Synergistic control of chloroplast biogenesis by *MYB-related* and *Golden2-like* transcription factors

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

[1]Department of Plant Sciences, University of Cambridge, Cambridge, CB3 EA, UK
[2]Present address: Crop Science Centre, University of Cambridge, 93 Lawrence Weaver Road, Cambridge CB3 0LE, UK
[3]Present address: Department of Biology, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

* Equal contribution
$ correspondence:
Julian M. Hibberd
jmh65@cam.ac.uk
Eftychios Frangedakis
ef391@cam.ac.uk

Keywords: *Chloroplast biogenesis, photosynthesis, Marchantia, Arabidopsis, transcription factors*

Abstract

Chloroplast biogenesis is dependent on master regulators from the Golden2-like (GLK) family of transcription factors, but higher order *glk* mutants contain residual chlorophyll and therefore other proteins must also 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 mutants of *GLK*. In *M. polymorpha* MYB-related genes act upstream of *GLK*, while in *A. thaliana* this relationship has been rewired. In both species, genes encoding enzymes of chlorophyll biosynthesis are controlled by both MYB-related and GLK proteins whilst those allowing CO_2 fixation, photorespiration as well as photosystem assembly and repair require the MYB-related proteins. Thus, *MYB-related* and *GLK* genes have overlapping as well as distinct targets. We conclude that MYB-related transcription factors act with GLK to orchestrate chloroplast development in land plants.

Introduction

Photosynthesis is fundamental to life on earth and in photosynthetic eukaryotes this process takes place in organelles known as chloroplasts. It is widely accepted that chloroplasts originated from an endosymbiotic event that took place more than one billion years ago and involved a photosynthetic prokaryote and a eukaryotic cell (Archibald, 2009; Gould et al., 2008; McFadden, 2014). However, since that time significant elaborations in the control of photosynthesis gene expression have taken place. For example, in extant land plants the majority of genes allowing chloroplast biogenesis are encoded in the nuclear genome such that thousands are post-translationally imported from cytosol to chloroplast (Abdallah et al., 2000; Martin et al., 2002). Despite these significant rearrangements to the genetics underpinning photosynthesis, in most plants including major crops, the photosynthetic process has not been optimised by natural selection (Long et al., 2015; Ort et al., 2015). One current limitation to improvement of photosynthesis in crops is that our knowledge of the structure of the underlying gene regulatory networks is incomplete.

The expression of photosynthesis associated nuclear genes is responsive to light but also tuned by processes intrinsic to the cell. For example, in angiosperms light is required for chloroplast formation, and hormones including auxins, cytokinins, brassinosteroids and gibberellin amplify or repress this response (Cackett et al., 2022). These exogenous and endogenous inputs are integrated by key transcriptional regulators 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] (Cackett et al., 2022; Fitter et al., 2002; Naito et al., 2007; Waters et al., 2009; Zubo et al., 2018). However, multiple *glk* mutants in *Arabidopsis thaliana* (Waters et al., 2009), rice (Wang et al., 2013) and also non-seed plants such as the moss *Physcomitrium patens* (Yasumura et al., 2005) and the liverwort *Marchantia polymorpha* (Yelina et al., 2023) still contain chlorophyll. Moreover, mutants lacking both functional *GLK* and *GATA* genes are not albino (Chiang et al. 2012; Zubo et al., 2018). In summary, other actors must allow assembly of the photosynthetic apparatus in the absence of the known regulators.

Our incomplete understanding of the proteins controlling chloroplast biogenesis and therefore photosynthesis means re-engineering photosynthesis to improve crops is currently not predictable. We therefore sought to identify transcription factors that act with master regulators such as GLK to control chloroplast biogenesis. As forward genetics seems to have failed to identify such proteins, we rationalised that genetic redundancy had hindered their identification and that analysis of a species with a more compact genome would circumvent this issue. *M.*

polymorpha possesses a streamlined genome with many transcription factors represented by either one or two copies, and the dominant form of the lifecycle is haploid (Bowman et al., 2022). Moreover, the control of greening regulation is streamlined in *M. polymorpha* with for example only one copy of *GLK* being present, and orthologs of the *GATAs* that implicated in chloroplast biogenesis in other land plants (Schwechheimer et al., 2022) not being required (Yelina et al., 2023). We hypothesised that homologous transcription factors in *A. thaliana* and *M. polymorpha* act alongside *GLK* and so their expression should respond to light during photomorphogenesis. To test this, we examined publicly available RNA sequencing data from each species and selected transcription factors that responded to light. Plants with mutations in an RR-MYOBLASTOMA-related transcription factor (MpRR-MYB5, with a previously unknown role in chloroplast development and photosynthesis) showed a pale green phenotype when mutated. Chlorophyll levels were lower and chloroplasts significantly smaller than controls. Mutating the paralog Mp*RR-MYB2* exacerbated these phenotypes. Double mutants for Mp*RR-MYB5* and Mp*GLK* accumulated less chlorophyll than each single mutant, and null mutants for all three genes were lethal. In *M. polymorpha*, genetic analysis indicated that *MpRR-MYB2&5* act upstream of *GLK*, but in *A. thaliana* this relationship has been rewired. However, in contrast with *GLKs* that control the expression of genes allowing chlorophyll biosynthesis and function of photosystem I and II, in both species the RR-MYBs have a broader set of targets. For example, in addition to genes that support chlorophyll synthesis and the photosystems, the targets include genes associated with CO_2 fixation, photorespiration, photosystem assembly and repair. The RR-MYBs in *M. polymorpha* and *A. thaliana* show conservation in targets, and we conclude that these proteins likely function as master regulators of chloroplast biogenesis and photosynthesis gene expression across the land plants. 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, and the biosynthesis of amino acids, fatty acids and carotenoids.

Results

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

We interrogated publicly available gene expression data sampled during the transition from non-photosynthetic to photosynthetic growth in *M. polymorpha* (Flores-Sandoval et al., 2018) as well *A. thaliana* (Sullivan et al., 2014). This identified 108 and 144 transcription factors that were upregulated after exposure to light in *M. polymorpha* and *A. thaliana* respectively (Supplemental Table 1 and 2). We then selected orthologs upregulated in both datasets with an unknown or chlorophyll-related annotation, and that were represented by a multigene family in *A. thaliana* (Figure 1A). Fourteen candidates from *M. polymorpha* were identified (Figure 1B and Supplemental Table 2). Two of these (Mp*GLK,* Mp*GATA4*) are homologs of known photosynthesis regulators in *A. thaliana* (Bastakis et al., 2018; Waters et al., 2009) and Mp*GLK* has a confirmed role in *M. polymorpha* (Gangappa et al. 2013; Yelina et al. 2023). The remainder included a number of B-BOX (BBX) domain proteins known to interact with the master regulator of photomorphogenesis HY5 (Bursch et al., 2020), a homeobox-leucine zipper (HD-ZIP) protein *ATHB17* whose ortholog in *A. thaliana* regulates photosynthesis-associated nuclear genes in response to abiotic stress (Zhao et al., 2017), a *C2H2* type zinc finger transcription factor with unknown function in *A. thaliana,* and a *MYB-related* gene predicted to regulate photosynthesis gene expression (Tu et al., 2022; Halpape et al., 2023).

Taking advantage of the rapid and predominantly haploid life-cycle we used *M. polymorpha* as a testbed for each of these candidates and subjected each to CRISPR/Cas9-mediated editing. With the exception of Mp*GLK*, which has previously been reported to lead to a pale phenotype when mutated (Yelina et al., 2023) only one other candidate had low chlorophyll. This was Mp5g11830, annotated as Mp*RR-MYB5* in the *M. polymorpha* genome database*,* but previously also referred to as a CIRCADIAN CLOCK ASSOCIATED1-like RR-MYB-Related transcription factor (Du et al., 2013; Dubos et al., 2010). Mp*RR-MYB5* has a single paralog (Mp*RR-MYB2* - Figure 2A) that shows high similarity to Mp*RR-MYB5* at the amino acid level (Figure 2A) with for example the RR-MYB/CCA1-like domain being 92% identical (Figure 2A). Mutant alleles of Mp*RR-MYB5* but not Mp*RR-MYB2* appeared pale (Figure 2B, C, D, Supplemental Figure 1A-B) and analysis of chlorophyll content confirmed this (Figure 2F). All lines in which insertions or deletions introduced premature stop codons in Mp*RR-MYB5* (Supplemental Figure 1A) had 40-50% less chlorophyll than controls (Figure 2C and F). The Mp*rr-myb5* mutant was complemented when Mp*RR-MYB5* was expressed from its own promoter (Supplemental Figure 1B) confirming that the pale phenotype was unlikely associated with off-target CRISPR/Cas9 editing. Mutating

Mp*RR-MYB5* and Mp*RR-MYB2* simultaneously (Supplemental Figure 1C) led to extremely pale plants with chlorophyll content reduced to 95% compared with controls (Figure 2E and F, Supplemental Figure 2). To test whether the photosynthetic apparatus was functional in the single Mp*rr-myb5* and double Mp*rr-myb5,2* mutants we applied the inhibitor Di-Chlorophenyl Di-Methyl Urea (DCMU) that blocks photosynthetic electron transport (Trebst, 2007) and measured activity of photosystem II via chlorophyll fluorescence imaging. This showed that although these mutants had low levels of chlorophyll, the photosynthetic apparatus was operational (Figure 2G). Consistent with the very low chlorophyll levels in Mp*rr-myb5*, and Mp*rr-myb5,2* double mutants, chloroplasts were significantly smaller and thylakoids were underdeveloped (Figure 2H-P and Supplemental Figure 2). It was noticeable that poorly developed chloroplasts from double Mp*rr-myb5*,*2* mutants contained significant amounts of starch.

To determine whether Mp*RR-MYB5* and Mp*RR-MYB2* limit greening, we generated overexpression lines driven by the strong Mp*UBE2* constitutive promoter (Sauret-Güeto et al., 2020) and to facilitate analysis of chloroplast size per cell used GFP to mark the plasma membrane (Supplemental Figure 3A-E). Although quantitative polymerase chain reactions confirmed that each transgene was over-expressed (Supplemental Figure 3F-N) plants appeared similar to controls and there were no evident perturbations to chlorophyll content, chloroplast size or morphology (Supplemental Figure 3O-P). We conclude that Mp*RR-MYB5* and Mp*RR-MYB2* act redundantly and are necessary for chloroplast biogenesis but unlike Mp*GLK* (Yelina et al., 2023) they do not restrict it. Moreover, in the absence of both Mp*RR-MYB5* and Mp*RR-MYB2* assembly of the photosynthetic apparatus is very limited.

Mp*RR-MYB5&2* act with Mp*GLK* to control chloroplast biogenesis

As double Mp*rr-myb5*,*2* mutants showed residual chloroplast development and were viable, we hypothesised that the limited ability for photoautotrophic growth was associated with activity of the previously characterised master regulator *GLK*. To test this, we attempted to generate higher order mutants that combined mutant alleles of Mp*glk*, Mp*rr-myb5* and Mp*rr-myb2*. We were able to knock out Mp*RR-MYB5* in the presence of Mp*glk* mutant alleles (Supplemental Figure 4 and 5A). Such double mutants were paler than the single Mp*glk* mutant (Figure 3A-D) and contained less chlorophyll (Figure 3E). Application of DCMU confirmed that the photosystem II was functional in the Mp*glk*,*rr-myb5,2* double mutant (Figure 3F). Double Mp*glk,rr-myb5* mutants had smaller chloroplasts with fewer thylakoid membranes and reduced granal stacking compared with each single mutant (Figure 3G-O). Thus, in the absence of both Mp*RR-MYB5* and MpGLK, very limited assembly of the photosynthetic apparatus takes place.

We were unable to generate triple Mp*glk,rr-myb5,2* mutants implying that this allelic combination is lethal. For example, after super transforming Mp*glk,rr-myb5* double mutants with a vector allowing expression of the same guide RNA used to generate the single Mp*rr-myb2* mutants reported above, 91 lines were obtained. However, none were paler than the double Mp*glk,rr-myb5* mutant, and when genotyped 86 lines had no edits in MpRR-MYB2. Of the five lines that were edited in Mp*RR-MYB2* as well as Mp*GLK* and Mp*RR-MYB5* genotyping showed that the mutations had limited impact on the MpRR-MYB2 protein. For example, these edits altered one, two, three or seven amino acids in a poorly conserved region of the protein, and in all cases the reading frame was maintained (Supplemental Figure 5C and D). In contrast, when the original single Mp*rr-myb2* mutants were identified (Figure 2D) 50% of plants contained mutations that introduced early stop codons or disturbed the reading frame. We conclude that absence of all three proteins (MpGLK, MpRR-MYB5 and MpRR-MYB2) is lethal likely because chloroplast biogenesis is abolished.

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

To provide insight into the types of genes regulated by Mp*RR-MYB5* and Mp*RR-MYB2* we performed RNA sequencing of overexpressing lines as well as single and multiple mutants. Overexpression of Mp*RR-MYB2* and Mp*RR-MYB5* led to the upregulation of 71 and 11 genes respectively (*padj*-value ≤0.01, LFC ≥ 1-fold) (Supplemental Figure 6A and C) and there was limited overlap between these two datasets (Figure 4A and Supplemental Figure 7). This contrasts with overexpression of Mp*GLK* that led to the upregulation of 492 genes (Figure 4A and Yelina et al 2023).

In loss of function mutants for Mp*RR-MYB2* or Mp*RR-MYB5*, 65 and 823 genes, respectively, showed reductions in transcript abundance compared with controls (*padj*-value ≤0.01, LFC ≥ 1-fold) (Figure 4B and Supplemental Figure 6B and D). Knocking out Mp*GLK* had greater impact with 1065 genes being downregulated (Figure 4B and Yelina et al 2023). In double Mp*glk,rr-myb5* mutants, 1161 genes had lower transcript abundance than controls, and in the double Mp*rr-myb5,2* mutants this was increased to 1744 (Figure 4B). The largest overlap in changes to transcript abundance between genotypes (524 genes) was detected for Mp*glk,rr-myb5* and Mp*rr-myb5,2* mutants (Figure 4B). This finding further supports synergistic action of Mp*RR-MYBs* and Mp*GLK*. Gene Ontology (GO) terms were used to provide insight into the classes of genes impacted by overexpression or loss of function of Mp*RR-MYBs* and Mp*GLK*. Consistent with the lack of a detectable phenotype after overexpression of Mp*RR-MYB5* or Mp*RRMYB2,* or loss of

Mp*RR-MYB2* function, no distinct GO terms were impacted in these lines. Although the response to oxidative stress GO term was over-represented in both Mp*rr-myb5* and Mp*glk* mutants, other terms were distinct (Figure 4C and Supplemental Figure 7D). For example, Mp*rr-myb5* mutants showed changes to protein phosphorylation and peroxidase activity terms, whilst in Mp*glk* photosynthesis, light harvesting and chlorophyll biosynthesis terms were affected (Figure 4C and Yelina et al., 2023). It was notable that very similar GO terms responded in Mp*rr-myb5,2*, Mp*glk* and Mp*glk,rr-myb5* mutants (Figure 4C). For example, in all genotypes the top five biological processes impacted were response to oxidative stress, hydrogen peroxide catabolism, photosynthesis, light harvesting and chlorophyll biosynthesis (Figure 4C). Thus, loss of function alleles for Mp*rr-myb5,2,* Mp*glk* and Mp*glk,rr-myb5* all caused changes in GO terms that are primarily associated with photosynthesis.

