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

Nataliya E. Yelina[1,2,*], Eftychios Frangedakis[1,*], Tina B. Schreier[1], Jenna Rever[1], Marta Tomaselli[1], Jim Haseloff[1] and Julian M. Hibberd[1,$]

[1]Department of Plant Sciences, University of Cambridge, Cambridge, CB3 EA, UK

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

* equal contributions

$ correspondence: Julian M. Hibberd jmh65@cam.ac.uk

Keywords: Chloroplast biogenesis, chlorophyll, Marchantia, transcription factors

Abstract

Photosynthesis in eukaryotic cells takes place in specialised plastids. The regulation of plastid development is crucial for multicellular systems such as plants. Two families of transcription factors known as Golden2-like (GLK) and GATA regulate plant chloroplast development, and the miR171-targeted SCARECROW-LIKE (SCL) GRAS transcription factors regulate chlorophyll biosynthesis. The extent to which these proteins carry out conserved roles in non-seed plants such as the liverworts is not known. Here we determine the degree of functional conservation of the GLK, GATA and SCL proteins in controlling chloroplast development in the model liverwort *Marchantia polymorpha*. Our results indicate that GATA and SCL do not play a detectable role in chloroplast biogenesis but loss of GLK function leads to reduced chloroplast size, underdeveloped thylakoid membranes and lower chlorophyll accumulation. These findings suggest that the functioning of GATA and SCL in chloroplast development either evolved after the divergence of vascular plants from bryophytes, that both roles were secondarily lost in *M. polymorpha*, or that functional redundancy is masking their roles. In contrast, and consistent with its presence in algae, GLK plays a conserved role in chloroplast biogenesis of liverworts and vascular plants.

Introduction

Photosynthesis sustains the majority of life on Earth. In eukaryotes it is performed in subcellular organelles known as chloroplasts that are thought to have originated from an endosymbiotic event between a photosynthetic prokaryote similar to cyanobacteria and a eukaryotic cell (Archibald 2009; Gould et al. 2008). Chloroplasts develop in response to light from undifferentiated proplastids in embryos, the shoot apical meristem or gametophytic cells. Key processes during chloroplast biogenesis include synthesis of chlorophyll, assembly of the thylakoid membranes and the photosynthetic apparatus and accumulation of enzymes of the Calvin Benson Bassham cycle in the chloroplast stroma (Jarvis and López-Juez 2013). Underpinning all of these events is the targeting and accumulation of transporter proteins in the chloroplast envelope because, since the endosymbiotic event, the majority of the genes controlling chloroplast biogenesis have transferred from the plastid to the nucleus and therefore need to be imported from the cytosol into the chloroplast (de Vries and Gould 2018; Gould et al. 2008). Chloroplast biogenesis therefore needs to be responsive to the external environment as well as the cell and so nuclear-encoded photosynthesis genes are regulated by light and hormones. Key intermediaries allowing responsiveness are transcription factors.

Our understanding of transcription factors acting on photosynthesis and chloroplast biogenesis is based primarily on analysis of the model flowering plant *Arabidopsis thaliana* (Figure 1A). For example, regulators that belong to the GARP and GATA families of transcription factors play key roles in chloroplast biogenesis (Cackett et al. 2022) and members of the SCARECROW-LIKE (SCL) GRAS family impact on chlorophyll accumulation (Ma et al. 2014). Within the GARP superfamily, Golden2-like (GLK) are positive regulators of nuclear-encoded chloroplast and photosynthesis related genes (Waters et al. 2008; Waters et al. 2009; Fitter et al. 2002; Bravo-Garcia et al. 2009). In the GATA family, GATA Nitrate-inducible Carbon metabolism-involved (GNC) and Cytokinin-Responsive GATA Factor 1 (CGA1) induce genes involved in chlorophyll biosynthesis and suppress phytochrome interacting factors as well as brassinosteroid (BR) related genes to promote chloroplast biogenesis and division (Chiang et al. 2012; Hudson et al. 2011; Naito et al. 2007; Bi et al. 2005). Moreover, GATA2 promotes photomorphogenesis by directly binding to light-responsive promoters (Luo et al. 2010) and in the absence of light the BR activated transcription factor BRASSINAZOLE RESISTANT1 (BZR1) represses GATA2 expression to inhibit photomorphogenesis. Lastly, three *A. thaliana* targeted SCL transcription factors redundantly regulate chlorophyll biosynthesis by suppressing the expression of *Protochlorophyllide Oxidoreductase C,* a key enzyme for chlorophyll production (Ma et al. 2014).

Land plants evolved from aquatic green algae, and approximately 500 MYA diverged into two major monophyletic clades, the vascular plants (which include angiosperms, gymnosperms, ferns and lycophytes) and the bryophytes (which include hornworts, liverworts and mosses) (Figure 1B) (Li et al. 2020; de Vries and Archibald 2018). Despite our detailed knowledge of chloroplast biogenesis in angiosperms, our understanding of this process in other land lineages including the bryophytes is poor. With the exception of a detailed study of GLK function in the moss *Physcomitrium patens* (Yasumura et al. 2005) little is known about chloroplast biogenesis in bryophytes. Consequently, how this fundamental biological process may have evolved is unclear. Bryophytes are of particular importance to infer more accurately the ancestral state of a trait. Whilst earlier evolutionary hypotheses proposed that bryophytes represent a collection of paraphyletic lineages (Qiu et al. 2006), current phylogenetic analyses indicate that the liverworts and mosses form a monophyletic group that split from hornworts approximately 400 MYA (Li et al. 2020; One Thousand Plant Transcriptomes Initiative 2019). This revised land plant phylogeny supports the notion that traits of the common ancestor of land plants could have diversified not only within the vascular plant clade but also in the three deeply divergent groups of bryophytes. Thus, studying representative species from more than one bryophyte clade is useful to gain insight into the likely ancestral state of any trait of interest. We therefore selected the model liverwort *Marchantia polymorpha* as a system to further investigate processes underpinning the evolution of chloroplast biogenesis. *M. polymorpha* has a small and well-annotated genome, key steps in its development are easily accessible for observation, and an extensive set of genetic manipulation tools are available (Bowman et al. 2022; Sauret-Güeto et al. 2020).

Phylogenetic analysis identified putative homologs for GATAs (GNC/CGA1 and GATA2), the SCL-miR171 module as well as GLK and suggested the regulatory network controlling chloroplast biogenesis in *A. thaliana* may be conserved in *M. polymorpha*. To test this we generated knock-out mutants of each protein. Although thallus growth was perturbed by editing the *GATA* and *SCL* genes, no differences in chlorophyll content were detectable. However, when *GLK* was mutated chlorophyll accumulation was reduced and chloroplast development perturbed. We conclude that in *M. polymorpha* the regulatory network controlling chloroplast biogenesis may be simpler than in angiosperms and that of the genes tested in this study *GLK* is the only conserved component. To better understand the extent to which *GLK* controls similar sets of genes in *M. polymorpha* and *A. thaliana* we generated constitutive over-expressors and performed RNA sequencing on over-expressing lines as well as *glk* mutant alleles. As previously reported in *A. thaliana*, the abundance of transcripts derived from genes associated with chlorophyll biogenesis and Photosystems I and II were impacted when *GLK* was mis-expressed in *M. polymorpha*. We

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

Results

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

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

GNC, *CGA1* and *GATA2* belong to the GATA superfamily of transcription factors that comprise a family of zinc finger proteins present in all eukaryotes. Plant GATAs can be categorised into four distinct classes (A, B, C and D) based on sequence conservation in the zinc finger motif, additional protein domains, and intron positions (Reyes et al. 2004) (Figure 1B, C and Supplemental Figure 2). *GATA2* is an A-Class GATA, which is characterized by the presence of a conserved glutamine (Q) and a threonine (T) within the zinc finger motif (Figure 1C and Supplemental Figure 2A). *GNC* and *CGA1* belong to the LLM subclass of B-Class GATAs due to a conserved Leucine–Leucine–Methionine (LLM) domain at their C-terminus (Behringer and Schwechheimer 2015) (Figure 1B, C and Supplemental Figure 2B). Our phylogenetic analysis showed that *M. polymorpha* has a total of six GATA genes, among which we identified *Mp7g03490* (annotated as Mp*GATA4*) and *Mp1g03950* (annotated as Mp*GATA2*) as single orthologs of *GNC/CGA1* and *GATA2* respectively (Figure 1B, C, Supplemental Figure 2C and D, and Supplemental Tables file 1). *SCL* is a member of the GRAS family of transcription factors which have a conserved C-terminal domain (Figure 1B, C). *A. thaliana AtSCL6, AtSCL22* and *AtSCL27* redundantly control chlorophyll biosynthesis and are regulated by miR171. The *M.*

polymorpha genome encodes eleven GRAS transcription factors. Genome-wide bioinformatic and next generation sequencing analysis of *M. polymorpha* miRNAs and their potential targets suggests that *Mp8g03980* (annotated as Mp*GRAS10*) is a target of miR171 (Lin and Bowman 2018). Although *Mp8g03980* is located in a sister clade to *A. thaliana* At*SCL6*, At*SCL22* and At*SCL27* (Lin and Bowman 2018), based on the conservation of its miR171-mediated regulation we refer to *Mp8g03980* as Mp*SCL*. Lastly, *GLK* genes belong to the GARP (GOLDEN2-LIKE, ARR-B, Psr1) family of transcription factors (Safi et al. 2017) and can be divided into two main subfamilies comprising the GARP-G2-like and GARP-ARR-B transcription factors. *GLK* genes belong to the GARP-G2-like group and are distinguished by a characteristic C-terminal domain called GOLDEN2 C-terminal (GCT) box (Figure 1C and Supplemental Figure 3A) (Fitter et al. 2002). Our phylogenetic analysis showed that *GLK* is present in one or two copies in most land plants, but multiple copies are present in some moss and fern species. *GLK* genes were thought to be land plant specific, however, we were able to detect *GLK* orthologs in two Zygnematophyceae green algae for which genome assemblies became recently available (Cheng et al. 2019) (two genes in *Spirogloea muscicola* [SM000086S23019 and SM000015S01186] and one gene in *Mesotaenium endlicherianum* [ME000101S10682]) (Supplemental Fig3B-C). Although *M. polymorpha* has eight GARP-G2 genes, we identified *Mp7g*09740 (annotated as Mp*GARP8*) as a single ortholog of *GLK* (Figure 1B, C, Supplemental Figure 3A-C, and Supplemental Tables file 1). In summary, our analysis indicates that *M. polymorpha* contains single orthologs of *GNC/CGA1*, *GATA2*, *SCL* and *GLK* that we hereafter refer to as Mp*GATA4*, Mp*GATA2*, Mp*SCL* and Mp*GLK* respectively.

