, jkac105.

1289 61. Sugano, S.S., Nishihama, R., Shirakawa, M., Takagi, J., Matsuda, Y., Ishida, S., Shimada,
1290 T., Hara-Nishimura, I., Osakabe, K., and Kohchi, T. (2018). Efficient CRISPR/Cas9-based
1291 genome editing and its application to conditional genetic analysis in *Marchantia polymorpha*.
1292 PLoS One 13, e0205117.

1293 62. Romani, F., Sauret-Güeto, S., Rebmann, M., Annese, D., Bonter, I., Tomaselli, M.,
1294 Dierschke, T., Delmans, M., Frangedakis, E., Silvestri, L., et al. (2024). The landscape of
1295 transcription factor promoter activity during vegetative development in *Marchantia*. Plant Cell
1296 36, 2140-2159.

1297 63. Fauser, F., Schiml, S., and Puchta, H. (2014). Both CRISPR/Cas-based nucleases and
1298 nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. Plant J. 79,
1299 348–359.

1300 64. Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing
1301 capability with the endogenous tRNA-processing system. Proc. Natl. Acad. Sci. U. S. A. 112,
1302 3570–3575.

1303 65. Ursache, R., Fujita, S., Dénervaud Tendon, V., and Geldner, N. (2021). Combined
1304 fluorescent seed selection and multiplex CRISPR/Cas9 assembly for fast generation of
1305 multiple *Arabidopsis* mutants. Plant Methods 17, 111.

1306 66. Logemann, E., Birkenbihl, R.P., Ülker, B., and Somssich, I.E. (2006). An improved method
1307 for preparing *Agrobacterium* cells that simplifies the *Arabidopsis* transformation protocol.
1308 Plant Methods 2, 16.

1309 67. Edwards, K., Johnstone, C., and Thompson, C. (1991). A simple and rapid method for the
1310 preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res. 19, 1349.

1311 68. Schreier, T.B., Müller, K.H., Eicke, S., Faulkner, C., Zeeman, S.C., and Hibberd, J.M. (2023).
1312 Plasmodesmal connectivity in *C₄ Gynandropsis gynandra* is induced by light and dependent
1313 on photosynthesis. New Phytol. 241, 298-313

1314 69. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-
1315 seq quantification. Nat. Biotechnol. 34, 525–527.

1316 70. Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S.,
1317 Nishihama, R., Nakamura, Y., Berger, F., et al. (2017). Insights into Land Plant Evolution
1318 Garnered from the *Marchantia polymorpha* Genome. *Cell* 171, 287–304.e15.

1319 71. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and
1320 dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.

1321 72. Zheng, Y., Jiao, C., Sun, H., Rosli, H.G., Pombo, M.A., Zhang, P., Banf, M., Dai, X., Martin,
1322 G.B., Giovannoni, J.J., et al. (2016). iTAK: A Program for Genome-wide Prediction and
1323 Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein
1324 Kinases. *Mol. Plant* 9, 1667–1670.

1325 73. Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., and Gao, G. (2017). PlantTFDB
1326 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic
1327 Acids Res.* 45, D1040–D1045.

1328 74. Li, F.-W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P.-M., Eily, A.,
1329 Koppers, N., Kuo, L.-Y., Li, Z., et al. (2018). Fern genomes elucidate land plant evolution and
1330 cyanobacterial symbioses. *Nat Plants* 4, 460–472.

1331 75. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version
1332 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.

1333 76. Li, F.-W., Nishiyama, T., Waller, M., Frangedakis, E., Keller, J., Li, Z., Fernandez-Pozo, N.,
1334 Barker, M.S., Bennett, T., Blázquez, M.A., et al. (2020). *Anthoceros* genomes illuminate the
1335 origin of land plants and the unique biology of hornworts. *Nat Plants* 6, 259–272.

1336 77. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for
1337 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25,
1338 1972–1973.

1339 78. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and
1340 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.*
1341 32, 268–274.

1342 79. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermin, L.S. (2017).
1343 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14,
1344 587–589.

- 1345 80. Hoang, D., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., (2018) UFBoot2:
1346 Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 35, 518–
1347 522,
- 1348 81. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010).
1349 New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the
1350 performance of PhyML 3.0. *Syst. Biol.* 59, 307–321.
- 1351 82. Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new
1352 developments. *Nucleic Acids Res.* 47, W256–W259.
- 1353 83. Szövényi, P., Frangedakis, E., Ricca, M., Quandt, D., Wicke, S., and Langdale, J.A. (2015).
1354 Establishment of *Anthoceros agrestis* as a model species for studying the biology of
1355 hornworts. *BMC Plant Biol.* 15, 98.
- 1356 84. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat.*
1357 *Methods* 9, 357–359.
- 1358 85. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A.,
1359 Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and
1360 BCFtools. *Gigascience* 10. 10.1093/gigascience/giab008.
- 1361 86. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C.,
1362 Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS).
1363 *Genome Biol.* 9, R137.
- 1364 87. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing
1365 genomic features. *Bioinformatics* 26, 841–842.
- 1366 88. Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP
1367 peak annotation, comparison and visualization. *Bioinformatics* 31, 2382–2383.
- 1368 89. Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A.,
1369 Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D., et al. (2020). JASPAR 2020: update
1370 of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 48,
1371 D87–D92.
- 1372 90. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W.,

1373 and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids*
1374 *Res.* 37, W202–W208.

1375 91. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given
1376 motif. *Bioinformatics* 27, 1017–1018.

1377 92. Page, M.T., Parry, M.A.J., and Carmo-Silva, E. (2019). A high-throughput transient
1378 expression system for rice. *Plant Cell Environ.* 42, 2057–2064.

1379 93. Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: beta-glucuronidase
1380 as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* 6, 3901–3907.

1381 94. Delmans, M., Pollak, B., and Haseloff, J. (2017). MarpoDB: An Open Registry for *Marchantia*
1382 *Polymorpha* Genetic Parts. *Plant Cell Physiol.* 58, e5.

1383 95. Frangedakis, E., Waller, M., Nishiyama, T., Tsukaya, H., Xu, X., Yue, Y., Tjahjadi, M., Gunadi,
1384 A., Van Eck, J., Li, F.-W., et al. (2021). An Agrobacterium-mediated stable transformation
1385 technique for the hornwort model *Anthoceros agrestis*. *New Phytol.* 232, 1488–1505.