Since chlorophyll content was reduced in Mp*rr-myb5* and Mp*rr-myb5,2* double mutants we examined impact on transcript abundance derived from genes associated with the nineteen annotated chlorophyll biosynthesis genes (Figure 4D). With the exception of *HEMA* gene in Mp*rr-myb5* mutants*,* knocking out either Mp*RR-MYB5* or Mp*RR-MYB2* did not significantly affect transcript abundance from chlorophyll biosynthesis genes. In contrast, in Mp*glk* mutant alleles transcript abundance from seventeen chlorophyll biosynthesis genes was reduced, and in the Mp*rr-myb5,2* double mutant all nineteen genes were significantly downregulated (Figure 4D). We next examined the impact of loss of the Mp*RR-MYBs* on approximately 200 other genes annotated (Supplemental Dataset 1) as photosynthesis related in *M. polymorpha*. This group included genes associated with CO_2 fixation and the light harvesting apparatus as well as their assembly and repair. In the single Mp*rr-myb5* and Mp*rr-myb2* mutant alleles there was limited effect on photosynthesis associated genes (Figure 5A and B). For example, in Mp*rr-myb2* expression of only a single photosynthesis gene (petE Mp4g02720) was perturbed (Supplemental Dataset 2). In Mp*rr-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 5A and Supplemental Dataset 2). As expected, changes to photosynthesis transcripts were more evident in the Mp*glk* mutant (Figure 5A and B), and even more severe when both Mp*RR-MYB5* and Mp*GLK* were mutated (Figure 5A and B). Strikingly, when Mp*RR-MYB2* and Mp*RR-MYB5* were simultaneously knocked out, the effect on photosynthesis associated genes was extensive and more widespread than in the *Mpglk,rr-myb5* double mutant. For example, in Mp*rr-myb5,2* double mutants, the majority of genes encoding enzymes involved in the Calvin Benson Bassham cycle and photorespiration were downregulated (Figure 5A and Supplemental Figure 8A). Moreover, genes encoding components of both

photosystems and their respective light harvesting complexes as well as the Cytochrome *b_6f* complex were downregulated (Figure 5B). We also found that genes associated with assembly of RuBisCO, non-photochemical quenching, as well as granal stacking and repair of photosystem II were impacted in Mp*rr-myb5,2* double mutants (Supplemental Figure 8B). This contrasts with Mp*glk* in which only genes encoding enzymes of chlorophyll biosynthesis as well as components of the photosystems and their light harvesting complexes were mis-regulated (Figure 5B).

MpRR-MYB transcription factors condition expression of *GLK*

Consistent with the alterations in transcript abundance described above, consensus binding sites for the GLK and RR-MYBs proteins (Figure 6A) were found in promoters of 69% of photosynthesis genes. 24% and 27% possessed motifs associated with GLK or RR-MYB binding respectively, and 18% contained both motifs (Figure 6B and Dataset 3). For example, genes encoding RuBisCO Activase, Phosphoglycolate Phosphatase and a component of Light Harvesting Complex II were predicted to contain RR-MYB as well as GLK binding sites (Figure 6C). Overall, these data are consistent with overlapping as well as distinct roles for these two classes of transcription factor. Our data also support the notion that in contrast with Mp*GLK*, the Mp*RR-MYBs* activate genes allowing CO_2 fixation as well as light harvesting.

To better understand this interplay between Mp*RR-MYB2,* Mp*RR-MYB5* and Mp*GLK* we first examined the expression of Mp*RR-MYB2&5* in the *Mpglk* mutant background. Mp*RR-MYB5* was upregulated in Mp*glk* (Figure 6D) indicating that either Mp*GLK* suppresses Mp*RR-MYB5* or that Mp*RR-MYB5* is upregulated in the absence of a functional MpGLK protein. Similarly, Mp*RR-MYB2* was upregulated in the absence of MpGLK. In contrast, in the Mp*rr-myb5,2* double mutant Mp*GLK* was downregulated (Figure 6E). We were not able to identify any strong binding motifs for GLK in either Mp*RR-MYB5* or Mp*RR-MYB2* but in the promoter of Mp*GLK* consensus binding sites for *RR-MYBs* were clearly evident (Figure 6F). The promoters of Mp*RR-MYB5* and Mp*GLK* each contained their own binding site indicating the potential for self-regulation.

We next tested the extent to which MpRR-MYB2, MpRR-MYB5 or MpGLK transcription factors could rescue the pale phenotype of single Mp*rr-myb5,* Mp*glk,* and double Mp*glk,rr-myb5* or Mp*rr-myb5,2* mutants (Figure 7A-F). Quantitative polymerase chain reactions confirmed that each transgene was over-expressed (Supplemental Figure 9). Both Mp*RR-MYB5* and Mp*RR-MYB2* complemented Mp*rr-myb5* mutants (Figure 7B, C and G). This further argues for functional redundancy between MpRR-MYB5 and MpRR-MYB2. However, neither Mp*RR-MYB* rescued single Mp*glk* or double Mp*glk,rr-myb5* mutants (Figure 7B, D, E and G). When Mp*GLK* was expressed in the Mp*rr-myb5,2* double mutant background chlorophyll levels were increased by

~10% but the absolute levels were still 90% lower than wild type (Figure 7F and H), and consistent with functional redundancy between MpRR-MYB5 and MpRR-MYB2, we also found that Mp*GLK* partially complemented Mp*rr-myb5* single mutants Figure 7C and G).

A conserved role for RR-MYBs in *Arabidopsis thaliana*

The RR-MYB/CCA1-like subfamily of MYB-related transcription factors (Du et al., 2013; Pu et al., 2020; Yanhui et al., 2006) containing Mp*RR-MYB5* and Mp*RR-MYB2* are characterised by a conserved SHAQK(Y/F)F motif (Supplemental Figure 10A). Based on phylogenetic analysis we identified eleven members of this group in *A. thaliana* (Figure 8A, Supplemental Figure 10B and C) of which At*MYBS1,* At*MYBS2* and AT5G23650 were the closest homologs of Mp*RR-MYB5* and Mp*RR-MYB2* (Supplemental Figure 11-12). Re-analysis of publicly available data indicated that AT5G23650 is expressed in photosynthetic tissues and so we focused analysis on At*MYBS1* and At*MYBS2*. Due to the functional redundancy evident for Mp*RR-MYB5* and Mp*RR-MYB2* above, double At*mybs1,mybs2* mutants were identified after CRISPR/Cas9-mediated gene editing (Supplemental Figure 13) and analysed in parallel with previously generated single At*mybs1* and At*mybs2* mutants. There were no detectable changes to rosette phenotype in the single mutants but double At*mybs1,mybs2* mutants were pale (Figure 8C-F), and this was most noticeable after bolting (Figure 8G). Confocal laser scanning microscopy revealed no detectable changes in chloroplast size or number in mesophyll cells of single mutants, but chloroplasts were smaller in the double At*mybs1,2* mutant (Figure 8F). In support of these findings, chlorophyll content in the single mutants was indistinguishable from wild-type, but was ~40% lower in the double At*mybs1,mybs2* mutant (Figure 8H), and quantification of chloroplast size demonstrated a 50% reduction in the double mutants (Figure 8I).

To gain insight into the types of photosynthesis genes regulated by At*MYBS1* and At*MYBS2* we performed RNA sequencing of the double mutant. 841 genes were down-regulated (*padj*-value ≤0.01, LFC ≥ 1-fold) and GO term analysis indicated that the two top biological terms impacted were response to light stimulus and photosynthesis (Supplemental Figure 14A and B). Thus, consistent with lower levels of chlorophyll, loss of *MYBS1* and *MYBS2* in *A. thaliana* led to down-regulation of transcripts associated with photosynthesis. Notably, promoters of the At*MYBS1&2* transcription factors in *A. thaliana* contained motifs recognised by GLK as well as RR-MYBs (Figure 9A) implying regulatory interplay. Although the promoter of *GLK1* in *A. thaliana* has a motif associated with GLK binding (Figure 9A), in contrast with *M. polymorpha,* neither At*GLK1* or At*GLK2* contained any binding sites for MYBS1 or MYBS2 (Figure 9A). Promoters of the *CGA1* and *GNC* transcription factors that regulate photosynthesis gene expression in *A.*

thaliana (Bastakis et al., 2018; Zubo et al., 2018) also contained GLK but not MYBS predicted binding sites (Figure 9A). Analysis of *GLK*, *CGA1* and *GNC* transcript abundance in the At*mybs1,2* indicated no change in *GLK1* and *CGA1* but increases in *GLK2* and *GNC* (Figure 9B). This implies that cryptic or more distant MYBS1&2 binding sites exist for *GLK2* and *GNC,* or that indirect regulation links expression of these transcription factors.

Despite these changes to the upstream regulatory network, multiple lines of evidence point to a role for At*MYBS1&2* in the control of photosynthesis gene expression in *A. thaliana*. For example, the promoters from 16% and 27% of photosynthesis associated genes contained predicted MYBS1&2 and GLK binding sites respectively (Figure 9C and Supplemental Dataset 3), and 15% contained both (Figure 9C and Supplemental Dataset 3). Thus, in both *M. polymorpha* and *A. thaliana* the majority of photosynthesis genes contain *cis*-elements predicted to be bound by these regulators. Promoters with RR-MYB motifs included components of the Calvin Benson Bassham cycle, photorespiration and light harvesting complexes (Figure 9D). Moreover, abundance of transcripts associated with chlorophyll biosynthesis, both photosystems and their light harvesting apparatus, as well as the Calvin Benson Bassham cycle were perturbed when At*MYBS1&2* were knocked out in *A. thaliana* (Figure 9E and Supplemental Figure 14C).

Discussion

Due to the importance of chloroplasts in photosynthesis, nitrogen and sulphur assimilation, as well as the biosynthesis of amino acids, fatty acids and carotenoids, understanding the biogenesis of these organelles has long been of interest (Jarvis & López-Juez, 2013; López-Juez, 2007; Wang et al., 2017; Waters & Langdale, 2009; Witte & Herde, 2020). Moreover, it is now widely recognised that photosynthesis has not been optimised by evolution, and so a targeted re-engineering of the process could contribute to crop development (Long et al., 2015; Ort et al., 2015). Indeed, significant improvements in yield have been reported after over-expression of *SEDOHEPTULOSE BISPHOSPHATASE* (Driever et al., 2017; Lefebvre et al., 2005), faster relaxation of non-photochemical quenching of photosystem II (Kromdijk et al., 2016) and re-routing photorespiratory metabolism (South et al., 2019). A complementary approach to improving photosynthesis that is predicted to increase yield by up to 50% would be to convert C_3 crops to use the more efficient C_4 photosynthetic pathway (Hibberd et al., 2008; von Caemmerer et al., 2012). Introducing C_4 photosynthesis in C_3 crops such as rice would require a remodelling of chloroplast biogenesis in mesophyll and bundle sheath cells (Langdale, 2011).

In land plants, the GLK family of transcription factors are known as master regulators of chloroplast biogenesis with CGA1 and GNC from the GATA family considered ancillary players (Bastakis et al., 2018; Cackett et al., 2022; Waters et al., 2009). 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 C_4 pathway (Wang et al., 2017). However, we have an incomplete understanding of the transcription factors allowing chloroplast development. For example, *GLK* and *CGA1/GNC* loss of function mutants still possess small chloroplasts (Zubo et al., 2018) 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 of these transcription factors also control chloroplast biogenesis in *A. thaliana* indicating functional conservation between these distantly related species. One interesting output of the analysis reported here is that we were unable to identify null mutants lacking both GLK and RR-MYB function in *M. polymorpha*. This apparent lethality of triple Mp*glk,rr-myb5,2* mutants mirrors *A. thaliana* loss-of-function mutations in genes involved in plastidial pathways such as amino acid, vitamin, nucleotide or fatty acid biosynthesis, as well as those involved in protein translation in the chloroplast that result in an arrest of embryo development (Bryant et al., 2011; Budziszewski et al., 2001). Interestingly, this often appears to take place at the globular-to-heart transition stage

of embryogenesis when chloroplasts start to differentiate in the ground tissue of the cotyledons. Moreover, mutants in genes encoding plastidial proteins that are required for the import, modification and localisation of indispensable proteins within the chloroplast are also often associated with embryo lethality (Bryant et al., 2011; Budziszewski et al., 2001).

The importance of these RR-MYBs from *M. polymorpha* and *A. thaliana* in controlling chloroplast biogenesis is supported by previous work in other systems. For example, the tomato RR-MYB-related transcription factor LeMYBI has been reported to bind the promoter of the *RBCS* gene (Rose et al., 1999). And, although no effect on chlorophyll content of chloroplast biogenesis was reported, a reduction in the expression of *RBCS* and the *chlorophyll a/b binding protein 1* (*CAB1*) genes has been reported in At*mybs1* mutants (Chen et al., 2017). Along with transcription factors belonging to the GLK, B-Box and Nuclear Factor-Y families, random forest analysis of gene expression in *A. thaliana* and wheat recently predicted that RR-MYBs regulate photosynthesis gene expression (Halpape et al., 2023). Moreover, RNA sequencing of an inducible At*MYBS1* over-expressor line showed upregulation of genes that we also detected as downregulated in the At*mybs1,mybs2* mutant alleles. Based on these findings we conclude that the RR-MYB class of transcription factors likely control photosynthesis in a wide range of land plants.

Most land plants contain multiple *GLK* genes. With the exception of species such as maize in which neofunctionalization of distinct copies of *GLK* allows differential control of chloroplast biogenesis in the mesophyll and bundle sheath (Fitter et al., 2002; Hall et al., 1998), *GLK* genes act redundantly (Waters et al., 2008, 2009). In the liverwort *M. polymorpha* that contains only one *GLK* gene such that redundancy of function is not possible, our analysis indicates that there is instead redundancy over the control of photosynthesis because both the two RR-MYBs and GLK have overlapping functions. In *M. polymorpha* and *A. thaliana* double mutants are required to generate strong phenotypic perturbations to chloroplast biogenesis. For example, the single Mp*rr-myb5* mutant is pale, and this phenotype is exacerbated when combined with loss of either MpRR-MYB2 or MpGLK. In *A. thaliana* neither A*tmybs1* nor At*mybs2*mutants appear to have less chlorophyll, but the double mutant is pale. We infer that in both species the two RR-MYB proteins act redundantly and synergistically to control chlorophyll biosynthesis. Given the importance of being able to develop chloroplasts in response to light, it is perhaps not surprising that multiple regulators exist to ensure a robust response.