Mp*glk* mutants show compromised chlorophyll accumulation

To test whether the Mp*GATA4*, Mp*GATA2*, Mp*SCL* and Mp*GLK* orthologs control chloroplast biogenesis in *M. polymorpha* we used CRISPR/Cas9 editing to generate knockout mutant alleles for each gene. For Mp*GATA4* two guide RNAs (gRNAs) were used to target the second and third exons, for Mp*GATA2* one gRNA targeted the second exon, for Mp*SCL* two gRNAs targeted the middle of its single exon, and for Mp*MIR171* two gRNAs flanking the miR171 gene were designed (Figure 1C). Finally, for Mp*GLK* a single gRNA was designed to target the sixth exon (Figure 1C). Plants transformed with the same vector but without a gRNA sequence were used as 'no guide RNA' controls. Each mutant line was clonally propagated through one gemmae generation to obtain isogenic plants, and for each targeted gene three independent lines selected for analysis. Mutant alleles of Mp*GATA4* contained deletions of 93 bp (line 1 /201), 862bp (line 2 /215) and 248bp (line 3/ 227) (Supplemental Figure 4). For Mp*GATA2* mutant alleles with two distinct 12bp

base pair deletions (lines 1 and 2 /106&109) and a 16bp insertion (line 3/ 62) were isolated (Supplemental Figure 4). Mutant alleles of Mp*SCL* contained 59bp or 7bp deletions that introduced frameshifts resulting in premature stop codons. miRNA genes contain a ~21-bp miRNA sequence and its reverse complement called a miRNA* separated by a spacer. With the exception of 10 bp of the Mp*MIR171** sequence, Mp*mir171* mutants had a 139 bp deletion and 1 bp insertion resulting in a deletion of the entire Mp*MIR171* gene (Supplemental Figure 4). When transcribed, miRNA genes produce a stem-loop or a 'hairpin' miRNA precursor which is cleaved to release a mature miRNA. Deletion of the miRNA171 sequence, the spacer and ~half of the miR171* sequence in our Mp*mir171* mutants, therefore, leads to a miR171 loss. Finally, for the Mp*GLK* mutant allele, a frameshift and a premature stop codon are predicted to generate a truncated protein of 201 amino acids (compared to the full-length 585 amino acid protein) (Supplemental Figure 4). We tested whether mRNA levels from the mutated genes were affected in the respective mutants but did not observe a clear correlation between mRNA levels and mutant phenotypes (Supplemental Figure 5).

Mp*mir171* and Mp*gata2* mutants did not show any morphological or developmental phenotypes and had similar chlorophyll levels and chlorophyll a/b ratios compared with wild type and 'no guide RNA' controls (Figure 1D, F, H, K and L). Mp*gata4* and Mp*scl* mutants showed altered thallus morphology. For example, Mp*gata4* mutants had narrower thallus lobes, while thalli of Mp*scl* mutants were 'stunted' with inward curling edges (Figure 1E and G). However, in both cases chlorophyll levels were not statistically different from controls with the exception of one Mp*gata4* line that had an ~10% increase. (Figure 1J, M). Mp*gata4* and Mp*scl* mutant chlorophyll a/b ratios were slightly reduced. In contrast, Mp*glk* mutants had an obvious pale green phenotype (Figure 1I). Quantitative assays showed that chlorophyll content was ~90% lower and chlorophyll a/b ratios higher than 'no guide RNA' controls (Figure 1N). Mp*glk* mutants also showed morphological changes with narrower thallus lobes and upward curling lobe edges (Figure 1I).

In summary, absence of functional Mp*gata4* and Mp*gata2* genes did not appear to affect chlorophyll biosynthesis or chloroplast biogenesis in *M. polymorpha*. As *GNC/CGA1* mutants of *A. thaliana* contain lower chlorophyll levels (Zubo et al. 2018), our data argue for lack of functional conservation between *A. thaliana* and *M. polymorpha* GNC/CGA1 and Mp*GATA4* in regulating chloroplast biogenesis. Similarly, we found a lack of conservation between the function of At*GATA2* and Mp*GATA2*. Moreover, despite conservation of a miR171 target site in the Mp*SCL*, it does not appear to control chlorophyll content. Again, this contrasts with the role in *A. thaliana* where triple At*scl6* At*scl22* At*scl27* mutants show increased chlorophyll levels and At*MIR171* mimic lines which are functional equivalents of At*mir171* knockouts have reduced chlorophyll

levels (Todesco et al. 2010; Ma et al. 2014). In contrast to *M. polymorpha GNC/CGA1, GATA2* and *SCL* orthologs, Mp*glk* mutants accumulated less chlorophyll indicating that the Mp*GLK* ortholog is required for proper chloroplast biogenesis. This implies a conserved function for GLK in chloroplast biogenesis in land plants from liverworts to angiosperms.

An allelic series of Mp*GLK* mutants

The Mp*GLK* gene has 11 exons with the conserved DNA-binding GARP domain encoded by exons 6 to 8 and the GCT domain by exon 11 (Figure 2A). To investigate the importance of different MpGLK domains we used CRISPR/Cas9 gene editing to mutate either the N-terminal or the DNA binding domains (Figure 2A). All generated Mp*glk* mutants were pale green compared with controls and had reduced chlorophyll levels (Figure 2B-F). However, they were still able to grow and produce gemmae and reproductive organs (Supplemental Figure 6A, B) without the need for supplemental carbon sources. Disruption of the Mp*GLK* genomic sequence between exons 1 and 3 resulted in 'weak' Mp*glk* alleles (Figure 2A-C, Mp*glk* g3+7 *lines 1-3*) with chlorophyll content from 72 to 55 % lower than in the controls (Figure 2F). In contrast, frameshifts in exon 6 (upstream of the GARP DNA-binding domain) leading to a premature stop codon (Figure 2A, B, D, Mp*glk* g7 *lines 1-3*), or deletion of a gene fragment between exons 6 and 8 encoding the conserved GARP DNA-binding domain (Figure 2A, E, Mp*glk* g14+g17 *lines 1-3*) resulted in 'strong' alleles with chlorophyll ~90% lower than in the controls (Figure 2F). We found that lower chlorophyll content in *M. polymorpha* Mp*glk* mutants was associated with increased chlorophyll a/b ratios (Figure 2F) consistent with previous reports for Arabidopsis At*glk1* At*glk2* mutants (Waters et al. 2009). We hypothesise that in the 'weak' Mp*glk* alleles disruption of Mp*GLK* genomic sequence between exons 1 and 3 allowed translation from an alternative start codon resulting in an N-terminally truncated but partially functional MpGLK protein, while in the 'strong' Mp*glk* alleles frameshifts in exon 6 resulting in a premature stop codon or deletion of an Mp*GLK* gene portion encoding a conserved GARP DNA binding domain, led to a loss-of-function MpGLK protein.

To better understand the effects of Mp*glk* mutations on chloroplast size and number we regenerated 'strong' Mp*glk* mutants combined with a transgene to mark cell boundaries via a plasma membrane-targeted eGFP fluorescent protein. These lines (Mp*glk* g7 MM *lines 1-3*) contained only ~85% chlorophyll compared with wild type (Figure 3A-D and Supplemental Figure 6C-E). We found that *Mpglk* g7 MM mutants had smaller chloroplasts compared with the controls while chloroplast number per cell was slightly increased (Figure 3E-H). We used Mp*glk* g7 (lines 1 and 2) as well as Mp*glk* g14+g17 (lines 1 and 2) that had a ~90% reduction in chlorophyll content

to analyse chloroplast ultrastructure via electron microscopy (Figure 3G, H). Consistent with our confocal microscopy observations, Mp*glk* mutants had smaller chloroplasts compared with the controls and showed perturbed ultrastructure. Specifically, chloroplasts had fewer thylakoid membranes that often did not occupy the full chloroplast area and had reduced granal stacking. Mp*glk* chloroplasts also showed visible signs of light stress such as increased numbers of plastoglobules and stromal extrusions (stromules) (Figure 3G, H and Supplemental Figure 6F).

We tested whether these poorly developed chloroplasts in the Mp*glk* 'strong' mutants were functional by treatment with di-chlorophenyl di-methyl urea (DCMU) which inhibits the photosynthetic electron transport chain (Figure 3I, J) (Tresbst 2007). Although they contained very little chlorophyll a substantial reduction in chlorophyll fluorescence following DCMU treatment was evident in the mutants indicating that the photosynthetic apparatus was functional (Figure 3I, J). This is consistent with the ability of 'strong' Mp*glk* alleles to grow under standard conditions without a carbon supplement.

In summary, in our Mp*glk* mutant alleles chlorophyll content was reduced between ~55% and ~90% compared with controls. To our knowledge, a ~90% chlorophyll loss represents the strongest reported so far in a *glk* mutant in any plant species. Mutations that either caused a deletion of the GARP DNA-binding domain or introduced a premature stop codon immediately upstream of it resulted in a more severe phenotype than mutations in the 5´ end of the Mp*GLK* gene. 'Strong' Mp*glk* mutants contained smaller chloroplasts with altered ultrastructure compared with control plants but the photosynthetic apparatus was functional. These Mp*glk* mutant phenotypes are consistent with the Mp*GLK* role in promoting chloroplast biogenesis and greening.

The Mp*GLK* 3´ UTR reduces accumulation of Mp*GLK* transcripts

Overexpression of *GLK* in angiosperms such as rice and *A. thaliana* can increase chlorophyll content and chloroplast size (Nakamura et al. 2009, Wang et al. 2017, Zubo et al. 2018). Although overexpression of rice *GLK* in rice is not maintained beyond the seedling stage (Nakamura et al. 2009) and overexpression of the maize *GLK* in rice can generate more sustained responses, it is reduced in a sequence-specific manner throughout plant development suggesting the potential for a post-transcriptional regulatory mechanism (Wang et al. 2017). Consistent with this hypothesis, previous analysis identified that Mp*GLK* mRNA may be subject to cleavage by miR11666.4 and another unidentified miRNA or siRNA via target sites in the 3´ untranslated region (UTR) (Supplemental Figure 7 and Lin et al. 2016).

To investigate whether Mp*GLK* is subject to post-transcriptional regulation we generated three constructs where the Mp*GLK* coding sequence (CDS) was driven by the strong promoter for a

constitutive ubiquitin-conjugating enzyme E2 gene (Mp*UBE2*) (Sauret-Güeto et al. 2020) (Figure 4A-E). In the first construct we used the native Mp*GLK* coding sequence, hereafter referred to as Mp*GLK-CDS*. The second construct was designed to test whether Mp*GLK* coding sequence contains motifs targeting its mRNA for post-transcriptional downregulation and so nucleotide sequence was mutated but amino acid sequence preserved (Supplemental Figure 8). We hypothesised that this would remove nucleotide sequence motifs targeting Mp*GLK* for post-transcriptional degradation. The resulting 're-written' coding sequence was named Mp*GLK-CDSrw*. In the third construct the native Mp*GLK* coding sequence was fused to the Mp*GLK* 3´ UTR sequence (Mp*GLK-CDS+3´UTR*) to test whether this would reduce greening. In all constructs we included a plasma-membrane-targeted eGFP (Sauret-Güeto et al. 2021) to visualise cell boundaries and quantify chloroplast numbers per cell. There were no differences in chlorophyll levels or morphology between plants in which the plasma membrane was marked with eGFP and those without the fluorescent marker (Supplemental Figure 9).