As well as this redundancy between RR-MYBs and GLK, our data argue for RR-MYBs targeting a broader range of photosynthesis genes than those previously associated with GLK. Thus, the RR-MYBs and GLK act in a genetically separable manner. In fact, our analysis indicated that

penetrance of the RR-MYBs over photosynthesis gene expression and chloroplast biogenesis in *M. polymorpha* is particularly striking. For example, chloroplasts of Mp*myb5,2* mutants were ~30% smaller than those of Mp*glk* mutants, and double Mp*myb5,2* mutants were paler than those of Mp*glk*. The underlying reason for this appears to be because the RR-MYBs have overlap with GLK in terms of targets, but also that they control a broader set of photosynthesis genes. Shared targets is supported by previous work that proposed co-binding of RR-MYB and GLK to photosynthesis genes (Tu et al., 2020). Such cooperative binding of transcription factors is thought to allow greater variety of gene expression outputs (Ibarra et al., 2020; Sönmezer et al., 2021). The reach of these RR-MYBs appears extensive in that they control genes encoding enzymes participating in the Calvin Benson Bassham cycle and photorespiration but also in assembly and repair of RuBisCO and photosystem II. 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. And, when combined with loss of GLK this may be lethality ensues. Overall, the data are consistent with overlapping as well as distinct roles for these two classes of transcription factor that likely generates a robust regulatory system with inbuilt redundancy. Our data also support the notion that in contrast with Mp*GLK*, the Mp*RR-MYBs* activate genes allowing CO_2 fixation as well as light harvesting.

The precise architecture of the gene regulatory network involving the *RR-MYBs* and *GLK* cannot currently be defined. In *M. polymorpha* there is some evidence that Mp*RR-MYB5&2* act upstream of Mp*GLK*. For example, *MpGLK* contains predicted binding sites for the MpRR-MYBs, and in Mp*rr-myb2,5* mutant alleles transcripts derived from Mp*GLK* are less abundant. Moreover, when the Mp*RR-MYB*s were overexpressed in the Mp*glk* mutant it remained pale, arguing against the RR-MYBs simply acting downstream of GLK. The stronger and broader down regulation of photosynthesis transcripts in the double Mp*rr-myb5,2* mutant compared with Mp*glk* could be due to the RR-MYBs acting upstream of GLK. Alternatively, the larger response of photosynthesis transcripts in Mp*rr-myb5,2* mutant compared with Mp*glk* could be because the RR-MYBs also have a broader set of targets. Although Mp*RR-MYB5* overexpression failed to upregulate Mp*GLK* transcripts, MYB transcription factors commonly act in multimeric complexes involving basic helix–loop–helix (bHLH) and WD40 repeat proteins (Ramsay & Glover, 2005), or with MYCs (Dickinson et al., 2020). It is therefore possible that additional partners need to be over-expressed in combination with the RR-MYBs in order to increase expression of Mp*GLK*. It is also well documented that *GLK* is subject to multiple levels of regulation that are most easily overcome when non-native versions of the gene are mis-expressed (Wang et al., 2017). Thus, it may be

that the lack of response in Mp*GLK* transcripts after over-expression of the Mp*RR-MYB*s is due to a similar regulatory system in *M. polymorpha*. When Mp*GLK* was mis-expressed in the double Mp*rr-myb5,2* mutant although chlorophyll content remained low (90% of wild type) it was significantly statistically increased. We interpret these data in two ways. Either Mp*GLK* is downstream of the Mp*RR-MYBs* and both classes of transcription factor are needed to drive full photosynthesis gene expression, or Mp*GLK* is permissive for very early stages of chloroplast biogenesis, but full assembly of the photosynthetic apparatus is strengthened by the RR-MYBs because they both positively regulate Mp*GLK* itself, but they also control the expression of photosynthesis genes allowing carbon fixation as well as light harvesting. A permissive role for GLK in initiating chloroplast biogenesis is consistent with its ability to convert cell types that normally accumulate no or a few chloroplasts to having a much larger chloroplast compartment (Wang et al., 2017). Based on all the above, our current favoured hypothesis is that the RR-MYBs condition expression of *GLK* to permit early stages of chloroplast biogenesis. Although this conditioning can partially be overcome when Mp*GLK* is over-expressed in the presence of RR-MYBs such that increased chlorophyll and larger chloroplasts develop (Yelina et al., 2013), in the absence of Mp*RR-MYBs* the impact of Mp*GLK* over-expression is much more limited. Thus, the interplay between the Mp*RR-MYBs* and Mp*GLK* allow full and effective assembly of the photosynthetic apparatus.

In summary, from functional analysis of *M. polymorpha* and *A. thaliana* whose last common ancestor diverged around 400 million years ago, we propose a model in which both the RR-MYBs and GLKs operate as master regulators of photosynthesis gene expression in land plants (Figure 10). Our data show that in both species the RR-MYBs play a conserved role in controlling photosynthesis gene expression. Their targets are broader than those documented for GLK, such that components allowing carbon fixation and photorespiration as well as assembly and repair of the photosystems can be activated in response to light (Figure 10). In *M. polymorpha* both the RR-MYB and GLK appear to regulate themselves, and the RR-MYBs condition expression of GLK (Figure 10). In *A. thaliana* this regulatory system has been rewired and is more complex. For example, regulation of At*GLK* by AtMYBS1&2 is less evident, but members of the GATA family (Bastakis et al., 2018; Zubo et al., 2018) also control photosynthesis gene expression. Despite evidence for an evolutionary rewiring of the targets of these transcription factors between *A. thaliana* and *M. polymorpha*, we propose that in land plants both the RR-MYBs and GLK are key actors in chloroplast biogenesis.

Despite these similarities between *A. thaliana* and *M. polymorpha* it is clear that the exact wiring of these transcription factors to each other and also to structural photosynthesis genes

differs. For example, inducible overexpression of At*MYBS1* in *A. thaliana* has been reported to increase expression of some photosynthesis genes (Halpape et al. 2023). Although similar sets of genes were down-regulated in loss of function Mp*rr-myb5,2* mutants, we did not detect widespread upregulation of photosynthesis genes after over-expression in *M. polymorpha.* Also consistent with rewiring between these species is the fact that compared with Mp*RR-MYB2*, the pale phenotype of Mp*RR-MYB5* indicates it plays a dominant role in chloroplast biogenesis while in *A. thaliana* neither single mutant was pale. It is not clear if this dominance of one paralog was gained in *M. polymorpha* or lost from *A. thaliana*. Lastly, changes to the structure of the gene regulatory network associated with photosynthesis is also supported by the fact that promoters of Mp*RR-MYB5&2* contain their own binding sites, but Mp*GLK* does not. In contrast, in *A. thaliana* the promoters of At*MYBS1* and At*MYBS2* possess the cognate *cis*-elements for both GLKs and the RR-MYBs. The output is that the RR-MYBs operate upstream of GLK in *M. polymorpha* but not in *A. thaliana*. Taken together these data support a model in which rewiring of the gene regulatory networks underpinning chloroplast biogenesis has taken place between *A. thaliana* and *M. polymorpha*.

As RR-MYBs appear to be ubiquitous in land plants (Du et al., 2013) and the angiosperms and bryophytes diverged around 400 million years ago it seems possible that RR-MYBs play a conserved and ancestral role in chloroplast biogenesis of land plants. Moreover, while we were not able to detect homologs of Mp*RR-MYBs* in Zygnematophyceae algae (Cheng et al., 2019) that is sister to the land plants, they are present in the Klebsormidiophyceae (*Klebsormidium flaccidum*) and Charophyceae (*Chara braunii*) (Hori et al., 2014; Nishiyama et al., 2018) (Supplemental Fig. 10). *GLK* homologs are present in green algae (Cheng et al., 2019) and both GLK and RR-MYB motifs are present in promoter regions of these genes in *K. flaccidum* and *C. braunii* (Supplemental Figure 16). These data imply that *RR-MYBs* operated alongside GLK to control chloroplast biogenesis before the colonisation of land by plants.

Material & Methods

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

Marchantia polymorpha accessions Cam-1 (male) and Cam-2 (female) were used. Plants were grown on half-strength Gamborg B5 medium plus vitamins (Duchefa Biochemie G0210, pH 5.8) and 1.2% (w/v) agar (Melford capsules, A20021), under continuous light at 22 °C with light intensity of 100 µmol $m^{-2}s^{-1}$. Transgenic *M. polymorpha* plants were obtained following an established protocol. All analysis of *Arabidopsis* thaliana used C*ol-0* expect the At*mybs2*mutant that was in *Col-3* background, and plants were grown on F2 soil (Levington, F20117800), under 16 hours light followed by 8 hours dark, at 20 °C, 60% humidity, with light intensity of 150 µmol $m^{-2}s^{-1}$. *Arabidopsis* T-DNA insertion mutants At*mybs1* (SAIL_1184_D04) and At*mybs2* (SALK_150774) were obtained from NASC and the presence of T-DNA was confirmed by PCR (Supplemental Figure 15 and Supplemental Table 3), as well homozygosity/heterozygosity. SAIL_1184_D04 and SALK_150774 were segregating, but all plants looked phenotypically similar to wild type.

For CRISPR/Cas9 gene editing, guide RNAs were *in silico* predicted using CasFinder tool (https://marchantia.info/tools/casfinder/). Several gRNAs per target gene were *in vitro* tested as described (Yelina et al., 2022) using oligonucleotides reported in Supplemental Table 3. gRNA sequences selected are listed in Supplemental Table 3. Single gRNAs were cloned using the OpenPlant toolkit (Sauret-Güeto et al., 2020) and the destination vector pMpGE013 (Sugano et al., 2018). For overexpression constructs, Mp*RR-MYB2* and Mp*RR-MYB5* coding sequences were synthesised (Integrated DNA Technologies) and cloned into the pUAP4 vector (Sauret-Güeto et al. 2020) with the OpenPlant toolkit (Sauret-Güeto et al., 2020). OpenPlant parts used: OP-023 CDS12-eGFP, OP-020 CDS_hph, OP-037 CTAG_Lti6b, OP-054 3TER, _Nos-35S, OP-049 PROM_35S, OP-47 PROM_MpUBE2, OP-48 5UTR_MpUBE2, OP-063, L1_HyR-Ck1, OP073, OP-062 L1_CsR-Ck1, L1_Cas9-Ck4, OP-076 L2_lacZgRNA-Cas9-CsA and L1_35S_s::eGFP-Lti6b (Frangedakis et al., 2021).

To generate *A. thaliana* MYBS1/MYBS2 mutants two gRNAs for each gene were cloned into the pEn-Chimera vector (Fauser et al., 2014) using a modification of the gRNA-tRNA cloning approach (Xie et al., 2015) (Supplemental Figure 17) followed by a Gateway LR reaction. This placed the guides into the pRU294 vector that has a codon optimised and intron-containing version of *Cas9* for *Zea mays* (*zCas9i*) driven by the egg-cell specific *pEC1.2* promoter (Ursache et al., 2021). *A. thaliana* was transformed by floral dipping according to (Logemann et al., 2006).

A. thaliana genotyping was performed as reported previously (Edwards et al., 1991) T2 plants with confirmed edits were analysed (Supplemental Figure 13). For gemmae transformation, 5 mL LB media were inoculated with 3-4 *Agrobacterium* colonies (*GV3101:* 50 µg/mL rifampicin, 25 µg/mL gentamicin) and the plasmid-specific selection antibiotic. The preculture was incubated at 28°C for 2 days at 110 rpm. 5 mL of 2 d *Agrobacterium* culture were centrifuged for 7 min at 2000 x *g*. The supernatant was removed and pellet re-suspended in 5 mL liquid KNOP (0.25g/L KH_2PO_4, 25g/L KCl, 25g/L $MgSO_4$ $7H_2O$, 100g/L $Ca(NO_3)_2$ $4H_2O$, 12.5mg $FeSO_4 7H_2O$, 30mM MES and pH5.5) plus 1% (w/v) sucrose and 100 µM acetosyringone. The culture was then incubated with shaking (120 rpm) at 28°C for 3-4 hours. Around 100 gemmae were transferred into a 6-well plate with 5 mL of liquid KNOP medium supplemented with 1% (w/v) sucrose and 30 mM MES, pH 5.5, 80 µL of *Agrobacterium* culture and acetosyringone at final concentration of 100 µM. The tissue was co-cultivated with the *Agrobacterium* for 3 days on a shaker at 110 rpm, at 22°C with only ambient light. Using a sterile plastic pipette, the liquid was removed from the well and gemmae then transferred using sterile scalpels onto plates with growth media containing the appropriate antibiotic (Chlorsulfuron 0.5 µM). To facilitate spreading of the gemmae, 1-2 mL of sterile water was added to the petri dish. To genotype *M. polymorpha* small pieces (3 × 3 mm) of thalli from individual plants were placed in 1.5 mL Eppendorf tubes and crushed with an autoclaved micropestle in 100 µL genotyping buffer (100 mM Tris-HCl, 1 M KCl, 1 M KCl, and 10 mM EDTA, pH 9.5). Tubes were then placed at 70 °C for 15-20 min. 380 µL of sterile water was added to each tube. 5 µL aliquots of the extract were used as a template for polymerase chain reactions and products checked on a 1.5% (w/v) agarose gel.

Chlorophyll determination, fluorescence measurements and imaging analysis

For chlorophyll measurements of *M. polymorpha* ~30-50mg of 10-14 days old gemmalings were used with five biological replicates per genotype. The tissue was blotted on tissue paper before weighing to remove excess water and then 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 with 300 rpm shaking 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) and the image processing software provided by the manufacturer (Schreier et al., n.d.). *M. polymorpha* plants were placed in the dark for 20 mins and a minimum weak measuring light beam (<1 µmol m^{-2} s^{-1}) was applied to evaluate the dark-adapted minimum

fluorescence (F_o), and a subsequent saturating pulse of 3000 µmol m^{-2} s^{-1} was applied to evaluate dark-adapted maximum fluorescence (F_m), and then variable fluorescence F_v *was calculated according to the following formula:* $F_v = F_m – F_o$. All chlorophyll fluorescence images within each experiment were acquired at the same time in a single image, with a total of three plants measured per genotype and treatment. 20 µM DCMU (#45463, Sigma Aldrich) was added to the half-strength MS media before being poured into the individual petri dishes. Thalli were placed in DCMU containing media for 24 h before chlorophyll fluorescence measurements were obtained.

For confocal laser scanning microscopy, a gene frame (#AB0576, ThermoFisher) was positioned onto a glass slide. Five to seven gemma were placed within the medium-filled gene frame together with 30 µL of milliQ water and the frame then sealed with a cover slip. Plants were imaged immediately using a Leica SP8X spectral fluorescent confocal microscope. Imaging was conducted using either a 10× air objective (HC PL APO 10×/0.40 CS2) or a 20× air objective (HC PL APO 20×/0.75 CS2). Excitation laser wavelength and captured emitted fluorescence wavelength window were as follows: for eGFP (488 nm, 498−516 nm) and for chlorophyll autofluorescence (488 or 515nm, 670−700 nm). For electron microscopy ~2 mm^2 sections of 5-6 individual 3-week-old *M. polymorpha* thalli from each genotype were harvested, fixed, embedded and imaged using scanning electron microscopy as previously described (Schreier et al. 2022). Chloroplast area was measured using ImageJ and the Macro in Supplemental Information.