Mp*GLK-CDS* and Mp*GLK-CDSrw* overexpressing transformants were darker green compared with the 'empty vector' controls and showed perturbations to morphology such as stunted growth (Figure 4 C, D, G, H). qPCR analysis confirmed that the Mp*GLK-CDS* and the Mp*GLK-CDSrw* transgenes were overexpressed (Figure 4J, K). Lines expressing Mp*GLK-CDS* and the Mp*GLK-CDSrw* showed up to ~4 times higher chlorophyll levels compared with 'empty vector' controls (Figure 4L). Interestingly, chlorophyll content in *MpGLK-CDSrw* was not higher than the *MpGLK-CDS* lines (Figure 4L). Plants expressing Mp*GLK-CDS+3´UTR* showed intermediate chlorophyll accumulation between 'empty vector' controls and those expressing Mp*GLK*-CDS (Figure 4L). For all overexpressing lines chlorophyll a/b ratios were slightly reduced (~10%). qPCR analysis confirmed that the Mp*GLK-CDS+3´UTR* transgene was over-expressed (Figure 4J). Mp*GLK* mRNA levels were lower in Mp*GLK-CDS+3´UTR* compared with the *MpGLK-CDS* lines (Figure 4J), which is consistent with our hypothesis that Mp*GLK* is post-transcriptionally downregulated via its *3´UTR.* To understand which part of Mp*GLK* 3´UTR negatively regulates greening we tested a series of truncated Mp*GLK* 3´UTR versions covering positions 1-100, 1-200, 1-300, 1-400, 1-500 and 1-610 as well as the full length 671 bp sequence (Figure 5A). The 1-200 bp truncated version contained a predicted miR11666.4 target site located at positions 105-124 and an additional putative miRNA/siRNA cleavage site located between positions 135 and 136 (Lin and Bowman 2018). We then fused the truncated versions of Mp*GLK* 3´UTR to Mp*GLK*-CDS and overexpressed these constructs under the control of the M*pUBE2* promoter. Only the 1-500 bp and 1-610 bp fragments of the Mp*GLK* 3´UTR had the same negative effect on Mp*GLK*-mediated greening as the full-length Mp*GLK* 3´UTR (Figure 5A, B). This suggests a presence of an

unknown regulatory motif(s) within the 400-671 bp region of Mp*GLK* 3´UTR that are responsible for Mp*GLK* downregulation.

Finally, we analysed chloroplast morphology in gemmae of Mp*GLK* overexpression lines via confocal laser scanning microscopy. A mature gemma has two peripheral meristematic regions known as notches, while gemmae's central part is more mature and comprised of up to five cells layers including non-photosynthetic rhizoids precursor cells. Oil body cells is another non-photosynthetic cell type found on gemmae periphery (Kato et al. 2020). Confocal imaging of the central part of gemmae showed that cells overexpressing Mp*GLK-CDS* and Mp*GLK-CDSrw* but not Mp*GLK-CDS+3´UTR* contained more densely packed chloroplasts (Figure 5C-G and Supplemental Figure 10A-D). Average chloroplast number per cell area was the same as in the controls but chloroplasts were on average larger and in some case extremely large (Figure 5F, I and Supplemental Figure 10E-F and Supplemental Figure 11). Interestingly, particularly in the Mp*GLK-CDS* lines we observed an increase in chloroplast size in non-photosynthetic rhizoid precursor and oil body cells compared with control lines (Figure 5H). In summary, we show that Mp*GLK* 3´ UTR reduces greening that can be induced by Mp*GLK* overexpression in *M. polymorpha.*

GLK in *M. polymorpha* regulates thylakoid associated photosynthetic components and chlorophyll biosynthesis, similarly to *A. thaliana*

To identify genes that respond to GLK in both *A. thaliana* and *M. polymorpha*, and to understand GLK targets that may be unique to *M. polymorpha* we performed RNA sequencing in four Mp*glk* mutants (two 'strong' and two 'weak' alleles) and *MpGLK* overexpression lines alongside controls. We obtained ~45 million reads per accession, out of which 90-95% mapped to the *M. polymorpha* transcriptome (Supplemental Tables file 2). Principal component analysis of normalised counts shows that genotype background accounted for 62% of the variance detected in the data (Supplemental Figure 12A). 839 differentially expressed genes were detected in all four Mp*GLK* overexpressing lines compared with controls (*padj*-value =<0.01 test, LFC ≥ 1-fold) of which 493 were upregulated and 346 were downregulated (Supplemental Figure 12B, Supplemental Tables file 2). When only 'weak' Mp*glk* mutant alleles were assessed, there were 656 differentially expressed genes (*padj*-value =<0.01 test, LFC ≥ 1-fold) of which 190 and 466 were up- and downregulated, respectively (Figure 6C, Supplemental Tables file 2). As expected, 'strong' Mp*glk* mutant alleles had a greater impact on gene expression with 1536 differentially expressed genes (*padj*-value =<0.01 test, LFC ≥ 1-fold), of which 471 and 1065 were up- and downregulated, respectively (Supplemental Figure 12D, Supplemental Tables file 2). 32 genes

were differentially expressed in the Mp*GLK* overexpression lines as well as 'weak' and 'strong' mutant alleles (Supplemental Figure 12E). Gene Ontology analysis showed that in both overexpression lines and 'strong' Mp*glk* mutants, oxidation and reduction as well as carbohydrate metabolic processes were most impacted by mis-regulation of *GLK* (Figure 6A). When we specifically interrogated the effects of Mp*GLK* overexpression on chlorophyll and photosynthesis genes we found significant overlap with known GLK targets in *A. thaliana* (Tu et al. 2022; Waters et al. 2009). For example, in *M. polymorph*a GLK mis-expression affected nine of the 14 chlorophyll biosynthesis genes including *HEMB*, *HEMC*, *HEME*, *HEMY*, *DVR* and *CAO* (Figure 6B and C). Moreover, similarly to *A. thaliana*, changes in *M. polymorpha GLK* expression affected Photosystem II (PSII) genes including *psbP, psbR, LHCB1-3* and *LHCB*-6 (Figure 6D and E). In *A. thaliana* GLK targets include Photosystem I (PSI) light harvesting genes *LHCA1*&*4* and *psaD*; in *M. polymorpha* we identified additional GLK targets including *LHCA2&3, psaG, L, O, H, K* and *N* (Figure 6D and E). Additionally, consistent with *A. thaliana* (Tu et al. 2022) we found that genes encoding petD and petC components of the cytochrome b_6f complex were affected by GLK in *M. polymorpha* (Supplemental Figure 13). We conclude that there is significant functional overlap between photosynthesis gene types regulated by GLK in both *M. polymorpha* and *A. thaliana*. Finally, to test whether Mp*GATA4*, Mp*GATA2* and Mp*SCL* act redundantly with Mp*GLK,* we examined if their expression was affected when Mp*GLK* is misregulated. We did not find such evidence as there were no statistically significant changes in Mp*GATA4*, Mp*GATA2* and Mp*SCL* expression in either Mp*GLK* overexpression lines or Mp*glk* mutants (Supplemental Figure 14).

Discussion

Loss of GATAs and SCL alone do not impact on chloroplast biogenesis

The monophyletic group of bryophytes diverged from vascular plants approximately 400MYA. Since then, vascular plants and particularly angiosperms have undergone major morphological and physiological changes including elaborations in chloroplast biogenesis and photosynthesis. Very little is known about the evolution of the underlying genetic networks. We used *M. polymorpha* to test the extent of functional conservation for known regulators of chloroplast biogenesis compared to angiosperms and to shed light on the likely ancestral state.

Our analysis indicated that chloroplast accumulation was not detectably perturbed in mutant alleles of Mp*gata4 or* Mp*gata2*. This contrasts with a ~30-40% decrease in chlorophyll in *A. thaliana gnc*/*cga1* mutants (Zubo et al. 2018), and reduced photomorphogenesis when At*GATA2* was suppressed via RNA silencing or artificial microRNAs (Luo et al. 2010). The simplest and

most parsimonious explanation for the lack of impact of Mp*gata4* or Mp*gata2* loss on chlorophyll content in bryophytes such as *M. polymorpha* is that these proteins do not play a role in chloroplast biogenesis, and that this represents the ancestral state. It is also possible that other proteins compensate and so redundancy in the gene regulatory networks masks loss of function. However, no statistically significant effects of Mp*GLK* overexpression or downregulation on Mp*GATA4* and Mp*GATA2* transcript abundance most likely suggests that Mp*GLK* does not control Mp*GATA4* and Mp*GATA2,* at least at the transcript level. Another possibility is that GATA transcription factors regulate chloroplast biogenesis ancestrally in some but not all bryophytes. This possibility is supported by some limited analysis of Pp*GATA1,* one of the two *GNC/CGA1* orthologs in *P. patens*, where two Pp*GATA1* overexpression lines showed a ~10-20% increase in chlorophyll content (Luan et al. 2023). A similar effect was also reported when Pp*GATA1* was mis-expressed in *A. thaliana* (Luan et al. 2023). However, these increases in chlorophyll content induced by Pp*GATA1* are modest compared with those reported in *A. thaliana* where *GNC* over-expression increased chlorophyll tenfold in seedlings and ~30% in the leaf (Chiang et al. 2012). Moreover, while *GNC/CGA1* overexpression in *A. thaliana* led to an increase in chloroplast number per cell, this was not reported in *P. patens* (Luan et al. 2023). The role of *PpGATA* orthologs (including both Pp*GATA1* and Pp*GATA2*), therefore, needs additional analysis to confirm or refute their role in chloroplast biogenesis. We also note that editing Mp*GATA4* had an effect on gametophyte morphology such that thalli of Mp*gata4* mutants had narrower lobes. This contrasts with *A. thaliana* where to our knowledge there have been no reports of perturbations to leaf morphology in *gnc*/*cga1* mutants. It is therefore possible that the role of this protein has been repurposed from one in affecting development of photosynthetic tissue in *M. polymorpha*, to one in regulating chloroplast biogenesis in *A. thaliana*.

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

409 to GATA2 and GATA4 discussion above, it is also possible that an ancestral role of SCL relates
410 to development of photosynthetic tissue. For SCL6, SCL22 and SCL27 this function seems to be
411 retained in *A. thaliana*, and a role in repressing chlorophyll synthesis has been acquired.
412
413 **Loss of GLK limits chloroplast biogenesis in *M. polymorpha***
414 Unlike *GATAs* and the *MIR171-SCL* module, three lines of evidence indicate that GLK function
415 is conserved in *M. polymorpha*. Firstly, Mp*glk* mutants have significantly reduced chlorophyll
416 accumulation and smaller chloroplasts with underdeveloped thylakoids. These perturbations to
417 phenotype have been observed in other land plants (Bravo et al. 2009; Hall et al. 1998; Rossini
418 et al. 2001; Fitter et al. 2002; Yasumura et al. 2005; Waters et al. 2009). Secondly, constitutive
419 overexpression of Mp*GLK* resulted in increased chlorophyll accumulation, increased chloroplast
420 size and ectopic chloroplast development similar to reports in angiosperms (Kobayashi et al.
421 2013; Nakamura et al. 2009, Powell et al. 2012; Wang et al. 2017, Zubo et al. 2018). Thirdly, our
422 analysis of Mp*GLK* overexpression and mutant RNA transcriptomes revealed an overlap between
423 *GLK* targets in *M. polymorpha* and *A. thaliana* but also rice, tobacco, tomato and maize (Waters
424 et al. 2009; Tu et al. 2022). Most of these shared GLK targets are photosynthesis-related genes
425 such as the *LHCA*, *LHCB*, *PsbQ* and genes encoding chlorophyll biosynthesis enzymes (Tu et al.
426 2022). Finally, and in contrast to previous reports, our phylogenetic analysis identified *GLK*
427 orthologs in two Zygnematophyceae green algae for which genome assemblies became recently
428 available (Cheng et al. 2019). Taken together our results suggest that the GLK function is
429 ancestral to all land plants and the protein appears to have evolved before the transition of plants
430 to terrestrial ecosystems.
431 We also examined how Mp*GLK* is regulated at transcriptional and post-transcriptional levels.
432 First, we asked whether there are any motifs within Mp*GLK* CDS that could cause its
433 downregulation. To do so, we 're-wrote' the coding sequence changing its nucleotide sequence
434 while preserving amino acids. The 're-written' Mp*GLK* did not increase chlorophyll content over
435 and above that in plants over-expressing the native Mp*GLK* CDS. We believe that there are two
436 possible explanations for why Mp*GLK-CDSrw* did not enhance greening more strongly than the
437 native Mp*GLK*-CDS. Firstly, the Mp*GLK* coding sequence does not contain nucleotide sequence
438 motifs responsible for post-transcriptional regulation. Secondly, although we codon-optimised
439 Mp*GLK-CDSrw* for *M. polymorpha*, its expression may be less efficient than that of the native
440 Mp*GLK-CDS*. Finally, we found that although Mp*GLK* is negatively regulated by its 3´ UTR it is
441 unlikely to be regulated by a predicted miRNA/siRNA cleavage site, but rather some as yet
442 unknown motif located between 400 and 671 of Mp*GLK* 3´ UTR.