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

For *M. polymorpha*, RNA was extracted from 3-4 two-week old gemmae using the RNeasy Plant kit (#74903, Qiagen) according to the manufacturer's protocol (RLT buffer supplemented with beta-merchaptoethanol was used) and residual genomic DNA was removed using the Turbo DNA-free kit (# AM1907, Invitrogen). 500 ng of DNase-treated RNA was used as template for cDNA preparation using the SuperScript™ IV First-Strand Synthesis System (#18091050, Invitrogen) according to manufacturer's instructions with the exception that reverse transcriptase reaction was modified to 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 (Supplemental Tables 3) were used at a final concentration of 0.5 µM. Reaction conditions were as follows: 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 Supplemental Table 3. Library preparation and RNA sequencing was performed by Novogene (Cambridge, UK). Briefly, messenger RNA was purified from total RNA using poly-T oligo-

attached magnetic beads. After fragmentation, the first strand cDNA was synthesised using random hexamer primers followed by second strand cDNA synthesis. cDNA end repair, A-tailing, adapter ligation, size selection, amplification, and purification were then performed. Library concentration was measured on a Qubit instrument following the manufacturer's procedure (Thermo Fisher Scientific) followed by real-time qPCR quantification. Library size distribution was analysed on a bioanalyzer (Agilent) following the manufacturer's protocol. Quantified libraries were pooled and sequenced on a NovaSeq PE150 Illumina platform and 6G raw data per sample obtained. Adapter sequences were: 5' Adapter: 5'-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-3'. 3' Adapter: 5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGATGACTATCTCGTATGCCGTCTTCTGCTTG-3'

FastQC was used to assess read quality and TrimGalore (https://doi.org/10.5281/zenodo.5127899) to trim low-quality reads and remaining sequencing adapters. Reads were pseudo-aligned using Kallisto (Bray et al., 2016) to the *M. polymorpha* Genome version 5 (primary transcripts only, obtained from MarpolBase) (Bowman et al., 2017). Kallisto estimates the abundance of each transcript in units of transcripts per million (TPM). Mapping statistics for each library are provided in Supplemental Dataset 2. DGE analysis was performed with DESeq2 (Love et al., 2014), with padj-values < 0.01, Plots were generated using R.

Phylogenetic analysis

To identify RR Myb-related/CCA1-like genes three different approaches were combined. Firstly, the RR Myb-related/CCA1-like genes for twenty-one plant genomes were mined from the iTAK (Zheng et al., 2016) and PlantTFDB databases (Jin et al., 2017), Phytozome, Fernbase (Li et al., 2018), Phycozome and PhytoPlaza. Sequences for each individual species were aligned with the MpRR-MYB5 and *A. thaliana* MYBS1 and MYBS2 amino acid sequences using MAFFT (Katoh & Standley, 2013). 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, we performed BLASTP searches against plant genomes on Phytozome v13, fern genomes (fernbase.org), hornworts genome (www.hornworts.uzh.ch) (Li et al., 2020), green algae genomes on PhycoCosm (/phycocosm.jgi.doe.gov), and 1KP using the MpRR-MYB5 and *A. thaliana* MYBS1 and MYBS2 amino acid amino acid sequence as queries. Results were filtered manually as described above. The identified RR-Myb-related/CCA1-like

protein sequences were aligned using MAFFT. Alignments were then trimmed using TrimAl (Capella-Gutiérrez et al., 2009). A maximum likelihood phylogenetic tree was inferred using iQTree (Nguyen et al., 2015), ModelFinder (Kalyaanamoorthy et al., 2017) and ultrafast approximation for phylogenetic bootstrap (Hoang et al., 2020) and SH-aLRT test (Guindon et al., 2010). The tree was visualised using iTOL (Letunic & Bork, 2019). Full list of sequences in Supplemental Dataset 4.

RR-MYB and GLK binding site analysis

AtGLK1 and GLK2 transcription factor binding motifs were taken from previously reported ChIP-seq data (Tu et al., 2022). AtMYBS1 and MYBS2 binding sites were obtained from the JASPAR motif database (Fornes et al., 2020) (motifs MA1186.1 and MA1399.1 respectively). GLK and MYBS motifs were merged to create GLK combined and RR-MYB combined motifs and these were visualised using the Ceqlogo tool from the MEME suite (Bailey et al., 2009). The FIMO tool (Grant et al., 2011) was used to scan promoter sequences defined as 1500 bp upstream of the transcriptional start site (TSS), of *A. thaliana* and *M. polymorpha GLK, RR-MYB* and other photosynthesis related transcription factors for matches to transcription factor binding motifs found in the JASPAR motif database (Fornes et al., 2020). To account for input sequence composition, a background model was generated using the fasta-get-markov tool from the MEME suite (Bailey et al., 2009). FIMO was then run with the 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. To assess occurrence of GLK and RR-MYB motifs in photosynthesis associated genes promoter sequences defined as 1500 bp upstream of the TSS from *A. thaliana* and *M. polymorpha* were scanned using the FIMO tool (Grant et al., 2011) as described above. Promoters were then scored for presence or absence of these motifs, and percentage of photosynthesis of genes containing neither motif, GLK only, RR-MYB only or both motifs calculated.

Acknowledgements

This work was funded as part of the BBSRC/EPSRC OpenPlant Synthetic Biology Research Centre Grant BB/ L014130/1 to J.P.H., BBSRC BB/F011458/1 for confocal microscopy to J.P.H and BBSRC sLOLA BBP0031171 to J.M.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. We thank Karin H. Müller, Filomena Gallo and Georgina E. Lindop from the Cambridge Advanced Imaging Centre for the electron microscopy sample preparation as well as the support during the image acquisition.

REFERENCES

Abdallah, F., Salamini, F., & Leister, D. (2000). A prediction of the size and evolutionary origin of the proteome of chloroplasts of Arabidopsis. *Trends in Plant Science*, *5*(4), 141–142.

Archibald, J. M. (2009). The Puzzle of Plastid Evolution. In *Current Biology* 19(2), 81-88.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, *37*, 202–208.

Bastakis, E., Hedtke, B., Klermund, C., Grimm, B., & Schwechheimer, C. (2018). LLM-Domain B-GATA Transcription Factors Play Multifaceted Roles in Controlling Greening in Arabidopsis. *The Plant Cell*, *30*(3), 582–599.

Bowman, J. L., Arteaga-Vazquez, M., Berger, F., Briginshaw, L. N., Carella, P., Aguilar-Cruz, A., Davies, K. M., Dierschke, T., Dolan, L., Dorantes-Acosta, A. E., Fisher, T. J., Flores-Sandoval, E., Futagami, K., Ishizaki, K., Jibran, R., Kanazawa, T., Kato, H., Kohchi, T., Levins, J., … Zachgo, S. (2022). The renaissance and enlightenment of Marchantia as a model system. *The Plant Cell*, *34*(10), 3512–3542.

Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., Adam, C., Aki, S. S., Althoff, F., Araki, T., Arteaga-Vazquez, M. A., Balasubrmanian, S., Barry, K., Bauer, D., Boehm, C. R., … Schmutz, J. (2017). Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. *Cell*, *171*(2), 287–304.

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

Bryant, N., Lloyd, J., Sweeney, C., Myouga, F., & Meinke, D. (2011). Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in

Arabidopsis. *Plant Physiology*, *155*(4), 1678–1689.

Budziszewski, G. J., Lewis, S. P., Glover, L. W., Reineke, J., Jones, G., Ziemnik, L. S., Lonowski, J., Nyfeler, B., Aux, G., Zhou, Q., McElver, J., Patton, D. A., Martienssen, R., Grossniklaus, U., Ma, H., Law, M., & Levin, J. Z. (2001). Arabidopsis genes essential for seedling viability: isolation of insertional mutants and molecular cloning. *Genetics*, *159*(4), 1765–1778.

Bursch, K., Toledo-Ortiz, G., Pireyre, M., Lohr, M., Braatz, C., & Johansson, H. (2020). Identification of BBX proteins as rate-limiting cofactors of HY5. *Nature Plants*, *6*(8), 921–928.

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

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

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

Dickinson, P. J., Kneřová, J., Szecówka, M., Stevenson, S. R., Burgess, S. J., Mulvey, H., Bågman, A.-M., Gaudinier, A., Brady, S. M., & Hibberd, J. M. (2020). A bipartite transcription factor module controlling expression in the bundle sheath of Arabidopsis thaliana. *Nature Plants*, *6*(12), 1468–1479.

Driever, S. M., Simkin, A. J., Alotaibi, S., Fisk, S. J., Madgwick, P. J., Sparks, C. A., Jones, H. D., Lawson, T., Parry, M. A. J., & Raines, C. A. (2017). Increased SBPase activity

improves photosynthesis and grain yield in wheat grown in greenhouse conditions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1730).

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., & Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. In *Trends in Plant Science*,15(10), 573-581

Du, H., Wang, Y.-B., Xie, Y., Liang, Z., Jiang, S.-J., Zhang, S.-S., Huang, Y.-B., & Tang, Y.-X. (2013). Genome-wide identification and evolutionary and expression analyses of MYB-related genes in land plants. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *20*(5), 437–448.

Edwards, K., Johnstone, C., & Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research*, *19*(6), 1349.

Fauser, F., Schiml, S., & Puchta, H. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. *The Plant Journal: For Cell and Molecular Biology*, *79*(2), 348–359.

Fitter, D. W., Martin, D. J., Copley, M. J., Scotland, R. W., & Langdale, J. A. (2002). *GLK*gene pairs regulate chloroplast development in diverse plant species. In *The Plant Journal*, 31(6), 713-727

Flores-Sandoval, E., Romani, F., & Bowman, J. L. (2018). Co-expression and Transcriptome Analysis of Transcription Factors Supports Class C ARFs as Independent Actors of an Ancient Auxin Regulatory Module. *Frontiers in Plant Science*, *9*, 1345.

Fornes, O., Castro-Mondragon, J. A., Khan, A., van der Lee, R., Zhang, X., Richmond, P. A., Modi, B. P., Correard, S., Gheorghe, M., Baranašić, D., Santana-Garcia, W., Tan, G., Chèneby, J., Ballester, B., Parcy, F., Sandelin, A., Lenhard, B., Wasserman, W. W., & Mathelier, A. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, *48*(D1), 87–92.

Frangedakis, E., Waller, M., Nishiyama, T., Tsukaya, H., Xu, X., Yue, Y., Tjahjadi, M., Gunadi, A., Van Eck, J., Li, F.-W., Szövényi, P., & Sakakibara, K. (2021). An Agrobacterium-mediated stable transformation technique for the hornwort model Anthoceros agrestis. *The New Phytologist*, *232*(3), 1488–1505.

Gangappa, S. N., Crocco, C. D., Johansson, H., Datta, S., Hettiarachchi, C., Holm, M., & Botto, J. F. (2013). The Arabidopsis B-BOX Protein BBX25 Interacts with HY5, Negatively Regulating BBX22 Expression to Suppress Seedling Photomorphogenesis. *The Plant Cell*, *25*(4), 1243–1257.

Gould, S. B., Waller, R. F., & McFadden, G. I. (2008). Plastid Evolution. In *Annual Review of Plant Biology* 56(1), 491-517

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* , *27*(7), 1017–1018.

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

Hall, L. N., Rossini, L., Cribb, L., & Langdale, J. A. (1998). GOLDEN 2: a novel transcriptional regulator of cellular differentiation in the maize leaf. *The Plant Cell*, *10*(6), 925–936.

Hibberd, J. M., Sheehy, J. E., & Langdale, J. A. (2008). Using C4 photosynthesis to increase the yield of rice-rationale and feasibility. *Current Opinion in Plant Biology*, *11*(2), 228–231.

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

Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T., Mori, H., Tajima, N., Moriyama, T., Ikeuchi, M., Watanabe, M., Wada, H., Kobayashi, K., Saito, M., Masuda, T., Sasaki-Sekimoto, Y., Mashiguchi, K., … Ohta, H. (2014).

Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. *Nature Communications*, *5*, 3978.

Ibarra, I. L., Hollmann, N. M., Klaus, B., Augsten, S., Velten, B., Hennig, J., & Zaugg, J. B. (2020). Mechanistic insights into transcription factor cooperativity and its impact on protein-phenotype interactions. *Nature Communications*, *11*(1), 124.

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

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

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. In *Nature Methods* 14(6), 587-589.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.

Kromdijk, J., Głowacka, K., Leonelli, L., Gabilly, S. T., Iwai, M., Niyogi, K. K., & Long, S. P. (2016). Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science*, *354*(6314), 857–861.

Langdale, J. A. (2011). C4 cycles: past, present, and future research on C4 photosynthesis. *The Plant Cell*, *23*(11), 3879–3892.

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

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*, *47*(W1), 256–259.

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

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

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

Long, S. P., Marshall-Colon, A., & Zhu, X.-G. (2015). Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell*, *161*(1), 56–66.

López-Juez, E. (2007). Plastid biogenesis, between light and shadows. *Journal of Experimental Botany*, *58*(1), 11–26.

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

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., & Penny, D. (2002). Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proceedings of the National Academy of Sciences of the United States of*

America, *99*(19), 12246–12251.

McFadden, G. I. (2014). Origin and Evolution of Plastids and Photosynthesis in Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, *6*(4), a016105.

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

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

Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K. K., Haas, F. B., Vanderstraeten, L., Becker, D., Lang, D., Vosolsobě, S., Rombauts, S., Wilhelmsson, P. K. I., Janitza, P., Kern, R., Heyl, A., Rümpler, F., Villalobos, L. I. A. C., Clay, J. M., … Rensing, S. A. (2018). The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization. *Cell*, *174*(2), 448–464.

Ort, D. R., Merchant, S. S., Alric, J., Barkan, A., Blankenship, R. E., Bock, R., Croce, R., Hanson, M. R., Hibberd, J. M., Long, S. P., Moore, T. A., Moroney, J., Niyogi, K. K., Parry, M. A. J., Peralta-Yahya, P. P., Prince, R. C., Redding, K. E., Spalding, M. H., van Wijk, K. J., … Zhu, X. G. (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(28), 8529–8536.

Pu, X., Yang, L., Liu, L., Dong, X., Chen, S., Chen, Z., Liu, G., Jia, Y., Yuan, W., & Liu, L. (2020). Genome-Wide Analysis of the MYB Transcription Factor Superfamily in Physcomitrella patens. In *International Journal of Molecular Sciences*, 21(3), 975

Ramsay, N. A., & Glover, B. J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends in Plant Science*, *10*(2), 63–70.

Rose, A., Meier, I., & Wienand, U. (1999). The tomato I-box binding factor LeMYBI is a member of a novel class of myb-like proteins. *The Plant Journal: For Cell and Molecular Biology*, *20*(6), 641–652.

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

Schreier, T. B., Müller, K. H., Eicke, S., Faulkner, C., Zeeman, S. C., & Hibberd, J. M. (n.d.). *Plasmodesmal connectivity in C_4Gynandropsis gynandrais induced by light and dependent on photosynthesis*. In *bioRxiv*

Schwechheimer, C., Schröder, P. M., & Blaby-Haas, C. E. (2022). Plant GATA Factors: Their Biology, Phylogeny, and Phylogenomics. *Annual Review of Plant Biology*, *73*, 123–148.