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

Materials & Methods

Phylogenetic analysis

To identify GATA B-Class and A-Class genes, three different approaches were combined: Firstly, the GATA protein sequences for 21 plant genomes were mined from the iTAK (Zheng et al. 2016) and PlantTFDB databases (Jin et al. 2017), Phytozome, Fernbase, Phycozome and PhytoPlaza. Sequences for each individual species were aligned with the AtGNC and AtCGA1 amino acid sequences using MAFFT (Katoh et al. 2005). Results were filtered manually to identify GNC/CGA1 (B-Class) orthologs distinguished from other GATA family genes by the presence of conserved serine (S) residue, a conserved IRX(R/K)K motif (I: Isoleucine, R: Arginine, X: any amino acid and K: Lysine), and the presence or absence of conserved LLM- (leucine– leucine–methionine) domain at their C-terminus (Behringer et al. 2014). GATA2 (A-Class) orthologs, were distinguished by the presence of a conserved glutamine (Q) and a threonine (T) within the zinc finger motif. Secondly, we performed BLASTP searches against plant genomes on Phytozome v13, fern genomes (fernbase.org), hornworts genome (hwww.hornworts.uzh.ch), green algae genomes on PhycoCosm (/phycocosm.jgi.doe.gov), and 1KP using the AtGNC/CGA1 amino acid

sequence as a query. Results were filtered manually as described above. Finally, the combined results from the above to approaches were checked against Orthofinder searches (Emms and Kelly, 2019). The identified GATA protein sequences were aligned using MAFFT. Alignments were then trimmed using TrimAl (Capella-Gutiérrez et al., 2009). A maximum likelihood phylogenetic tree was inferred using iQTree, ModelFinder (Kalyaanamoorthy et a 2017) and ultrafast approximation for phylogenetic bootstrap (Hoang et al. 2018) and SH-aLRT test (Guindon et al. 2010). The tree was visualised using iTOL.

To identify *GLK* genes, three different approaches were combined: Firstly, the G2-GARP protein sequences for 21 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 AtGLK1/2 amino acid sequences using MAFFT (Katoh et al. 2019). Results were filtered manually to identify GLK orthologs distinguished from other G2-GARP family genes by three characteristic motifs: AREAEAA motif (consensus motif) in the DNA-binding domain, VWG(Y/H)P and the PLGL(R/K)(P/S)P in the GCT-box domain. Secondly, we performed BLASTP searches against plant genomes on Phytozome v13, fern genomes (fernbase.org), hornworts genome (hwww.hornworts.uzh.ch), green algae genomes on PhycoCosm (/phycocosm.jgi.doe.gov), and 1KP (One Thousand Plant Transcriptomes Initiative 2019) using the AtGLK1/2 amino acid sequence as a query. Results were filtered manually as described above. Finally, the combined results from the above two approaches were checked against Orthofinder searches (Emms and Kelly 2019). The identified GLK protein sequences were aligned using MAFFT. Alignments were then trimmed using TrimAl (Capella-Gutiérrez et al. 2009). A maximum likelihood phylogenetic tree was inferred using iQTree (Nguyen et al. 2015), ModelFinder (Kalyaanamoorthy et al. 2017) and ultrafast approximation for phylogenetic bootstrap (Hoang et al. 2018) and SH-aLRT test (Guindon et al. 2010). The tree was visualised using iTOL (Letunic and Bork 2021) .

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

Marchantia polymorpha accessions Cam-1 (male) and Cam-2 (female) were used in this study (Delmans et al. 2017). 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 (Sauret-Güeto et al. 2020).

For CRISPR/Cas9 gene editing, guide RNAs were *in silico* predicted using CasFinder tool (https://marchantia.info/tools/casfinder/). Several gRNAs per target gene were *in vitro* tested as described (Yelina et al. 2022) using oligonucleotides in Supplemental Tables File 3. gRNA sequences that were selected to generate Mp*glk*, Mp*gata2*, Mp*gata4*, Mp*scl* and Mp*mir171* mutants are listed in Supplemental Tables File 3. Single gRNA7 to mutate Mp*GLK* was cloned using OpenPlant toolkit (Sauret-Güeto et al. 2020). Multiple gRNAs to mutate Mp*GATA4*, Mp*SCL* and Mp*MIR171* and a single gRNA to mutate Mp*GATA2* were cloned as described in Yelina et al. 2022, using oligonucleotides listed in Supplemental Tables File 3 and the destination vector pMpGE013 (Sugano et al. 2018). For the overexpression constructs, Mp*GLK and* Mp*GLKrw* CDS were synthesised (Integrated DNA Technologies), Mp*GLK* 3´UTR was amplified from *M. polymorpha* genomic DNA and cloned into the pUAP4 vector (Sauret-Güeto et al. 2020). Constructs were generated using the OpenPlant toolkit (Sauret-Güeto et al. 2020). OpenPlant parts used: OP-023 CDS12-eGFP, OP-020 CDS_hph, OP-037 CTAG_Lti6b, OP-054 3TER, _Nos-35S, OP-049 PROM_35S, OP-47 PROM_MpUBE2, OP-48 5UTR_MpUBE2, OP-063, L1_HyR-Ck1, OP073 L1_Cas9-Ck4, OP-076 L2_lacZgRNA-Cas9-CsA and L1_35S_s::eGFP-Lti6b (Frangedakis et al. 2021). Nucleotide sequence of Mp*GLKrw* CDS and oligonucleotide sequences used for cloning are listed in Supplemental Tables File 3.

Chlorophyll determination, fluorescence measurements and imaging analysis

For chlorophyll measurements, ~30-50mg of 10-14 days old gemmalings were used, with five biological replicates per genotype. The tissue was blotted on tissue paper before weighing to remove excess water and then was transferred into a 1.5mL Eppendorf tube containing 1 mL of dimethyl sulfoxide (DMSO) (D8418, Sigma Aldrich) and incubated in the dark at 65C with 300 rpm shaking for 45 minutes. Samples were let to cool down to room temperature for approximately one hour. Chlorophyll content was measured using a NanoDrop™One/One C Microvolume UV-Vis Spectrophotometer (ThermoFisher) following the manufacturer's protocol. Chlorophyll fluorescence measurements were carried out using a CF imager (Technologica Ltd, UK) and the image processing software provided by the manufacturer, as described previously (Schreier et al. 2022). *M. polymorpha* plants were placed in the dark for 20 min for dark adaptation to evaluate the dark-adapted minimum fluorescence (Fo), dark-adapted maximum fluorescence (*Fm*), variable fluorescence *Fv* (*Fv* = *Fm*–*Fo*). All chlorophyll fluorescence images within each experiment were acquired at the same time in a single image, measuring a total of three plants per genotype and treatment. For the DCMU treatment, 20 µM DCMU (#45463, Sigma Aldrich) was added to the half-strength MS media before it was poured into the individual petri dishes.

Thalli were placed for 24 h onto the DCMU-containing media before chlorophyll fluorescence measurements.

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

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

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

500 ng of DNase-treated RNA was used as a template for cDNA preparation using the SuperScript™ IV First-Strand Synthesis System (#18091050, Invitrogen) according to manufacturer's instructions (with only modifying reverse transcriptase reaction time to 40 minutes and using oligo (dT)18 primers). qPCR was performed using the SYBR Green JumpStart *Taq* Ready Mix (#S4438, Sigma Aldrich) and a CFX384 RT System (Bio-Rad) thermal cycler. cDNA was diluted 6 times, oligonucleotides listed in Supplemental Tables file 3 were used at a final concentration of 0.5 µM and reaction conditions were as follows: initial denaturation step of 94°C for 2 minutes followed by 40 cycles of 94°C for 15 seconds (denaturation) and 60°C for 1 minute (annealing, extension, and fluorescence reading).

Library preparation and RNA sequencing was performed by Novogene (Cambridge, UK). Briefly, messenger RNA was purified from total RNA using poly-T oligo-attached magnetic beads. After fragmentation, the first strand cDNA was synthesized using random hexamer primers followed by the second strand cDNA synthesis. cDNA end repair, A-tailing, adapter ligation, size selection, amplification, and purification were performed next. Library concentration was measured on a Qubit instrument following the manufacturer's procedure (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 6 G raw data per sample were obtained. Adapter sequences were: 5' Adapter: 5'-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-3'. 3' Adapter: 5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGATGACTATCTCGTATGCCGTCTTCTGCTTG-3'

FastQC was used to assess read quality and TrimGalore (https://doi.org/10.5281/zenodo.5127899) to trim low-quality reads and remaining sequencing adapters. Reads were pseudo-aligned using Kallisto (Bray et al. 2016) to the *M. polymorpha* Genome version 5 (primary transcripts only, obtained from MarpolBase) (Bowman et al. 2018). Kallisto estimates abundance of each transcript in units of transcripts per million (TPM). Mapping statistics for each library are provided in Supplemental Tables file 2. DGE analysis was performed with DESeq2 (Love et al. 2014), with padj-values < 0.01, Supplemental Tables file 2. Plots were generated using R.

Author contributions

N.Y.E., E.F., T.S., M.T. and J.R. carried out the work. N.Y.E., E.F. and J.M.H. designed the work. N.Y.E, E.F. and J.M.H. wrote the manuscript. J.M.H., N.Y.E. and J.M.H. initiated and oversaw the project.

Acknowledgements

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

REFERENCES

Archibald, John M. 2009. "The Puzzle of Plastid Evolution." *Current Biology*. 27;19(2):R81-8.

Behringer, Carina, and Claus Schwechheimer. 2015. "B-GATA Transcription Factors – Insights into Their Structure, Regulation, and Role in Plant Development." *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2015.00090.

Bi, Yong-Mei, Yu Zhang, Tara Signorelli, Rong Zhao, Tong Zhu, and Steven Rothstein. 2005. "Genetic Analysis of Arabidopsis GATA Transcription Factor Gene Family Reveals a Nitrate-Inducible Member Important for Chlorophyll Synthesis and Glucose Sensitivity." *The Plant Journal*.

Bolle, Cordelia. 2004. "The Role of GRAS Proteins in Plant Signal Transduction and Development." *Planta* 218 (5): 683–92.

Bowman, John L., Mario Arteaga-Vazquez, Frederic Berger, Liam N. Briginshaw, Philip Carella, Adolfo Aguilar-Cruz, Kevin M. Davies, et al. 2022. "The Renaissance and Enlightenment of Marchantia as a Model System." *The Plant Cell* 34 (10): 3512–42.

Bravo-Garcia, Armando, Yuki Yasumura, and Jane A. Langdale. 2009. "Specialization of the Golden2-like Regulatory Pathway during Land Plant Evolution." *The New Phytologist* 183 (1): 133–41.

Bray, Nicolas L., Harold Pimentel, Páll Melsted, and Lior Pachter. 2016. "Near-Optimal Probabilistic RNA-Seq Quantification." *Nature Biotechnology* 34, 525–527.

Cackett, Lee, Leonie H. Luginbuehl, Tina B. Schreier, Enrique Lopez-Juez, and Julian M. Hibberd. 2022. "Chloroplast Development in Green Plant Tissues: The Interplay between Light, Hormone, and Transcriptional Regulation." *New Phytologist*. 233 (5), 2000-2016.