Sönmezer, C., Kleinendorst, R., Imanci, D., Barzaghi, G., Villacorta, L., Schübeler, D., Benes, V., Molina, N., & Krebs, A. R. (2021). Molecular Co-occupancy Identifies Transcription Factor Binding Cooperativity In Vivo. *Molecular Cell*, *81*(2), 255–267.

South, P. F., Cavanagh, A. P., Liu, H. W., & Ort, D. R. (2019). Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science*, *363*(6422),

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

Sullivan, A. M., Arsovski, A. A., Lempe, J., Bubb, K. L., Weirauch, M. T., Sabo, P. J., Sandstrom, R., Thurman, R. E., Neph, S., Reynolds, A. P., Stergachis, A. B., Vernot, B., Johnson, A. K., Haugen, E., Sullivan, S. T., Thompson, A., Neri, F. V., Weaver, M., Diegel, M., … Stamatoyannopoulos, J. A. (2014). Mapping and Dynamics of Regulatory DNA and

Transcription Factor Networks in A. thaliana. In *Cell Reports8(*6), 2015–2030.

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

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

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

Ursache, R., Fujita, S., Dénervaud Tendon, V., & Geldner, N. (2021). Combined fluorescent seed selection and multiplex CRISPR/Cas9 assembly for fast generation of multiple Arabidopsis mutants. *Plant Methods*, *17*(1), 111.

von Caemmerer, S., Quick, W. P., & Furbank, R. T. (2012). The development of C_4rice: current progress and future challenges. *Science*, *336*(6089), 1671–1672.

Wang, P., Fouracre, J., Kelly, S., Karki, S., Gowik, U., Aubry, S., Shaw, M. K., Westhoff, P., Slamet-Loedin, I. H., Paul Quick, W., Hibberd, J. M., & Langdale, J. A. (2013). Evolution of GOLDEN2-LIKE gene function in C3 and C4 plants. *Planta*, *237*(2), 481.

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

Waters, M. T., & Langdale, J. A. (2009). The making of a chloroplast. *The EMBO Journal*, *28*(19), 2861–2873.

Waters, M. T., Moylan, E. C., & Langdale, J. A. (2008). GLK transcription factors regulate

chloroplast development in a cell-autonomous manner. *The Plant Journal: For Cell and Molecular Biology*, *56*(3), 432–444.

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

Witte, C.-P., & Herde, M. (2020). Nucleotide Metabolism in Plants. *Plant Physiology*, *182*(1), 63–78.

Xie, K., Minkenberg, B., & Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(11), 3570–3575.

Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhaofeng, G., Zhiqiang, L., Yunfei, Z., Xiaoxiao, W., Xiaoming, Q., Yunping, S., Li, Z., Xiaohui, D., Jingchu, L., Xing-Wang, D., Zhangliang, C., Hongya, G., & Li-Jia, Q. (2006). The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Molecular Biology*, *60*(1), 107–124.

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

Yelina, N. E., Frangedakis, E., Schreier, T. B., Rever, J., Tomaselli, M., Haseloff, J., & Hibberd, J. M. (2023). Streamlined regulation of chloroplast development in the liverwort Marchantia polymorpha. In *bioRxiv*

Yelina, N. E., Holland, D., Gonzalez-Jorge, S., Hirsz, D., Yang, Z., & Henderson, I. R. (2022). Coexpression of MEIOTIC-TOPOISOMERASE VIB-dCas9 with guide RNAs specific to a recombination hotspot is insufficient to increase crossover frequency in Arabidopsis. *G3*, *12*(7).

Zhao, P., Cui, R., Xu, P., Wu, J., Mao, J.-L., Chen, Y., Zhou, C.-Z., Yu, L.-H., & Xiang, C.-B. (2017). ATHB17 enhances stress tolerance by coordinating photosynthesis associated nuclear gene and ATSIG5 expression in response to abiotic stress. *Scientific Reports*, *7*, 45492.

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

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

Plant Biochemistry and Molecular Biology. (1998). Ed PJ Lea and RC Leegood. Wiley

FIGURES

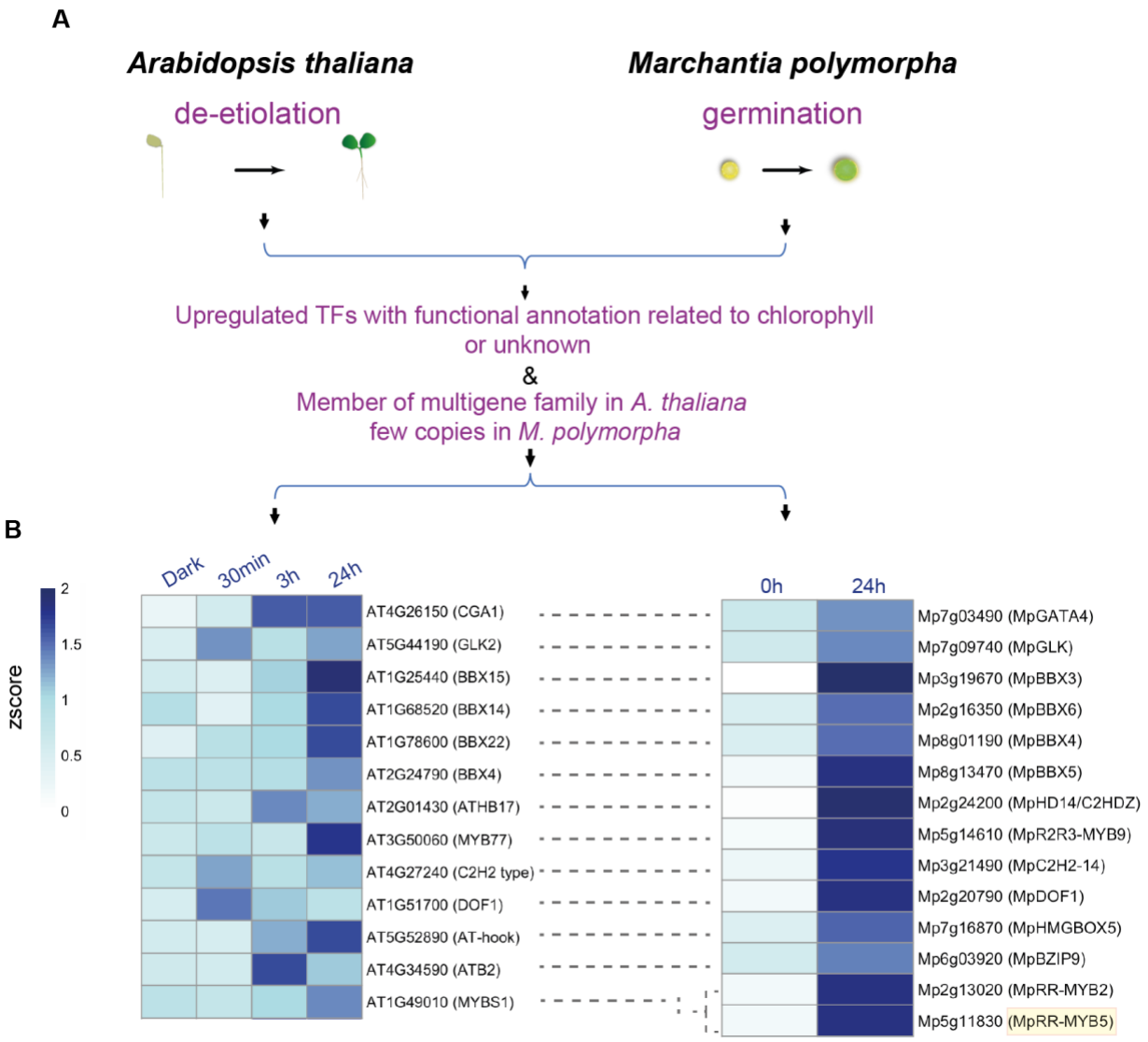


Figure 1: Schematic of pipeline used to identify candidate transcription factors regulating photosynthesis gene expression. A) RNA sequencing data from *Arabidopsis thaliana* during de-etiolation (Sullivan et al. 2014) and *Marchantia polymorpha* during spore germination (Flores-Sandoval et al. 2018) 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 hypothesized that functional redundancy had hindered identification of such regulators via forward genetic screens an additional criterion was that each family should be represented by multiple copies in *A. thaliana* 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. polymorp*ha.

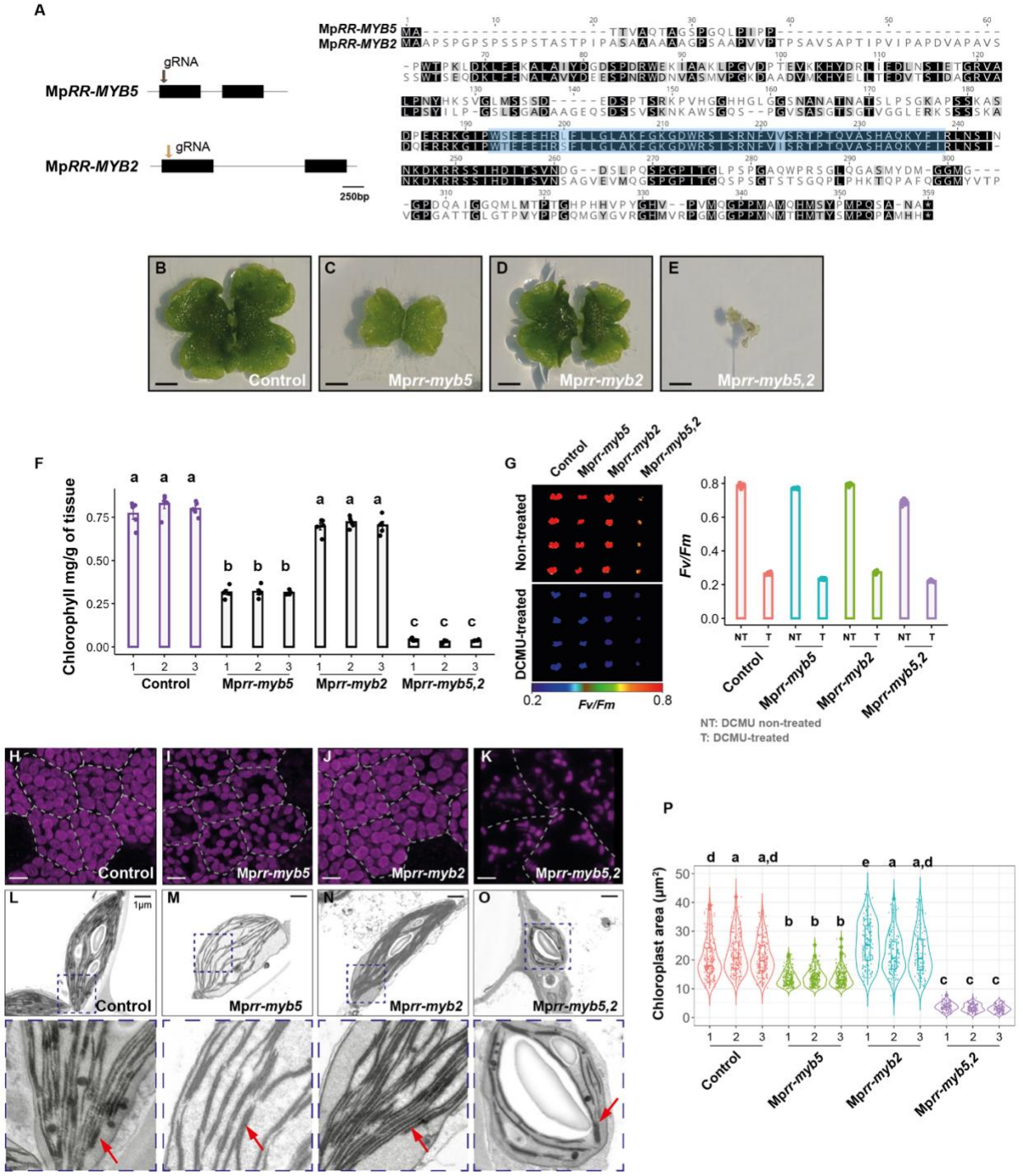


Figure 2: Mp*RR-MYB5* and Mp*RR-MYB2* act redundantly to regulate chloroplast biogenesis in *M. polymorpha*. A) Left: Schematic of Mp*RR-MYB5* and Mp*RR-MYB2* gene structure (exons represented as black boxes) with guide (g) RNAs for CRISPR represented by arrows. Right: Amino acid sequence alignments of MpRR-MYB5 and MpRR-MYB2 proteins with the characteristic RR-MYB/CCA1-like domain is highlighted in blue. B-E) Representative images of

control and Mp*rr-myb5,* Mp*rr-myb2* and Mp*rr-myb5,2* mutants. Scale bars represent 2 mm. F) Chlorophyll content of Mp*rr-myb5*, Mp*rr-myb2* and Mp*rr-myb5,2* mutants. Letters show statistical ranking using a *post hoc* Tukey test with different letters indicating statistically significant differences at $p<0.01$. Values indicated by the same letter are not statistically different. n=5. G) Representative images and quantification after imaging of the chlorophyll fluorescence parameter F_v/F_m with and without the inhibitor Di-Chlorophenyl Di-methyl Urea (DCMU). Although the Mp*rr-myb5,* Mp*rr-myb2* and Mp*rr-myb5,2* mutants have lower chlorophyll content, the drop in F_v/F_m values after treatment with DCMU indicates that photosystem II is functional. H-K) Representative images after confocal laser scanning microscopy of Mp*rr-myb5*, Mp*rr-myb2* and Mp*rr-myb5,2* mutants with chlorophyll autofluorescence shown in magenta. Cell borders are marked with dashed white lines. Scale bars represent 10 µm. L-O) Representative images after transmission electron microscopy of Mp*rr-myb5*, Mp*rr-myb2* and Mp*rr-myb5,2* mutants. Scale bars represent 1 µm. The dashed area depicted in each chloroplast is enlarged and grana stacks indicated with red arrows. P) Violin plots of chloroplast area for the Mp*rr-myb5,* Mp*rr-myb2* and Mp*rr-myb5,2* mutants. Box and whiskers represent the 25 to 75 percentile and minimum-maximum distributions of the data. Letters show statistical ranking using a *post hoc* Tukey test with different letters indicating statistically significant differences at $p<0.01$. Values indicated by the same letter are not statistically different, n=150.