Capella-Gutiérrez, Salvador, José M. Silla-Martínez, and Toni Gabaldón. 2009. "trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." *Bioinformatics* 25 (15): 1972–73.

Cheng, Shifeng, Wenfei Xian, Yuan Fu, Birger Marin, Jean Keller, Tian Wu, Wenjing Sun, et al. 2019. "Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant Evolution." *Cell* 179 (5): 1057–67.e14.

Chiang, Yi-Hsuan, Yan O. Zubo, Wiebke Tapken, Hyo Jung Kim, Ann M. Lavanway, Louisa Howard, Marinus Pilon, Joseph J. Kieber, and G. Eric Schaller. 2012. "Functional Characterization of the GATA Transcription Factors GNC and CGA1 Reveals Their Key Role

643 in Chloroplast Development, Growth, and Division in Arabidopsis." *Plant Physiology* 160 (1):
644 332–48.
645 Delmans, Mihails, Bernardo Pollak, and Jim Haseloff. 2017. "MarpoDB: An Open Registry for
646 Marchantia Polymorpha Genetic Parts." *Plant & Cell Physiology* 58 (1): e5.
647 Emms, David M., and Steven Kelly. 2019. "OrthoFinder: Phylogenetic Orthology Inference for
648 Comparative Genomics." *Genome Biology* 20 (1): 238.
649 Fitter, David W., David J. Martin, Martin J. Copley, Robert W. Scotland, and Jane A. Langdale.
650 2002. "*GLK* gene Pairs Regulate Chloroplast Development in Diverse Plant Species." *The*
651 *Plant Journal*. **31(6):713-27**
652 Frangedakis, Eftychios, Manuel Waller, Tomoaki Nishiyama, Hirokazu Tsukaya, Xia Xu, Yuling Yue,
653 Michelle Tjahjadi, et al. 2021. "An Agrobacterium-Mediated Stable Transformation Technique
654 for the Hornwort Model Anthoceros Agrestis." *The New Phytologist* 232 (3): 1488–1505.
655 Gould, Sven B., Ross F. Waller, and Geoffrey I. McFadden. 2008. "Plastid Evolution." *Annual*
656 *Review of Plant Biology* 59: 491–517.
657 Guindon, Stéphane, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and
658 Olivier Gascuel. 2010. "New Algorithms and Methods to Estimate Maximum-Likelihood
659 Phylogenies: Assessing the Performance of PhyML 3.0." *Systematic Biology* 59 (3): 307–21.
660 Hall, L. N., L. Rossini, L. Cribb, and J. A. Langdale. 1998. "GOLDEN 2: A Novel Transcriptional
661 Regulator of Cellular Differentiation in the Maize Leaf." *The Plant Cell* 10 (6): 925–36.
662 Harris, Brogan J., C. Jill Harrison, Alistair M. Hetherington, and Tom A. Williams. 2020.
663 "Phylogenomic Evidence for the Monophyly of Bryophytes and the Reductive Evolution of
664 Stomata." *Current Biology: CB* 30 (11): 2001–12.e2.
665 Hoang, Diep Thi, Olga Chernomor, Arndt von Haeseler, Bui Quang Minh, and Le Sy Vinh. 2018.
666 "UFBoot2: Improving the Ultrafast Bootstrap Approximation." *Molecular Biology and*
667 *Evolution* 35 (2): 518–22.
668 Hudson, Darryl, David Guevara, Mahmoud W. Yaish, Carol Hannam, Nykoll Long, Joseph D.
669 Clarke, Yong-Mei Bi, and Steven J. Rothstein. 2011. "GNC and CGA1 Modulate Chlorophyll
670 Biosynthesis and Glutamate Synthase (GLU1/Fd-GOGAT) Expression in Arabidopsis." *PloS*
671 *One* 6 (11): e26765.
672 Jarvis, Paul, and Enrique López-Juez. 2013. "Biogenesis and Homeostasis of Chloroplasts and
673 Other Plastids." *Nature Reviews. Molecular Cell Biology* 14 (12): 787–802.

Jin, Jinpu, Feng Tian, De-Chang Yang, Yu-Qi Meng, Lei Kong, Jingchu Luo, and Ge Gao. 2017. "PlantTFDB 4.0: Toward a Central Hub for Transcription Factors and Regulatory Interactions in Plants." *Nucleic Acids Research* 45 (D1): D1040–45.

Kalyaanamoorthy, Subha, Bui Quang Minh, Thomas K. F. Wong, Arndt von Haeseler, and Lars S. Jermiin. 2017. "ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates." *Nature Methods* 14 (6): 587–89.

Katoh, Kazutaka, John Rozewicki, and Kazunori D. Yamada. 2019. "MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization." *Briefings in Bioinformatics* 20 (4): 1160–66.

Kobayashi, Koichi, Daichi Sasaki, Ko Noguchi, Daiki Fujinuma, Hirohisa Komatsu, Masami Kobayashi, Mayuko Sato, et al. 2013. "Photosynthesis of Root Chloroplasts Developed in Arabidopsis Lines Overexpressing GOLDEN2-LIKE Transcription Factors." *Plant & Cell Physiology* 54 (8): 1365–77.

Letunic, Ivica, and Peer Bork. 2021. "Interactive Tree Of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation." *Nucleic Acids Research* 49 (W1): W293–96.

Li, Fay-Wei, Paul Brouwer, Lorenzo Carretero-Paulet, Shifeng Cheng, Jan de Vries, Pierre-Marc Delaux, Ariana Eily, et al. 2018. "Fern Genomes Elucidate Land Plant Evolution and Cyanobacterial Symbioses." *Nature Plants* 4 (7): 460–72.

Li, Fay-Wei, Tomoaki Nishiyama, Manuel Waller, Eftychios Frangedakis, Jean Keller, Zheng Li, Noe Fernandez-Pozo, et al. 2020. "Anthoceros Genomes Illuminate the Origin of Land Plants and the Unique Biology of Hornworts." *Nature Plants* 6 (3): 259–72.

Lin, Shih-Shun, and John L. Bowman. 2018. "MicroRNAs in Marchantia Polymorpha." *The New Phytologist* 220 (2): 409–16.

Llave, Cesar, Kristin D. Kasschau, Maggie A. Rector, and James C. Carrington. 2002. "Endogenous and Silencing-Associated Small RNAs in Plants." *The Plant Cell* 14 (7): 1605–19.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550.

Luan, Ji, Jianfang Ju, Xiaochen Li, Xiuling Wang, Yufei Tan, and Guangmin Xia. 2022. "Functional Identification of Moss PpGATA1 Provides Insights into the Evolution of LLM-Domain B-GATA Transcription Factors in Plants." *Gene* 855: 147103.

Luo, Xiao-Min, Wen-Hui Lin, Shengwei Zhu, Jia-Ying Zhu, Yu Sun, Xi-Ying Fan, Menglin Cheng, et al. 2010. "Integration of Light- and Brassinosteroid-Signaling Pathways by a GATA Transcription Factor in Arabidopsis." *Developmental Cell* 19 (6): 872–83.

Ma, Zhaoxue, Xupeng Hu, Wenjuan Cai, Weihua Huang, Xin Zhou, Qian Luo, Hongquan Yang, Jiawei Wang, and Jirong Huang. 2014. "Arabidopsis miR171-Targeted Scarecrow-Like Proteins Bind to GT Cis-Elements and Mediate Gibberellin-Regulated Chlorophyll Biosynthesis under Light Conditions." *PLoS Genetics* https://doi.org/10.1371/journal.pgen.1004519.

Naito, Takahito, Takatoshi Kiba, Nobuya Koizumi, Takafumi Yamashino, and Takeshi Mizuno. 2007. "Characterization of a Unique GATA Family Gene That Responds to Both Light and Cytokinin in Arabidopsis Thaliana." *Bioscience, Biotechnology, and Biochemistry* 71 (6): 1557–60.

Nakamura, Hidemitsu, Masayuki Muramatsu, Makoto Hakata, Osamu Ueno, Yoshiaki Nagamura, Hirohiko Hirochika, Makoto Takano, and Hiroaki Ichikawa. 2009. "Ectopic Overexpression of the Transcription Factor OsGLK1 Induces Chloroplast Development in Non-Green Rice Cells." *Plant & Cell Physiology* 50 (11): 1933–49.

Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32 (1): 268–74.

One Thousand Plant Transcriptomes Initiative. 2019. "One Thousand Plant Transcriptomes and the Phylogenomics of Green Plants." *Nature* 574 (7780): 679–85.

Powell, Ann L. T., Cuong V. Nguyen, Theresa Hill, Kalai Lam Cheng, Rosa Figueroa-Balderas, Hakan Aktas, Hamid Ashrafi, et al. 2012. "Uniform Ripening Encodes a Golden 2-like Transcription Factor Regulating Tomato Fruit Chloroplast Development." *Science* 336 (6089): 1711–15.

Qiu, Yin-Long, Libo Li, Bin Wang, Zhiduan Chen, Volker Knoop, Milena Groth-Malonek, Olena Dombrovska, et al. 2006. "The Deepest Divergences in Land Plants Inferred from Phylogenomic Evidence." *Proceedings of the National Academy of Sciences of the United States of America* 103 (42): 15511–16.

Rhoades, Matthew W., Brenda J. Reinhart, Lee P. Lim, Christopher B. Burge, Bonnie Bartel, and David P. Bartel. 2002. "Prediction of Plant microRNA Targets." *Cell* 110 (4): 513–20.

Rich, Mélanie K., and Pierre-Marc Delaux. 2020. "Plant Evolution: When Arabidopsis Is More Ancestral Than Marchantia." *Current Biology: CB*.

Rossini, L., L. Cribb, D. J. Martin, and J. A. Langdale. 2001. "The Maize golden2 Gene Defines a Novel Class of Transcriptional Regulators in Plants." *The Plant Cell* 13 (5): 1231–44.

Safi, Alaeddine, Anna Medici, Wojciech Szponarski, Sandrine Ruffel, Benoît Lacombe, and Gabriel Krouk. 2017. "The World according to GARP Transcription Factors." *Current Opinion in Plant Biology* 39: 159–67.

Saint-Marcoux D, Proust H, Dolan L, Langdale JA (2015) Identification of Reference Genes for Real-Time Quantitative PCR Experiments in the Liverwort *Marchantia polymorpha*. PLoS ONE 10(3): e0118678.

Sauret-Güeto, Susanna, Eftychios Frangedakis, Linda Silvestri, Marius Rebmann, Marta Tomaselli, Kasey Markel, Mihails Delmans, Anthony West, Nicola J. Patron, and Jim Haseloff. 2020. "Systematic Tools for Reprogramming Plant Gene Expression in a Simple Model,." *ACS Synthetic Biology* 9 (4): 864–82.

Schreier, Tina B., Karin H. Müller, Simona Eicke, Christine Faulkner, Samuel C. Zeeman, and Julian M. Hibberd. n.d. "Plasmodesmal Connectivity in C_4*Gynandropsis Gynandra*is Induced by Light and Dependent on Photosynthesis." https://doi.org/10.1101/2022.12.07.519530.

Schulze, Silke, Barbara Nicole Schäfer, Eneida Abreu Parizotto, Olivier Voinnet, and Klaus Theres. 2010. "LOST MERISTEMS Genes Regulate Cell Differentiation of Central Zone Descendants in Arabidopsis Shoot Meristems." *The Plant Journal: For Cell and Molecular Biology* 64 (4): 668–78.