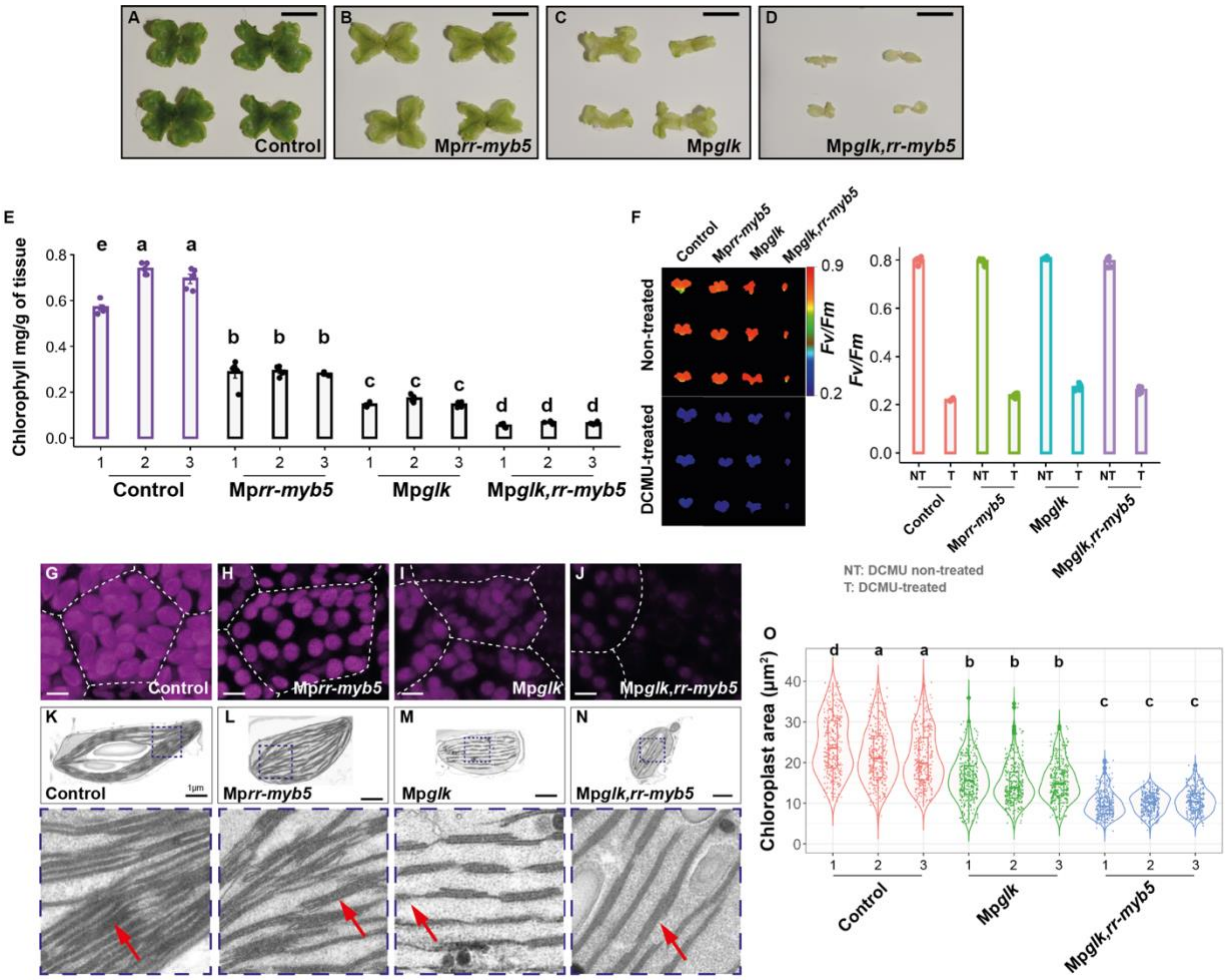


Figure 3: MpRR-MYB5 acts synergistically with MpGLK to control chloroplast biogenesis. A-D) Representative images of control, Mp*glk*, Mp*rr-myb5* and Mp*glk,rr-myb5* mutants. Scale bars represent 7 mm. E) Chlorophyll content is lower in Mp*glk*, Mp*rr-myb5* and the double Mp*glk,rr-myb5* mutants compared with controls. Letters show statistical ranking using a *post hoc* Tukey test with different letters indicating statistically significant differences at $p<0.01$. Values indicated by the same letter are not statistically different, n=5. F) Representative images and quantification after imaging of the chlorophyll fluorescence parameter F_v/F_m with and without the inhibitor Di-Chlorophenyl Di-methyl Urea (DCMU). Mp*rr-myb5,* Mp*glk* and Mp*glk,rr-myb5* mutants have lower chlorophyll content, the drop in F_v/F_m values after treatment with DCMU indicates that photosystem II is functional. G-J) Representative images after confocal laser scanning microscopy of control, Mp*rr-myb5,* Mp*glk* and Mp*glk,rr-myb5* mutants. Chlorophyll autofluorescence shown in magenta, cell borders marked with dashed white lines. Scale bars represent 10 µm. K-N) Representative images from transmission electron microscopy images of control, Mp*rr-myb5,* Mp*glk* and Mp*glk,rr-myb5* mutants. Scale bars represent 1 µm. The dashed area depicted in each chloroplast is enlarged and granal stacks indicated with red arrows. O) Chloroplast area of Mp*glk* and Mp*glk,rr-myb5* mutants. Box and whiskers represent the 25 to 75 percentile and minimum-maximum distributions of the data. Letters show statistical ranking using

a *post hoc* Tukey test with different letters indicating statistically significant differences at $p<0.01$. Values indicated by the same letter are not statistically different. n=330.

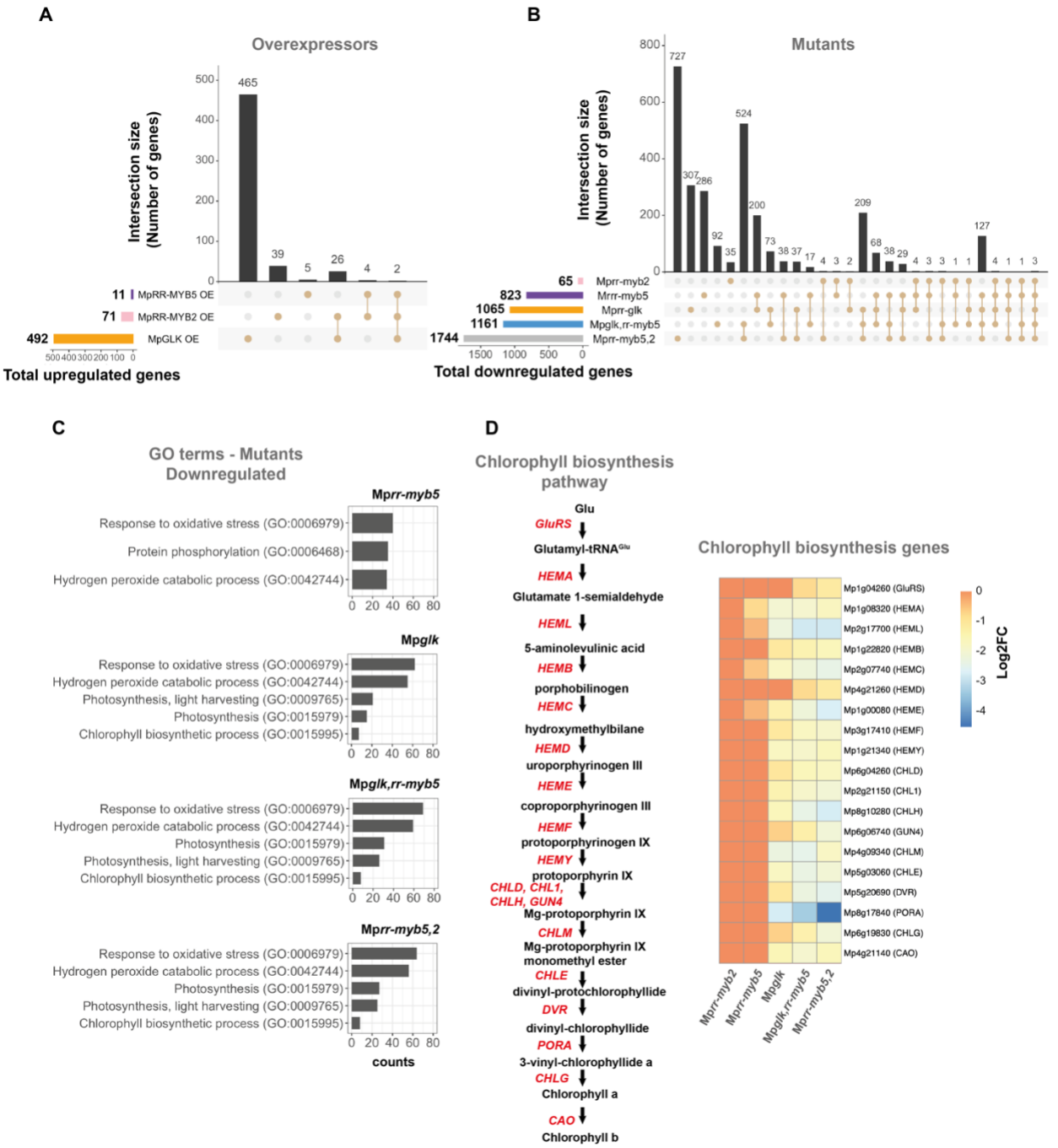

A
Overexpressors
Intersection size (Number of genes)
465
39
5
26
4
2
11 MpRR-MYB5 OE
71 MpRR-MYB2 OE
492 MpGLK OE
Total upregulated genes
B
Mutants
Intersection size (Number of genes)
727
307
286
92
35
524
200
73
38
37
17
4
3
2
209
68
38
29
4
3
3
1
1
127
4
1
1
3
65 Mprr-myb2
823 Mrrr-myb5
1065 Mprr-glk
1161 Mpglk,rr-myb5
1744 Mprr-myb5,2
Total downregulated genes
C
GO terms - Mutants Downregulated
Mprr-myb5
Response to oxidative stress (GO:0006979)
Protein phosphorylation (GO:0006468)
Hydrogen peroxide catabolic process (GO:0042744)
Mpglk
Response to oxidative stress (GO:0006979)
Hydrogen peroxide catabolic process (GO:0042744)
Photosynthesis, light harvesting (GO:0009765)
Photosynthesis (GO:0015979)
Chlorophyll biosynthetic process (GO:0015995)
Mpglk,rr-myb5
Response to oxidative stress (GO:0006979)
Hydrogen peroxide catabolic process (GO:0042744)
Photosynthesis (GO:0015979)
Photosynthesis, light harvesting (GO:0009765)
Chlorophyll biosynthetic process (GO:0015995)
Mprr-myb5,2
Response to oxidative stress (GO:0006979)
Hydrogen peroxide catabolic process (GO:0042744)
Photosynthesis (GO:0015979)
Photosynthesis, light harvesting (GO:0009765)
Chlorophyll biosynthetic process (GO:0015995)
counts
D
Chlorophyll biosynthesis pathway
Glu
GluRS
Glutamyl-tRNA^Glu
HEMA
Glutamate 1-semialdehyde
HEML
5-aminolevulinic acid
HEMB
porphobilinogen
HEMC
hydroxymethylbilane
HEMD
uroporphyrinogen III
HEME
coproporphyrinogen III
HEMF
protoporphyrinogen IX
HEMY
protoporphyrin IX
CHLD, CHL1, CHLH, GUN4
Mg-protoporphyrin IX
CHLM
Mg-protoporphyrin IX monomethyl ester
CHLE
divinyl-protochlorophyllide
DVR
divinyl-chlorophyllide
PORA
3-vinyl-chlorophyllide a
CHLG
Chlorophyll a
CAO
Chlorophyll b
Chlorophyll biosynthesis genes
Mp1g04260 (GluRS)
Mp1g08320 (HEMA)
Mp2g17700 (HEML)
Mp1g22820 (HEMB)
Mp2g07740 (HEMC)
Mp4g21260 (HEMD)
Mp1g00080 (HEME)
Mp3g17410 (HEMF)
Mp1g21340 (HEMY)
Mp6g04260 (CHLD)
Mp2g21150 (CHL1)
Mp8g10280 (CHLH)
Mp6g06740 (GUN4)
Mp4g09340 (CHLM)
Mp5g03060 (CHLE)
Mp5g20690 (DVR)
Mp8g17840 (PORA)
Mp6g19830 (CHLG)
Mp4g21140 (CAO)
Mprr-myb2
Mprr-myb5
Mpglk
Mpglk,rr-myb5
Mprr-myb5,2
Log2FC

Figure 4: Combined loss of Mp*RR-MYB5&2* has greater impact on abundance of transcripts encoding enzymes of chlorophyll biosynthesis than loss of Mp*GLK*. A) Upset diagrams showing sets of upregulated genes in Mp*RR-MYB5,* Mp*RR-MYB2* and Mp*GLK* over-expression lines. B) Upset diagrams showing sets of downregulated genes in Mp*glk*, Mp*rr-myb2,* Mp*rr-myb5,* and the double Mp*glk,rr-myb5* or Mp*rr-myb5,2* mutants. C) Enriched Gene Ontology terms for Mp*rr-myb5,* Mp*glk,rr-myb5,* Mp*glk* and Mp*rr-myb5,2* mutants. D) Heatmap illustrating the extent of down-regulation of transcripts encoding enzymes of chlorophyll biosynthesis in Mp*rr-myb2,* Mp*rr-myb5, Mpglk,* and the *Mpglk,rr-myb5* or Mp*rr-myb5,2* mutant alleles.

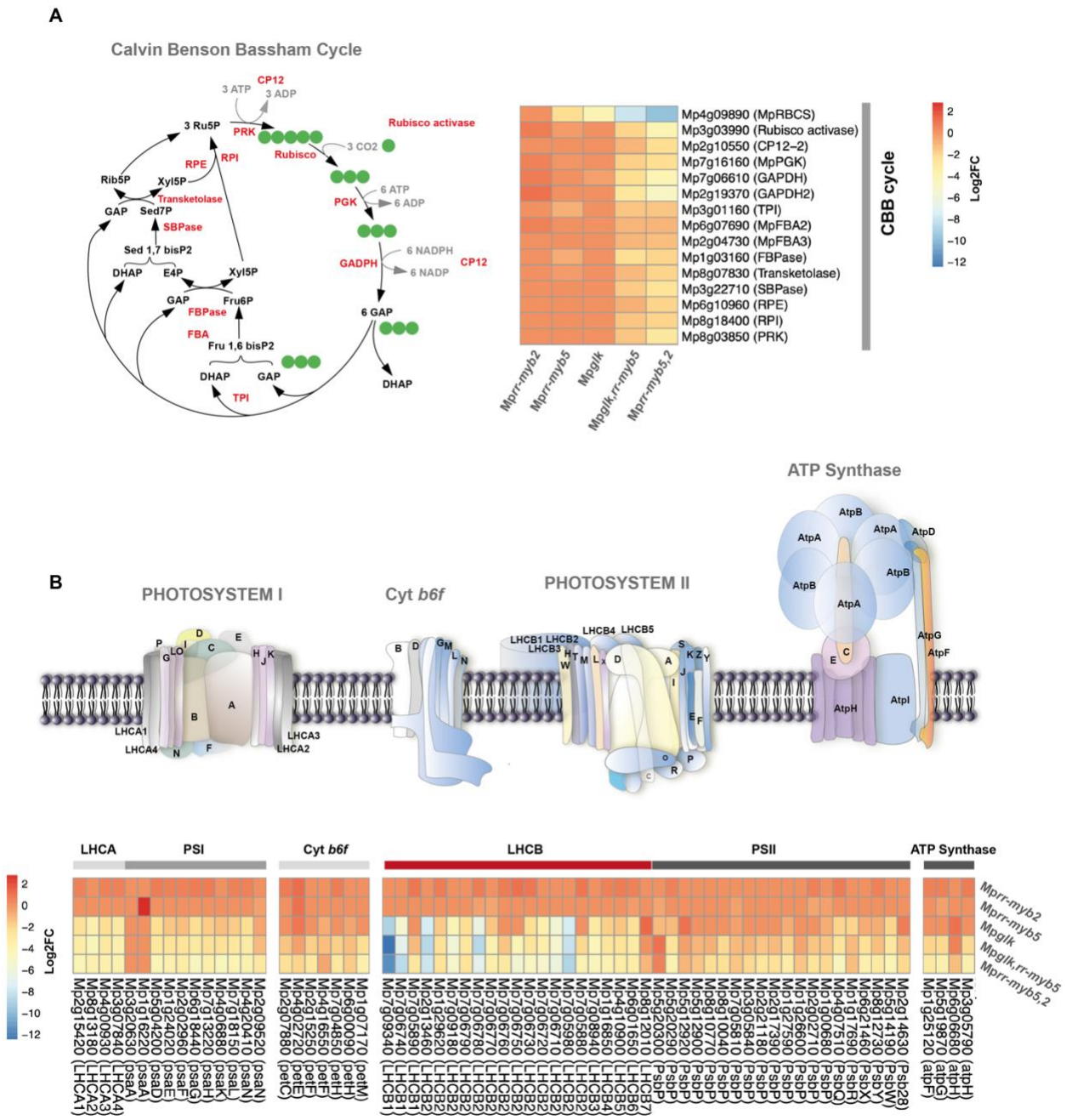


Figure 5: Loss of Mp*RR-MYB5* and Mp*RR-MYB2* downregulates expression of genes encoding the Calvin Benson Bassham cycle as well as those encoding the light harvesting apparatus. A) CCB cycle: Calvin-Benson-Bassham cycle, Heatmap indicating lower transcript abundance of genes encoding components of CCB in Mp*glk*, Mp*rr-myb5,* Mp*rr-myb2* as well as Mp*glk,rr-myb5*,*myb2* and Mp*rr-myb5*,*myb2* double mutants. Modified from Lea and Leegood, 1998. B) Heatmap indicating lower transcript abundance of genes encoding components of Photosystem II, Photosystem I and the cytochrome b_6f complex in Mp*glk*, Mp*rr-myb5,* Mp*rr-myb2* as well as Mp*glk,rr-myb5*,*myb2* and Mp*rr-myb5*,*myb2* double mutants. The greatest degree of

down regulation was detected in the double Mp*rr-myb5,2* double mutant. Schematic modified from Waters et al., 2009 and Tu et al., 2022.

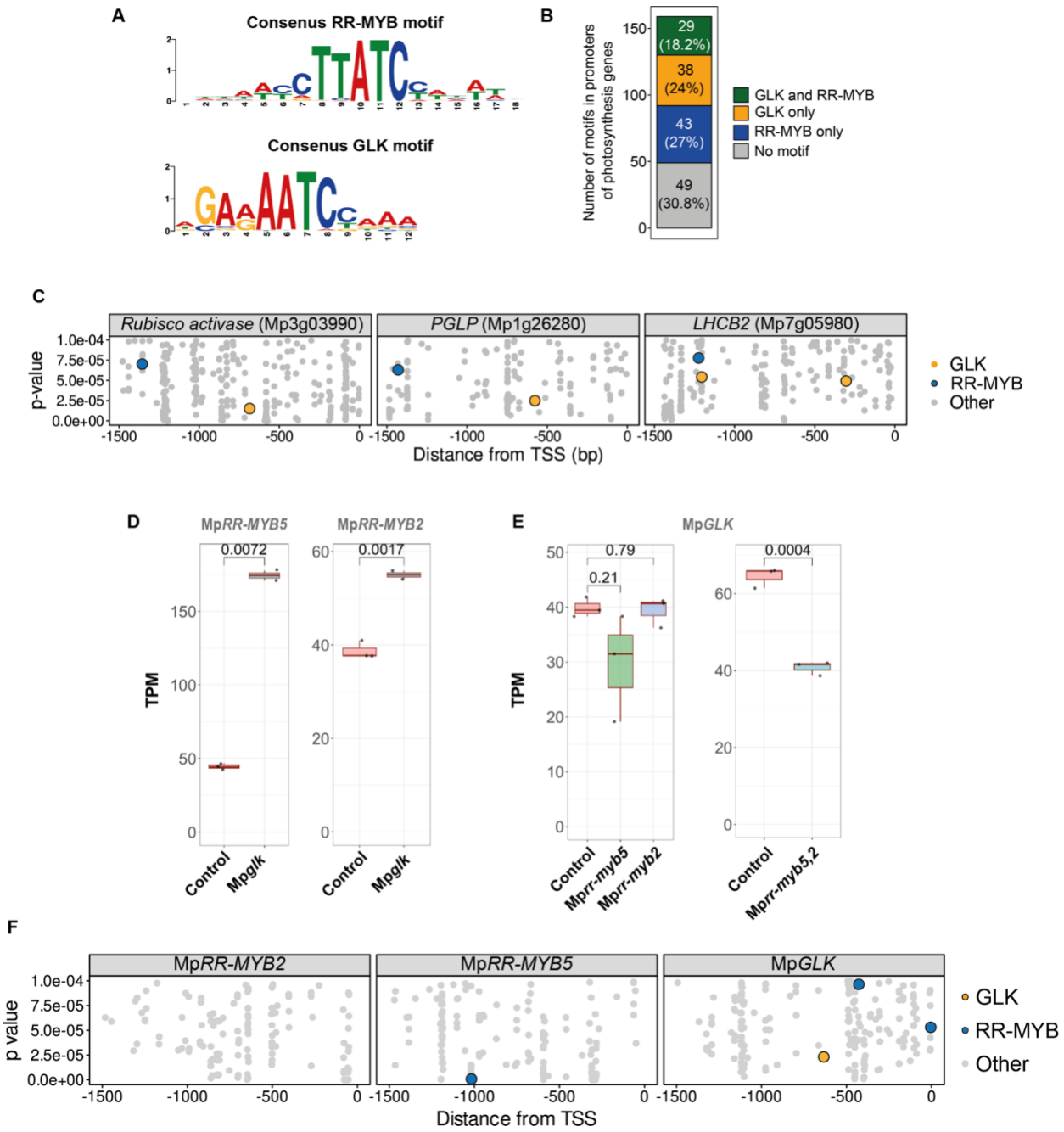


Figure 6: Combined loss of RR-MYB5&2 down-regulates GLK expression. A) Consensus motif sequence logo for RR-MYB and GLK transcription factors. B) Bar chart showing RR-MYB and GLK binding sites upstream of the predicted transcriptional start site (TSS) in *M. polymorpha* photosynthesis genes. C) Scatter plots showing position and predicted binding affinity of RR-MYB and GLK binding sites upstream of *RUBISCO ACTIVASE*, *PHOSPHOGLYCOLATE PHOSPHATASE (PGLP)* and *LIGHT HARVESTING COMPLEX B2 (LCHB2),* 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 TSS. P-values calculated by log-likelihood score using FIMO. D-E) Transcripts abundance of Mp*GLK,* Mp*RR-MYB2* and Mp*RR-MYB5* in control, Mp*rr-glk,* Mp*rr-myb5,* Mp*rr-myb2* and Mp*rr-myb5,2* mutant backgrounds. Data presented as

Transcript Per Million (TPM), and P-values of two-tailed *t*-test are shown. F) Scatter plots showing position and predicted binding affinity of GLK, RR-MYB binding and other motifs upstream of Mp*GLK,* Mp*RR-MYB2* and Mp*RR-MYB5* TSS, the 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 TSS. P-values calculated by log-likelihood score using FIMO.

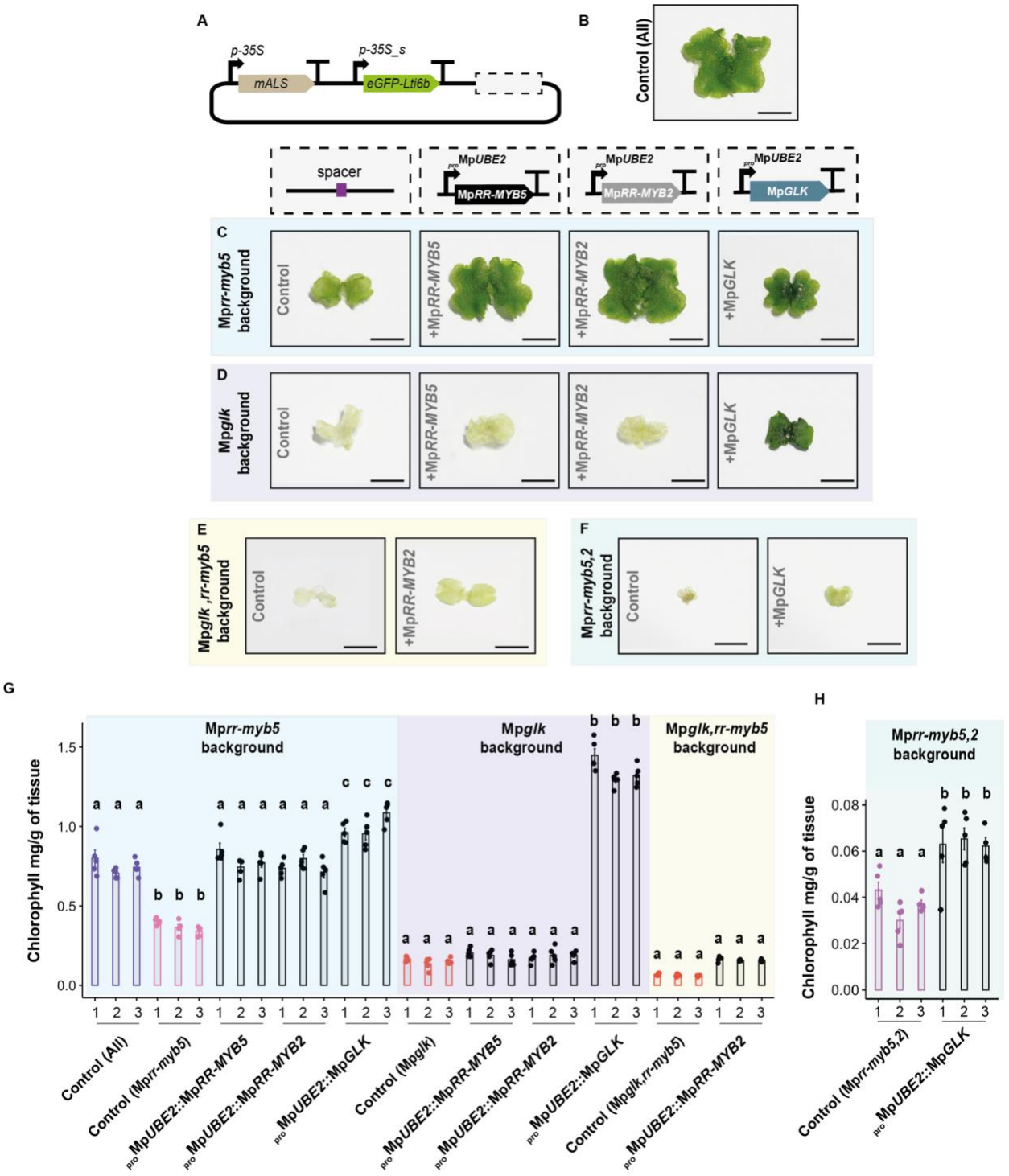


Figure 7: *M. polymorpha* RR-MYBs do not complement Mp*GLK*: A) Schematic representation of constructs used to overexpress Mp*RR-MYB5,* Mp*RR-MYB2* and Mp*GLK.* B) Representative images of control plants. Scale bar represents 5 mm. C-F) Representative images of Mp*rr-myb5,* Mp*glk,* Mp*glk,rr-myb5* and Mp*rr-myb5,2* mutants complemented with Mp*RR-MYB5,* Mp*RR-MYB2* and Mp*GLK.* Scale bars represent 2mm. G-H) Chlorophyll content in control, Mp*rr-myb5,* Mp*glk,* Mp*glk,rr-myb5* and Mp*rr-myb5,2* mutants complemented with Mp*RR-MYB5,* Mp*RR-MYB2* and

Mp*GLK.* Letters show statistical ranking using a *post hoc* Tukey test (comparisons are made between groups highlighted with the same colour rectangle, different letters indicating statistically significant differences at $P<0.01$). Values indicated by the same letter are not statistically different , n=5.