Sugano, Shigeo S., Ryuichi Nishihama, Makoto Shirakawa, Junpei Takagi, Yoriko Matsuda, Sakiko Ishida, Tomoo Shimada, Ikuko Hara-Nishimura, Keishi Osakabe, and Takayuki Kohchi. 2018. "Efficient CRISPR/Cas9-Based Genome Editing and Its Application to Conditional Genetic Analysis in Marchantia Polymorpha." *PloS One* 13 (10): e0205117.

Todesco, Marco, Ignacio Rubio-Somoza, Javier Paz-Ares, and Detlef Weigel. 2010. "A Collection of Target Mimics for Comprehensive Analysis of microRNA Function in Arabidopsis Thaliana." *PLoS Genetics* 6 (7): e1001031.

Trebst, Achim. 2007. "Inhibitors in the Functional Dissection of the Photosynthetic Electron Transport System." *Photosynthesis Research* 92 (2): 217–24.

766 Tu, Xiaoyu, María Katherine Mejía-Guerra, Jose A. Valdes Franco, David Tzeng, Po-Yu Chu, Wei
767 Shen, Yingying Wei, et al. 2020. "Reconstructing the Maize Leaf Regulatory Network Using
768 ChIP-Seq Data of 104 Transcription Factors." *Nature Communications* 11 (1): 5089.
769 Tu, Xiaoyu, Sibo Ren, Wei Shen, Jianjian Li, Yuxiang Li, Chuanshun Li, Yangmeihui Li, et al. 2022.
770 "Limited Conservation in Cross-Species Comparison of GLK Transcription Factor Binding
771 Suggested Wide-Spread Cistrome Divergence." *Nature Communications* 13 (1): 7632.
772 Vries, Jan de, and John M. Archibald. 2018. "Plant Evolution: Landmarks on the Path to Terrestrial
773 Life." *The New Phytologist* 217 (4): 1428–34.
774 Vries, Jan de, and Sven B. Gould. 2018. "The Monoplastidic Bottleneck in Algae and Plant
775 Evolution." *Journal of Cell Science* 131 (2).
776 Wang, Long, Yan-Xia Mai, Yan-Chun Zhang, Qian Luo, and Hong-Quan Yang. 2010.
777 "MicroRNA171c-Targeted SCL6-II, SCL6-III, and SCL6-IV Genes Regulate Shoot Branching
778 in Arabidopsis." *Molecular Plant* 3 (5): 794–806.
779 Wang, Peng, Roxana Khoshravesh, Shanta Karki, Ronald Tapia, C. Paolo Balahadia, Anindya
780 Bandyopadhyay, W. Paul Quick, Robert Furbank, Tammy L. Sage, and Jane A. Langdale.
781 2017. "Re-Creation of a Key Step in the Evolutionary Switch from C to C Leaf Anatomy."
782 *Current Biology: CB* 27 (21): 3278–87.e6.
783 Waters, Mark T., Elizabeth C. Moylan, and Jane A. Langdale. 2008. "GLK Transcription Factors
784 Regulate Chloroplast Development in a Cell-Autonomous Manner." *The Plant Journal: For*
785 *Cell and Molecular Biology* 56 (3): 432–44.
786 Waters, Mark T., Peng Wang, Muris Korkaric, Richard G. Capper, Nigel J. Saunders, and Jane A.
787 Langdale. 2009. "GLK Transcription Factors Coordinate Expression of the Photosynthetic
788 Apparatus in Arabidopsis." *The Plant Cell* 21 (4): 1109–28.
789 Yasumura, Yuki, Elizabeth C. Moylan, and Jane A. Langdale. 2005. "A Conserved Transcription
790 Factor Mediates Nuclear Control of Organelle Biogenesis in Anciently Diverged Land Plants."
791 *The Plant Cell* 17 (7): 1894–1907.
792 Yelina, Nataliya E., Daniel Holland, Sabrina Gonzalez-Jorge, Dominique Hirsz, Ziyi Yang, and Ian
793 R. Henderson. 2022. "Coexpression of MEIOTIC-TOPOISOMERASE VIB-dCas9 with Guide
794 RNAs Specific to a Recombination Hotspot Is Insufficient to Increase Crossover Frequency
795 in Arabidopsis." *G3* 12 (7).
796 Zheng, Yi, Chen Jiao, Honghe Sun, Hernan G. Rosli, Marina A. Pombo, Peifen Zhang, Michael
797 Banf, et al. 2016. "iTAK: A Program for Genome-Wide Prediction and Classification of Plant

Transcription Factors, Transcriptional Regulators, and Protein Kinases." *Molecular Plant* 9 (12): 1667–70.

Zubo, Yan O., Ivory Clabaugh Blakley, José M. Franco-Zorrilla, Maria V. Yamburenko, Roberto Solano, Joseph J. Kieber, Ann E. Loraine, and G. Eric Schaller. 2018. "Coordination of Chloroplast Development through the Action of the GNC and GLK Transcription Factor Families." *Plant Physiology* 178 (1): 130–47.

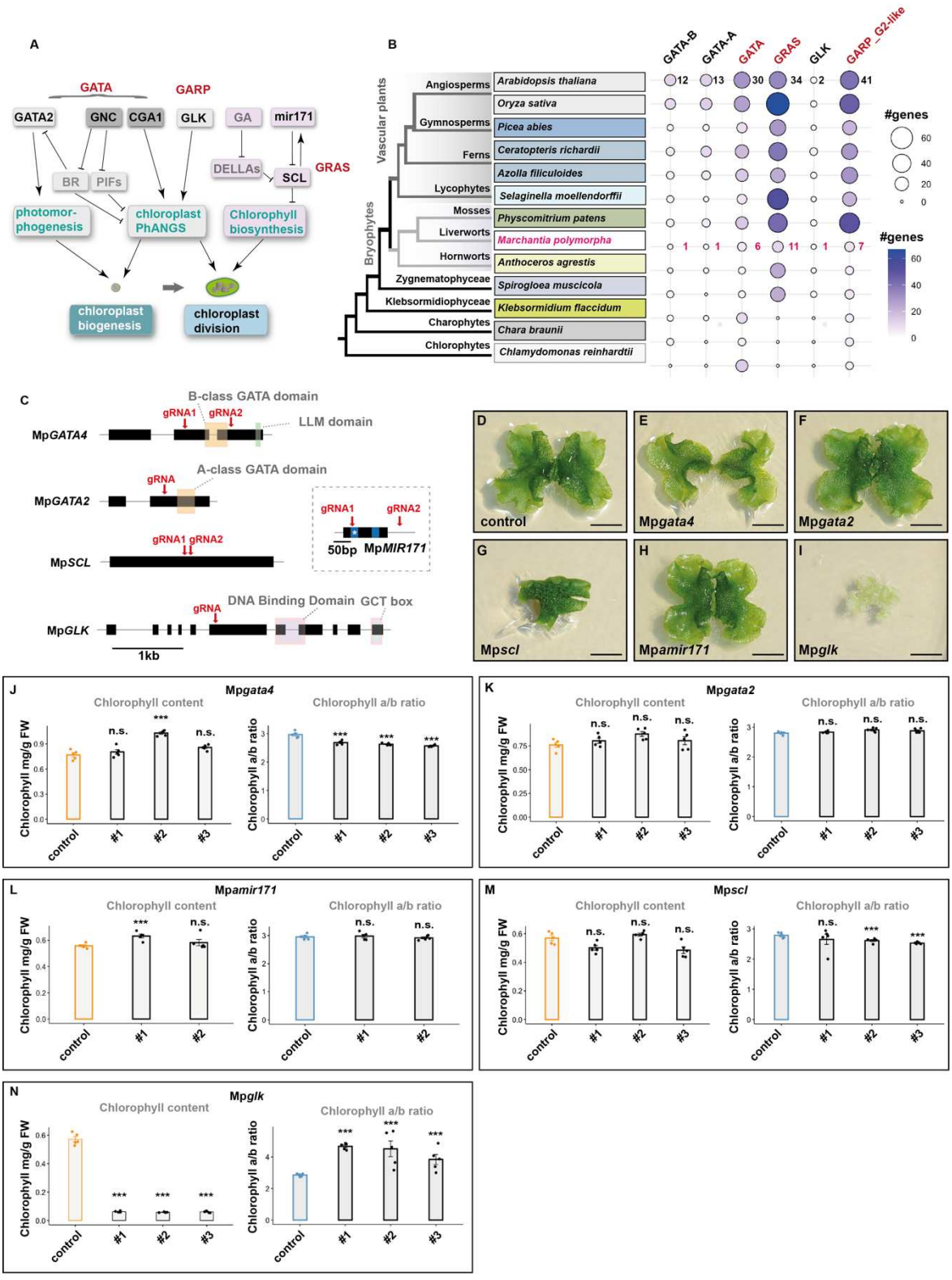

Figure 1: Mp*GLK* controls chlorophyll. A) Transcription factors known to regulate chloroplast development in angiosperms. GATA TFs - GATA NITRATE- INDUCIBLE CARBON METABOLISM-INVOLVED (GNC) and CYTOKININ- RESPONSIVE GATA FACTOR 1 (CGA1) control chloroplast development by suppressing phytochrome interacting factors and brassinosteroid (BR) related genes, promoting chloroplast biogenesis and division. GATA2 promotes photomorphogenesis in the presence of light by directly binding to light-responsive promoters. In the absence of light, the BR-activated transcription factor BRASSINAZOLE RESISTANT1 (BZR1) represses GATA2 expression, inhibiting photomorphogenesis. GOLDEN2-LIKE transcriptions factors (TFs) are positive regulators of nuclear-encoded photosynthesis related genes. SCARECROW-LIKE (SCL) GRAS TFs are negatively regulated by miR171 and GA-DELLA signalling, to control chlorophyll biosynthesis. B) Phylogenetic relationships of the major lineages of land plants and green algae (Li et al. 2020). C) Schematic representation of Mp*GLK,* Mp*GATA4,* Mp*GATA2* and Mp*SCL* gene structure showing exons as black rectangles. Characteristic gene domains are shown as coloured boxes. Mp*MIR171* genomic locus is shown in a dashed box. Mp*MIR171* gene is represented as a black rectangle, miR171* and miR171 shown as blue rectangles, miR171* indicated with a white star. gRNAs positions for CRISPR/Cas9 gene editing are shown as red arrows. D-I) Images of wild type control, Mp*gata4*, Mp*gata2*, Mp*scl*, Mp*mir171* and Mp*glk* mutant plants. Scale bars represent 2 mm. J-N) Barplots of chlorophyll content and chlorophyll a/b ratio for the Mp*gata4*, Mp*gata2*, Mp*scl*, Mp*mir171* and Mp*glk* mutant plants. Individual values are shown with dots. Error bars represent the standard deviation of the mean $n = 5$. Asterisks indicate statistically significant difference using a two-tailed *t*-test $P \leq 0.0001$ (***)

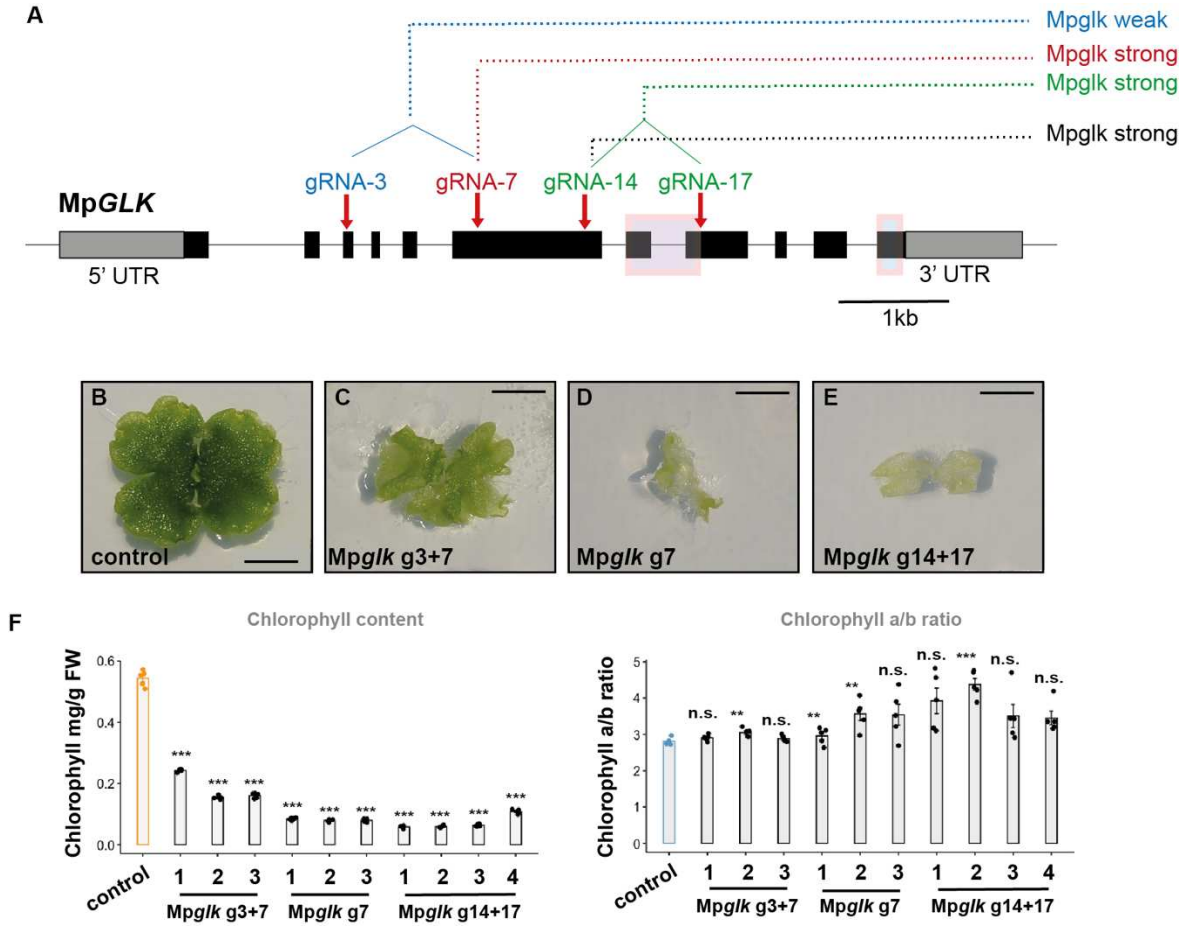

Figure 2: Mp*GLK* mutants. A) Schematic representation of Mp*GLK* gene structure showing exons as black rectangles, untranslated regions (UTRs) as grey rectangles and introns as black lines. Characteristic gene domains are highlighted by shaded boxes. Positions of gRNAs used for CRISPR/Cas9 gene editing are shown as red arrows. B-E) Images of control and Mp*glk* mutant plants. Scale bars represent 2mm. F) Barplots representing chlorophyll content and chlorophyll a:b ratio in Mp*glk* mutants. Individual values are shown as dots. Error bars, represent standard deviation of the mean form $n = 5$. Asterisks indicate statistically significant differences using a two-tailed *t*-test $P \leq 0.0001$ (***), $0.0001 \leq P \leq 0.001$ (**) and n.s.: non-significant.

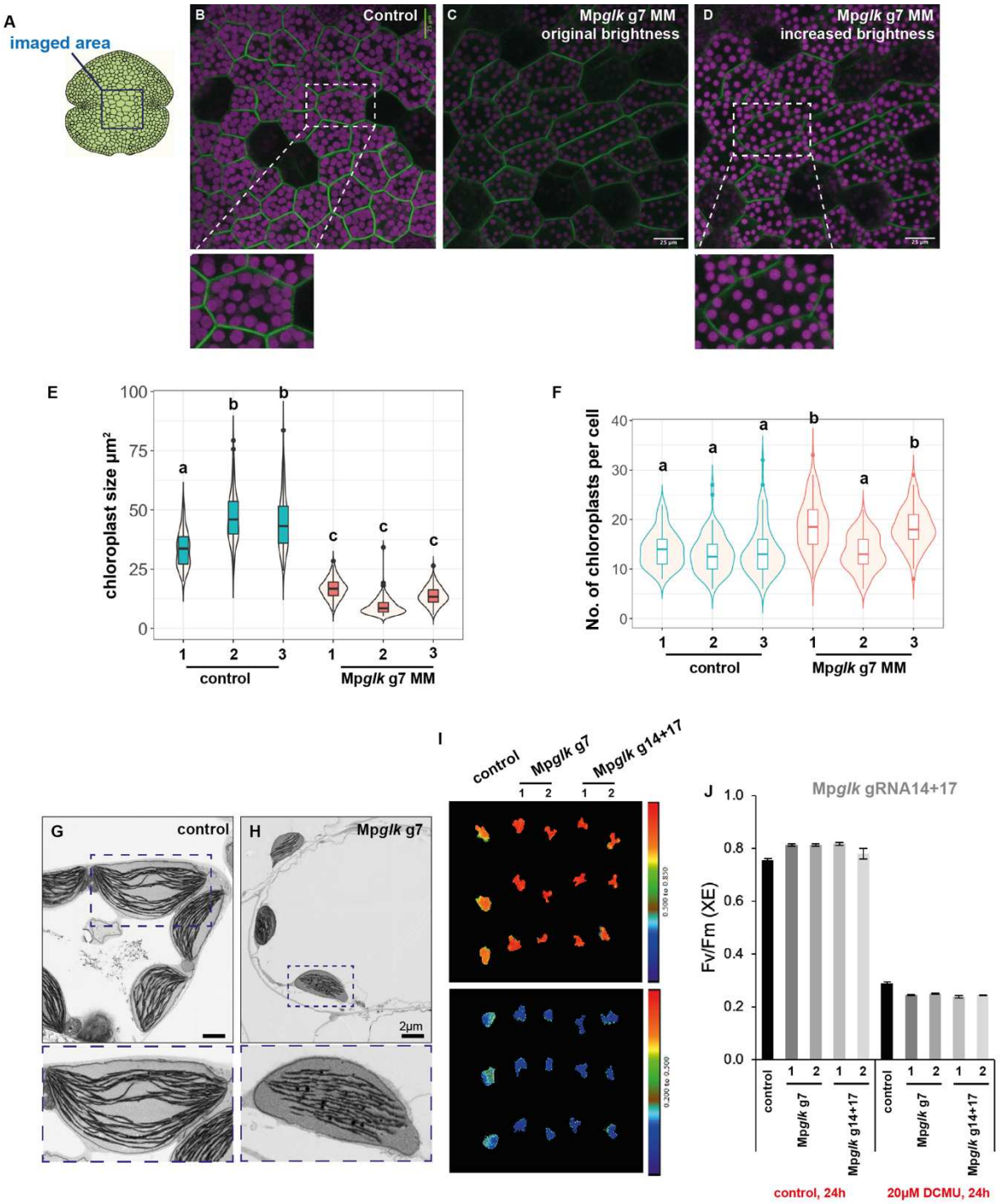

Figure 3: Mp*glk* controls chloroplast size and ultrastructure. A) Schematic of a gemmae, a dark blue square indicates imaged area. B-D) Representative confocal microscopy images of control (B) and *Mp*glk mutant (C, D) plant gemmae. Chlorophyll autofluorescence is shown in magenta and plasma membrane marked with eGFP shown in green. Magnified cells are shown at the bottom of panels B and D. Scale bars, represent 25 µm. E) Boxplot showing chloroplast size ranges in control and Mp*glk* plants. F) Boxplot showing chloroplast numbers per cell area in control and Mp*glk* plants. 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. G, H) Scanning electron micrographs of chloroplasts in control (G) and Mp*glk* mutant (H) plants. Dashed boxes highlight single chloroplasts that are shown in the insets at a higher magnification (bottom panels). Scale bars represent 2 µm. I) Chlorophyll fluorescence images of maximum quantum efficiency of Photosystem II photochemistry (*Fv/Fm*) of untreated (top) and 24 h DCMU-treated (bottom) control and Mp*glk* mutant plants. J) *Fv/Fm* measured in DCMU-treated and untreated control and Mp*glk* mutant plants. Bars represent mean ± standard error from $n = 3$ plants per genotype.

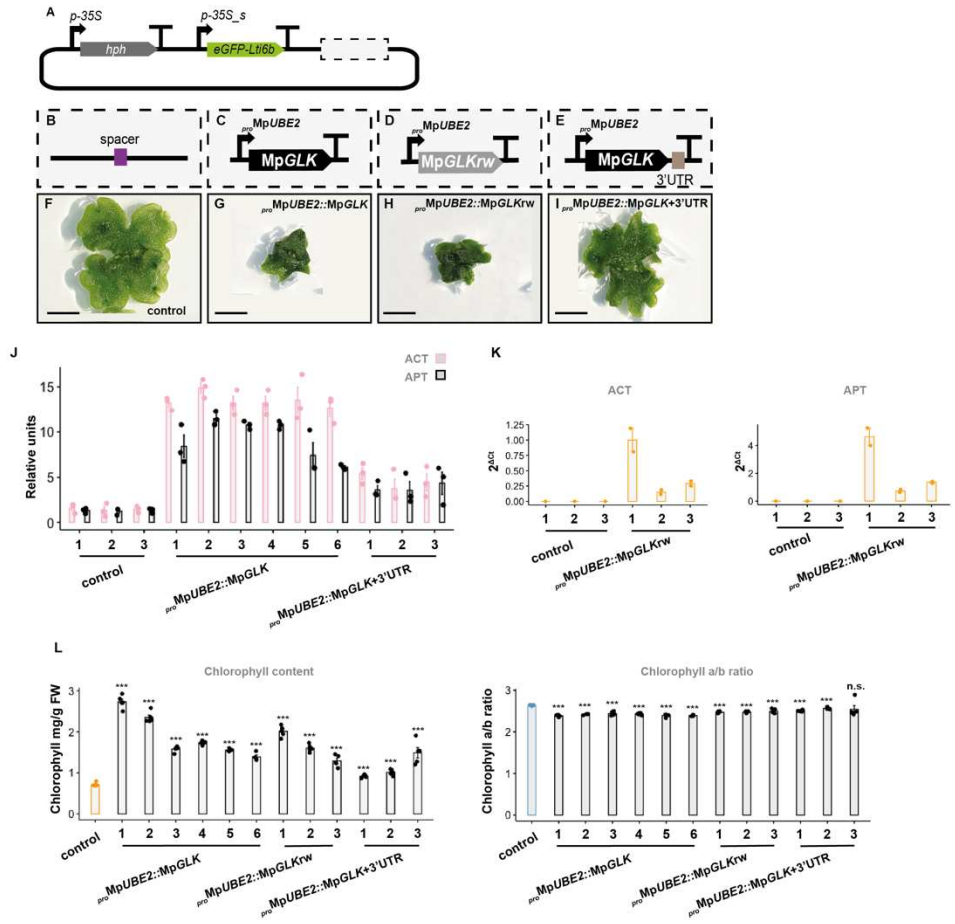

Figure 4: Mp*GLK* overexpression increases chlorophyll content. A-E) Schematic representation of transformation constructs used for Mp*GLK-CDS*, Mp*GLK-CDSrw* and Mp*GLK-CDS+3´UTR* overexpression as well as the 'empty vector' control. hph: hygromycin B phosphotransferase, 35S: cauliflower mosaic virus (CaMV) 35S promoter. F-I) Images of 'empty vector' control plants and plants transformed with Mp*GLK-CDS*, Mp*GLK-CDSrw* and Mp*GLK-CDS+3'UTR* overexpression constructs. Scale bars represent 2mm. J) Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis of Mp*GLK* levels in Mp*GLK-CDS* and Mp*GLK-CDS+3'UTR* overexpressing lines. (K) Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis of Mp*GLK*-CDSrw levels in Mp*GLK* CDSrw overexpressing lines. *ADENINE PHOSPHORIBOSYL TRANSFERASE 3* (*APT*) and *ACTIN 7* (ACT) were used as housekeeping gene controls (Saint-Marcoux et al. 2015). L) Barplots of chlorophyll content and chlorophyll a/b ratios in Mp*GLK-CDS*, Mp*GLK-CDSrw* and Mp*GLK-CDS+3'UTR* overexpressing lines compared to 'empty vector' controls. Individual values are shown with dots. Error bars, represent standard deviation of the mean from n = 3. Asterisks indicate statistically significant difference using a two-tailed t-test $P \leq 0.0001$ (***) and n.s.: non-significant.