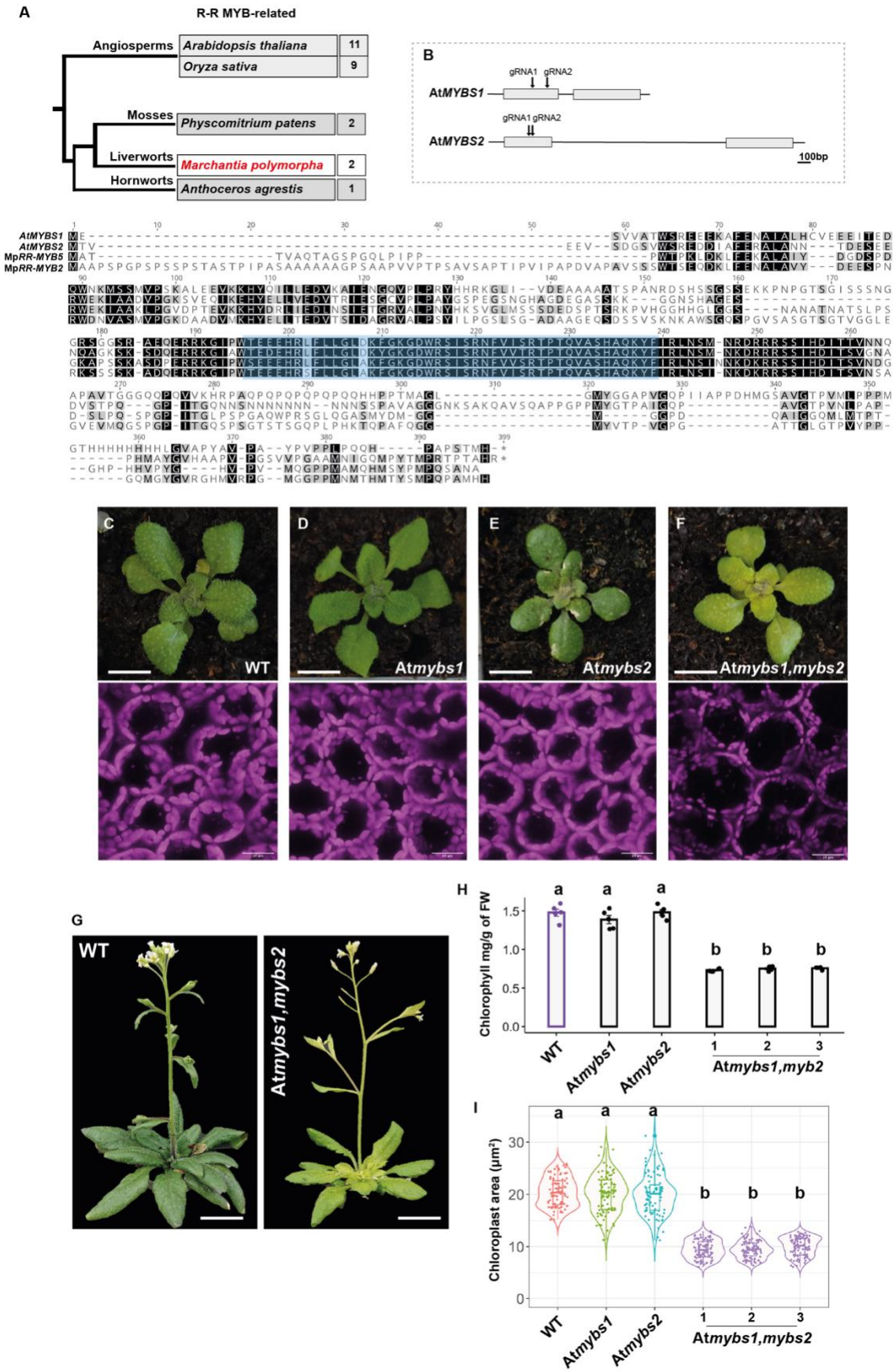
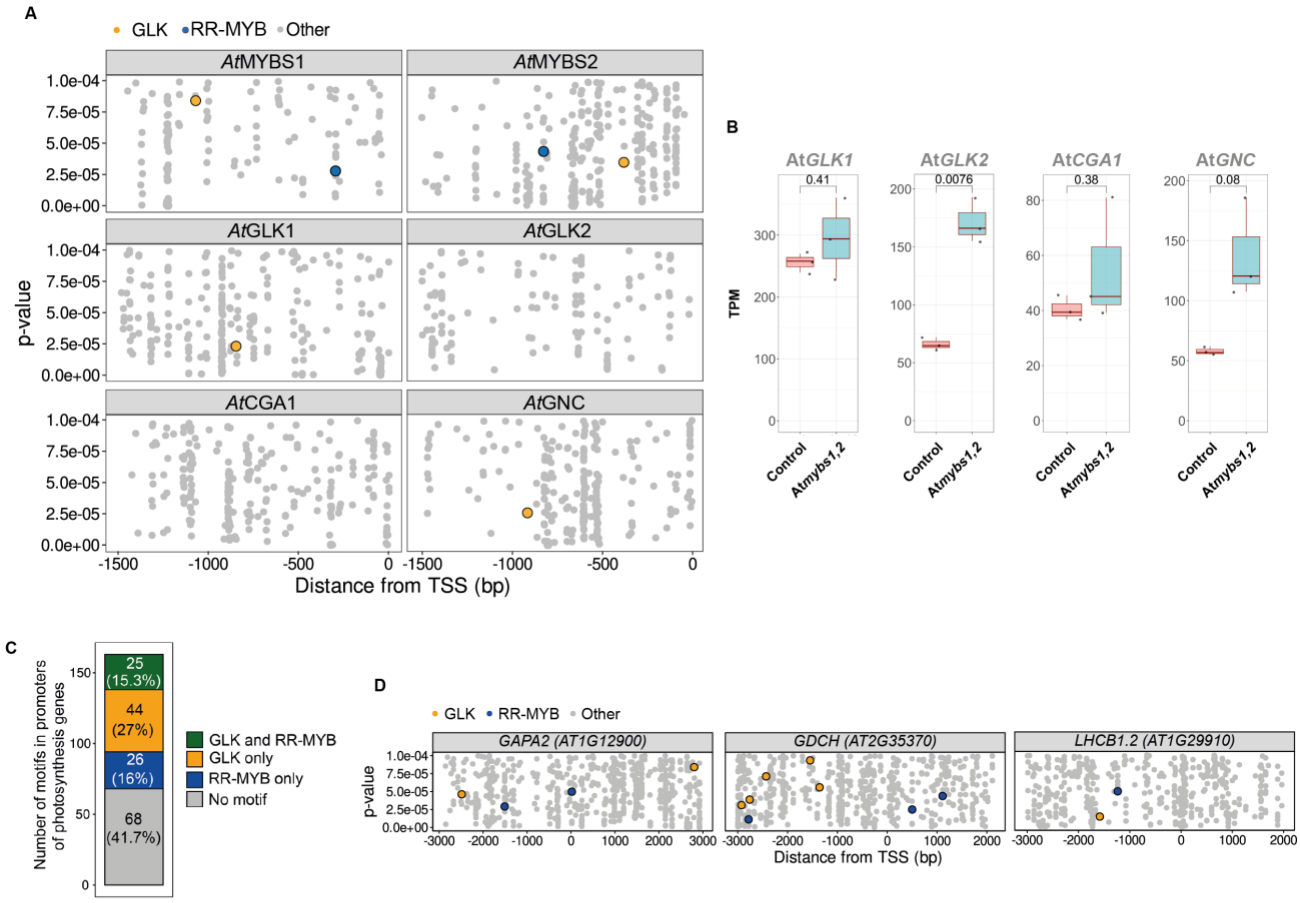

A
R-R MYB-related
Angiosperms
Arabidopsis thaliana
11
Oryza sativa
9
Mosses
Physcomitrium patens
2
Liverworts
Marchantia polymorpha
2
Hornworts
Anthoceros agrestis
1
B
gRNA1 gRNA2
AtMYBS1
gRNA1 gRNA2
AtMYBS2
100bp
AtMYBS1
AtMYBS2
MpRR-MYB5
MpRR-MYB2
C
WT
D
Atmybs1
E
Atmybs2
F
Atmybs1,mybs2
G
WT
Atmybs1,mybs2
H
Chlorophyll mg/g of FW
WT
Atmybs1
Atmybs2
Atmybs1,myb2
I
Chloroplast area (μm²)
WT
Atmybs1
Atmybs2
Atmybs1,mybs2

Figure 8: RR-MYBs control chlorophyll biogenesis in *A. thaliana.* A) Number of RR-Myb-Related transcription factors in *A. thaliana*, rice and bryophytes. Amino acid alignment the two MpRR-MYB factors in *M. polymorpha* (red arrowheads) and *A. thaliana* MYBS1 and MYBS2 (bottom). Characteristic gene domains are shown as coloured boxes. B) Schematic representation of At*MYBS1* and At*MYBS2* gene structure showing exons as grey rectangles. guide (g) RNAs positions for gene editing shown with arrows. C-F) Images of two-week old seedlings of wild type, At*mybs1,* At*mybs2* and At*mybs1,mybs2* mutants (scale bars represent 7 mm. Representative images after confocal laser scanning microscopy also shown (bottom) with scale bars representing 25 µm. G) Images of wild type and At*mybs1,mybs2* mutants with inflorescence. Scale bars represent 1.5 cm. H) Bar plots of chlorophyll content for wild type, At*mybs1,* At*mybs2* and At*mybs1,mybs2* mutants (three independent lines were used for measurements). Letters show statistical ranking using a *post hoc* Tukey test (with different letters indicating statistically significant differences at P<0.01). Values indicated by the same letter are not statistically different, n=5. I) Violin plots of chloroplast area for wild type, At*mybs1,* At*mybs2* and At*mybs1,mybs2* mutants. Box and whiskers represent the 25 to 75 percentile and minimum-maximum distributions of the data. Letters show statistical ranking using a *post hoc* Tukey test (with different letters indicating statistically significant differences at P<0.01). Values indicated by the same letter are not statistically different. n=100.



E

Chlorophyll biosynthesis

Gene	log2FC
AT3G48730 (GSA1)	-0.973404
AT5G08280 (HEMC)	-0.997152
AT2G40490 (HEME2)	-1.842105
AT5G18660 (DVR)	-0.908167
AT1G03630 (PORC)	-0.941689

Calvin Benson Bassham Cycle

Gene	log2FC
AT5G38410 (RBCS3B)	-1.158661
AT5G38420 (RBCS2B)	-1.34528
AT3G12780 (PGK1)	-1.033306
AT1G12900 (GAPA2)	-0.942342
AT1G55490 (CPNB1)	-0.8513
AT2G28000 (CPNA1)	-0.919179

Photorespiration

Gene	log2FC
AT2G35370 (GDHC)	-1.432987
AT1G32470 (GDC-H1)	-1.568028
AT1G32080 (L rgB)	-0.953238

Light harvesting

Gene	log2FC	
AT3G54890 (LHCA1)	-1.23513	LHCA
AT3G61470 (LHCA2)	-1.315061	LHCA
AT1G61520 (LHCA3)	-1.175695	LHCA
AT3G47470 (LHCA4)	-1.169918	LHCA
AT4G02770 (PsaD1)	-1.122695	PSI
AT4G28750 (PsaE1)	-1.125835	PSI
AT2G20260 (PsaE2)	-1.456162	PSI
AT1G31330 (PsaF)	-1.073336	PSI
AT1G55670 (PsaG)	-0.895592	PSI
AT3G16140 (PsaH1)	-1.077185	PSI
AT1G52230 (PsaH2)	-1.438544	PSI
AT1G08380 (PsaO)	-0.867905	PSI
AT1G30380 (PsaK)	-0.957567	PSI
AT2G46820 (PsaP)	-1.1607	PSI

Gene	log2FC	
AT1G29920 (LHCB1.1)	-1.10071	LHCB
AT1G29910 (LHCB1.2)	-0.904924	LHCB
AT1G29930 (LHCB1.3)	-1.761318	LHCB
AT2G34430 (LHCB1.4)	-6.082789	LHCB
AT2G34420 (LHCB1.5)	-1.049355	LHCB
AT5G54270 (LHCB3)	-1.386108	LHCB
AT5G01530 (LHCB4.1)	-1.3085	LHCB
AT3G08940 (LHCB4.2)	-0.929136	LHCB
AT4G10340 (LHCB5)	-0.811321	LHCB
AT1G15820 (LHCB6)	-1.561819	LHCB
AT5G66570 (PsbQ1)	-0.817722	PSII
AT4G05180 (PsbQ2)	-1.016175	PSII
AT2G30570 (PsbW)	-1.180948	PSII
AT2G06520 (PsbX)	-0.955055	PSII
AT3G21055 (PsbT)	-1.588563	PSII
AT4G21280 (PsbQ1)	-0.988605	PSII
AT1G51400 (Psb5)	-1.755429	PSII

Figure 9: Loss of At*MYBS1&2* downregulates expression of genes encoding chlorophyll biosynthesis enzymes, the Calvin Benson Bassham cycle as well as those encoding the light harvesting apparatus. A) Scatter plots showing position and predicted binding affinity of GLK, RR-MYB binding and other motifs upstream of At*GLK1,* At*GLK2,* At*MYBS1* and At*MYBS2* transcriptional start site (TSS). 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 TSS. P-values calculated by log-likelihood score using FIMO. B) Transcript abundance in Transcripts

Per Million (TPM) of At*GLK1,* At*GLK2,* At*CGA1* and At*GNC* in wild-type and At*mybs1,mybs2* mutant background. P-values of two-tailed *t*-test are shown. C) Barchart showing GLK and RR-MYB binding sites upstream of the transcriptional start site (TSS) in *A. thaliana* photosynthesis genes D) Scatter plots showing position and predicted binding affinity of GLK, RR-MYB binding and other motifs 3000 bp upstream of the transcriptional start site and 1000bp downstream of the end of the 3'UTR for*GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE A SUBUNIT 2 (GAPA2)*, *GLYCINE DECARBOXYLASE COMPLEX H (GDCH),* and *LIGHT HARVESTING COMPLEX B1.2 (LCHB1.2)* . 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. P-values calculated by log-likelihood score using FIMO. E) List of Chlorophyll biosynthesis, Calvin Benson Bassham (CBB) cycle, Photorespiration and Light harvesting genes downregulated in At*mybs1,mybs2* mutants.

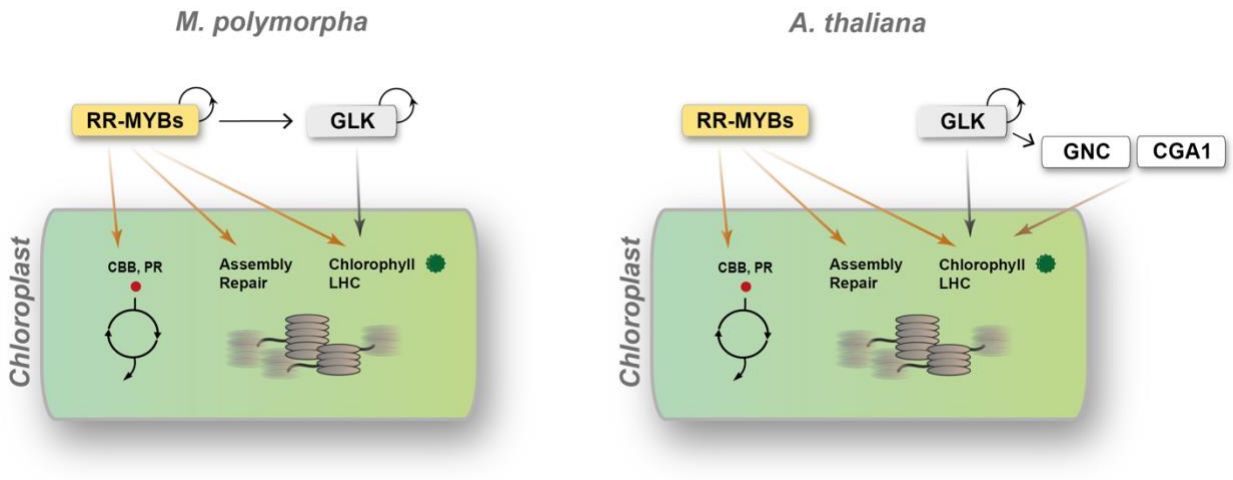


Figure 10: Model illustrating role of RR-MYBs in the control of chloroplast biogenesis. In the bryophyte *M. polymorpha* (left), RR-MYBs act upstream of GLK. Both MpRR-MYB and MpGLK appear to regulate themselves through feedforward loops. Whilst MpGLK regulates photosynthesis genes associated with chlorophyll biosynthesis and sub-units of the light harvesting complexes (LHC), RR-MYBs also regulate genes of the Calvin Benson Bassham (CBB) cycle and photorespiration (PR) as well as assembly and repair of the photosystems. In the dicotyledenous model *A. thaliana* (right) RR-MYBs no longer act upstream of *GLK* but they do regulate genes of the Calvin Benson Bassham (CBB) cycle and photorespiration (PR) as well as assembly and repair of the photosystems. At*GLK* regulates itself but also At*GNC*, and as with *M. polymorpha*, it controls photosynthesis genes associated with chlorophyll biosynthesis and sub-units of the light harvesting complexes (LHC).