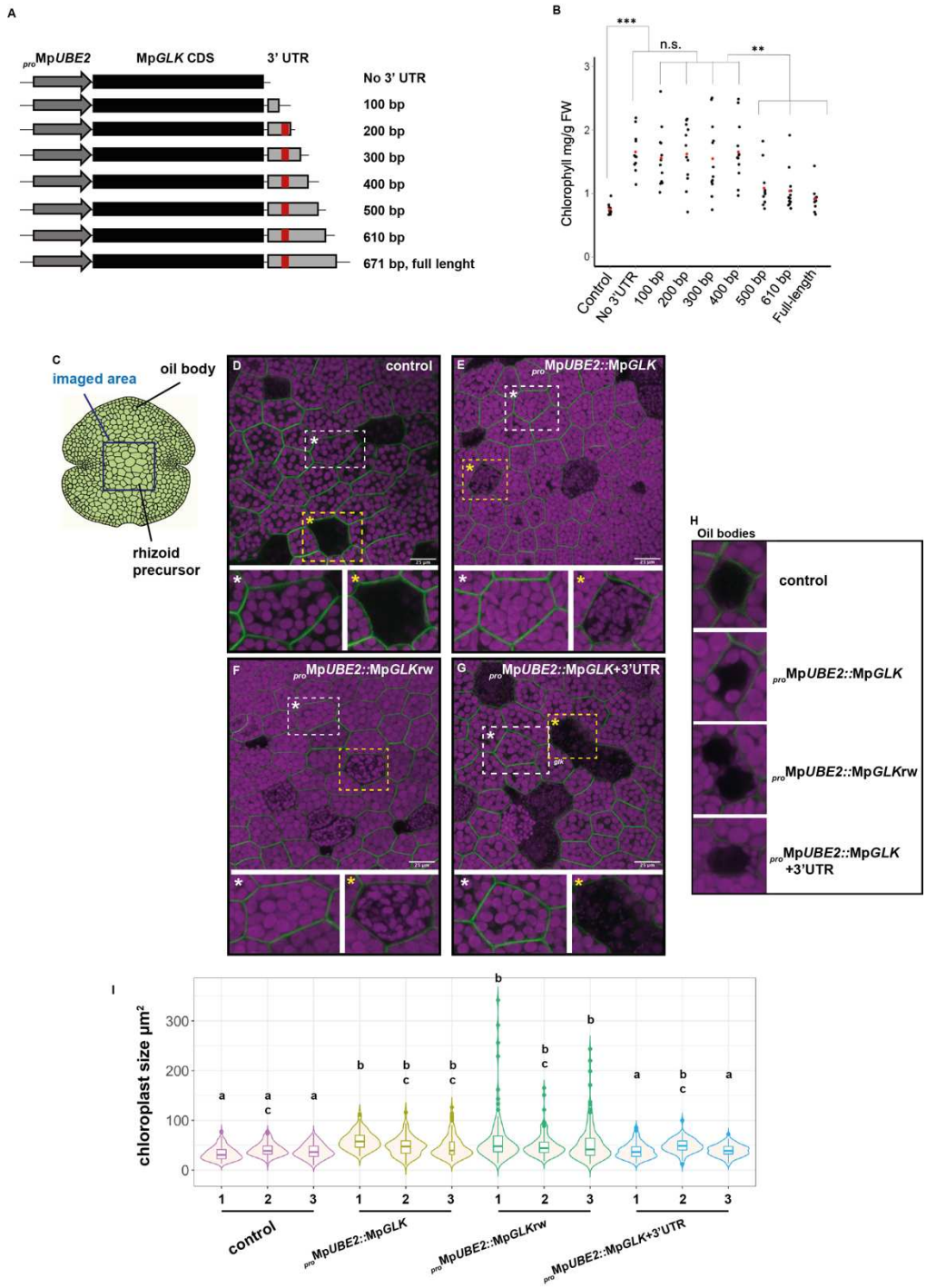

Figure 5: Mp*GLK* 3´UTR controls chlorophyll content. A) Schematic representation of overexpression constructs where Mp*GLK* CDS is fused to a full-length or a series of truncated Mp*GLK* 3´UTR versions. Grey arrows represent Mp*UBE2* promoter, black rectangles - Mp*GLK* CDS, grey rectangles – a full-length or truncated Mp*GLK* 3´UTR. Red boxes indicate Mp*GLK* 3´UTR region harbouring a predicted miR11666.4 and another putative miRNA/siRNA recognition/cleavage sites. B) Jitter plot showing chlorophyll content in transgenic lines overexpressing Mp*GLK* CDS with a full-length or truncated Mp*GLK* 3´ UTRs. Black dots represent individual transformants, red dots indicate mean average. Asterisks indicate statistically significant difference using a two-tailed *t*-test $P \leq 0.0001$ (***), $P \leq 0.001$ (**) and n.s.: non-significant. C) Schematic of a gemmae; a dark blue square indicates imaged area, black arrows indicate approximate positions of rhizoid precursor and oil body cells. D-G) Representative confocal microscopy images of control plant gemmae and Mp*GLK-CDS*, Mp*GLK-CDSrw* and Mp*GLK-CDS+3´UTR* overexpressing lines. Chlorophyll autofluorescence is shown in magenta and plasma membrane marked with eGFP shown in green. Magnified cells (white asterisks) and rhizoids precursors (yellow asterisks) are shown under each panel. Scale bars, represent 25 µm. H) Representative confocal microscopy images of oil body cells in a control plant and Mp*GLK-CDS*, Mp*GLK-CDSrw* and Mp*GLK-CDS+3´UTR* overexpressing lines. I) Boxplot showing chloroplast size range in control versus Mp*GLK-CDS*, Mp*GLK-CDSrw* and Mp*GLK-CDS+3´UTR* overexpressing lines. 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.

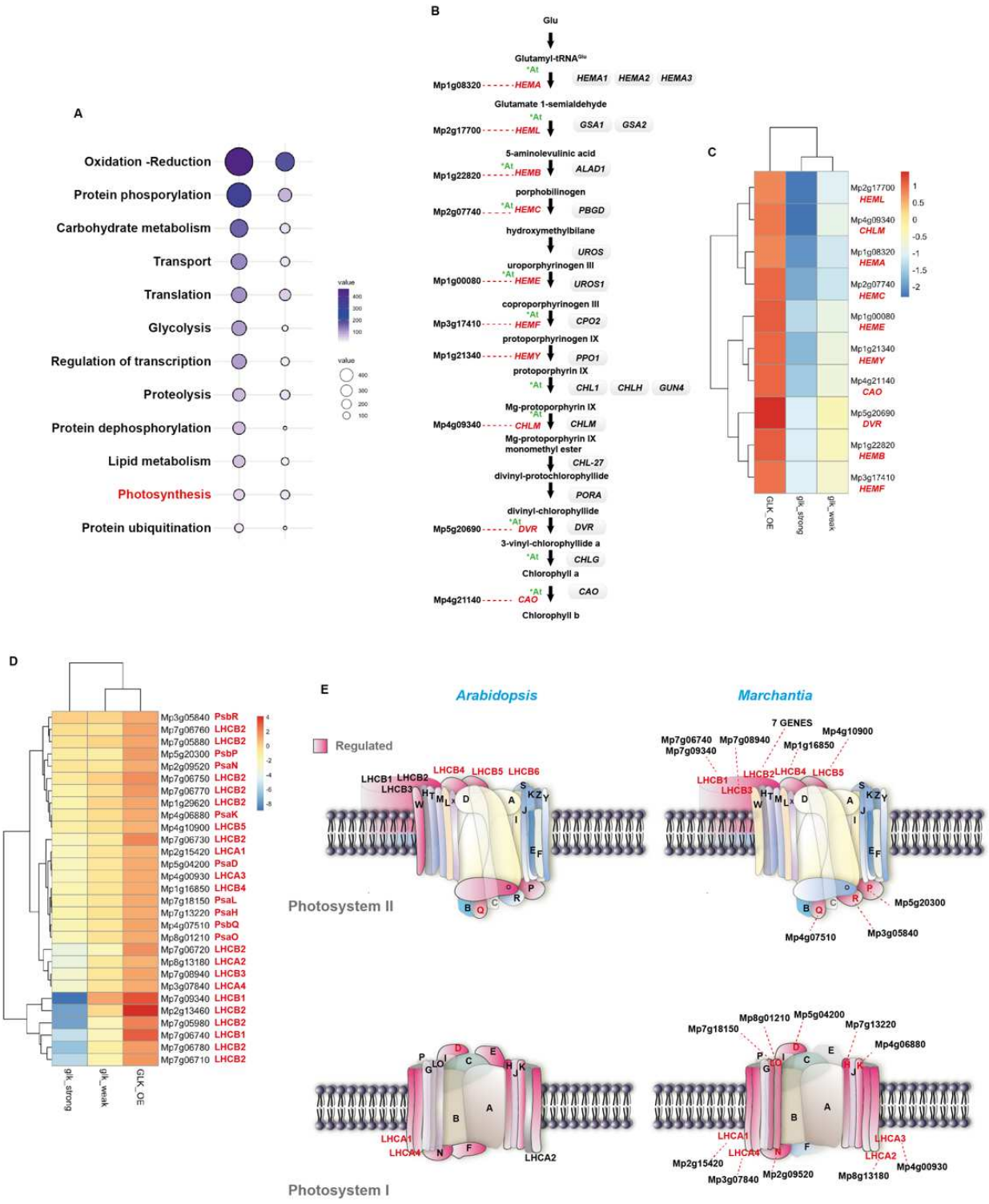

Figure 6: Mp*GLK* controls gene expression. A) Gene ontology (GO) term enrichment of Mp*GLK* overexpression plants versus 'strong' allele mutants. Differentially expressed genes with p-value=<0.01 were identified and were grouped according to their GO terms corresponding to biological processes. Terms with at least twenty genes are shown and the size of the dots represents the number of genes corresponding to a given term. B) Summary of the chlorophyll biosynthetic pathway. With "A*" genes that are regulated by *GLK1* and *GLK2* genes in *A. thaliana*. C) Heatmap of differentially expressed chlorophyll biosynthetic pathway genes (Log2FoldChange). D) Heatmap of differentially expressed of Photosystems I and II genes (Log2FoldChange). E) Schematic representation of Photosystem I and II. Subunits showing an increase (>1) in corresponding transcript levels are highlighted in pink. Genes that are differentially expressed are highlighted in red. Figure modified from Water et al. 2009 and Tu et al. 2022.