

# GOLDEN2-like1 is sufficient but not necessary for chloroplast biogenesis in mesophyll cells of C<sub>4</sub> grasses

Julia Lambret-Frotte<sup>†</sup>, Georgia Smith and Jane A. Langdale<sup>\*</sup> 

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

Received 4 August 2023; revised 25 September 2023; accepted 29 September 2023.

<sup>\*</sup>For correspondence (e-mail [jane.langdale@biology.ox.ac.uk](mailto:jane.langdale@biology.ox.ac.uk)).

<sup>†</sup>Present address: NIAB, Park Farm, Villa Road, Impington CB24 9NZ, Cambridge, UK

## SUMMARY

Chloroplasts are the site of photosynthesis. In land plants, chloroplast biogenesis is regulated by a family of transcription factors named *GOLDEN2-like* (*GLK*). In C<sub>4</sub> grasses, it has been hypothesized that genome duplication events led to the sub-functionalization of *GLK* paralogs (*GLK1* and *GLK2*) to control chloroplast biogenesis in two distinct cell types: mesophyll and bundle sheath cells. Although previous characterization of *golden2* (*g2*) mutants in maize has demonstrated a role for *GLK2* paralogs in regulating chloroplast biogenesis in bundle sheath cells, the function of *GLK1* has remained elusive. Here we show that, contrary to expectations, *GLK1* is not required for chloroplast biogenesis in mesophyll cells of maize. Comparisons between maize and *Setaria viridis*, which represent two independent C<sub>4</sub> origins within the Poales, further show that the role of *GLK* paralogs in controlling chloroplast biogenesis in mesophyll and bundle sheath cells differs between species. Despite these differences, complementation analysis revealed that *GLK1* and *GLK2* genes from maize are both sufficient to restore functional chloroplast development in mesophyll and bundle sheath cells of *S. viridis* mutants. Collectively our results suggest an evolutionary trajectory in C<sub>4</sub> grasses whereby both orthologs retained the ability to induce chloroplast biogenesis but *GLK2* adopted a more prominent developmental role, particularly in relation to chloroplast activation in bundle sheath cells.

**Keywords:** C<sub>4</sub>, chloroplasts, development, maize, green foxtail.

## INTRODUCTION

Chloroplasts are endosymbiotic organelles and are the site of photosynthetic reactions in a range of organisms from algae to land plants (Sagan, 1967). Given the importance of chloroplast function for autotrophic growth, development of the organelle is a tightly controlled process that involves many levels of regulation (Waters & Langdale, 2009). First described in the mid-1990s, the *golden2* (*g2*) mutant in *Zea mays* (maize) exhibits pale green leaves and impaired chloroplast development specifically in bundle sheath cells (Hall et al., 1998; Langdale & Kidner, 1994). Subsequent characterization revealed that *ZmG2* (sometimes referred to as *ZmGLK2*) encodes a transcription factor (Rossini et al., 2001), a founding member of the GARP transcription factor family (Riechmann et al., 2000). In most plant species examined, two *G2*-like (*GLK*) genes are present in the genome (Wang et al., 2013). *GLK* genes are defined by a characteristic helix-turn-helix DNA-binding domain related to MYB or MYB-like transcriptional factors, a conserved AREAEAA motif located immediately after the

last helix, and a GCT-box in the C-terminal region that is responsible for either homo- or heterodimerization (Fitter et al., 2002; Rossini et al., 2001; Safi et al., 2017). Functional characterization of *GLK* genes in maize and *Arabidopsis thaliana* (*Arabidopsis*) has shown that they are involved in the transcriptional activation of multiple photosynthesis-related genes including those encoding proteins of the light-harvesting complexes and enzymes of the chlorophyll biosynthesis pathway (Cribb et al., 2001; Waters et al., 2009). This pivotal role in chloroplast biogenesis has been associated with a myriad of physiological processes including stomatal movement, hormone signaling, defense responses, and fruit ripening (Kobayashi et al., 2012; Li et al., 2020; Nguyen et al., 2014). The emerging picture is one where *GLK* proteins act as linchpins in any metabolic or developmental pathway that requires chloroplast function to be modulated.

During the evolution of land plants, the *GLK* gene family underwent multiple independent duplication events (Wang et al., 2013). In cases such as *Arabidopsis*

where duplications likely occurred within species, the resulting paralogs are functionally redundant (Waters et al., 2009). However, duplication events in the monocots are more intriguing. The common ancestor of the Poales underwent a whole genome duplication that subdivided the *GLK* gene family into two sister clades: *GLK1* and *GLK2* (Wang et al., 2013). Notably, this event occurred before the many independent speciation events that led to the evolution of  $C_4$  photosynthesis within the Poales group (Christin & Osborne, 2013). In monocot species such as rice that perform  $C_3$  photosynthesis, the *GLK* paralogs are functionally redundant as in *Arabidopsis* (Wang et al., 2013; Waters et al., 2009). Unlike  $C_3$  photosynthesis, where carbon reactions occur predominantly in mesophyll cell chloroplasts, in  $C_4$  plants carbon assimilation and fixation reactions are compartmentalized in the chloroplasts of mesophyll and bundle sheath cells, respectively (Langdale, 2011). This biochemical compartmentalization is often accompanied by anatomical differences in chloroplast ultrastructure (Sage et al., 2011; Woo et al., 1970). In the  $C_4$  monocot species maize, *Setaria viridis* (setaria) and *Sorghum bicolor* (sorghum), *GLK1* is expressed preferentially in mesophyll cells whereas *GLK2* transcripts accumulate preferentially in bundle sheath cells, with the level of specificity differing between species (Chang et al., 2012; John et al., 2014; Li et al., 2010; Tausta et al., 2014; Wang et al., 2013). Notwithstanding that protein levels dictate functional capacity, defects in bundle sheath but not mesophyll chloroplast development in *g2* mutants of maize (Cribb et al., 2001; Langdale & Kidner, 1994; Rossini et al., 2001) demonstrated that the preferential localization of *ZmG2* transcripts in bundle sheath cells is functionally relevant and that *GLK* paralogs in  $C_4$  grasses can control chloroplast development in a cell-specific manner.

*GLK* paralogs have been extensively characterized in *Arabidopsis*, however, little is known about gene function in monocots, and in particular the function of *GLK1* genes has not been investigated. Because *GLK1* transcripts accumulate at much higher levels in mesophyll cells than bundle sheath cells of both maize (10-fold) and setaria (30-fold) (Chang et al., 2012; John et al., 2014; Figure S1), and *Zmg2* mutants of maize show bundle sheath specific chloroplast defects, it has been hypothesized that *GLK1* controls chloroplast biogenesis specifically in mesophyll cells of  $C_4$  grasses (Rossini et al., 2001). Here we test this hypothesis by generating and characterizing loss-of-function mutants in maize and setaria, two species that represent independent origins of  $C_4$ . Our findings reveal that contrary to expectations, neither *ZmGLK1* nor *SvGLK1* is necessary for chloroplast development in mesophyll cells whereas loss of *SvGLK2* function perturbs chloroplast development in both bundle sheath and mesophyll cells.

## RESULTS

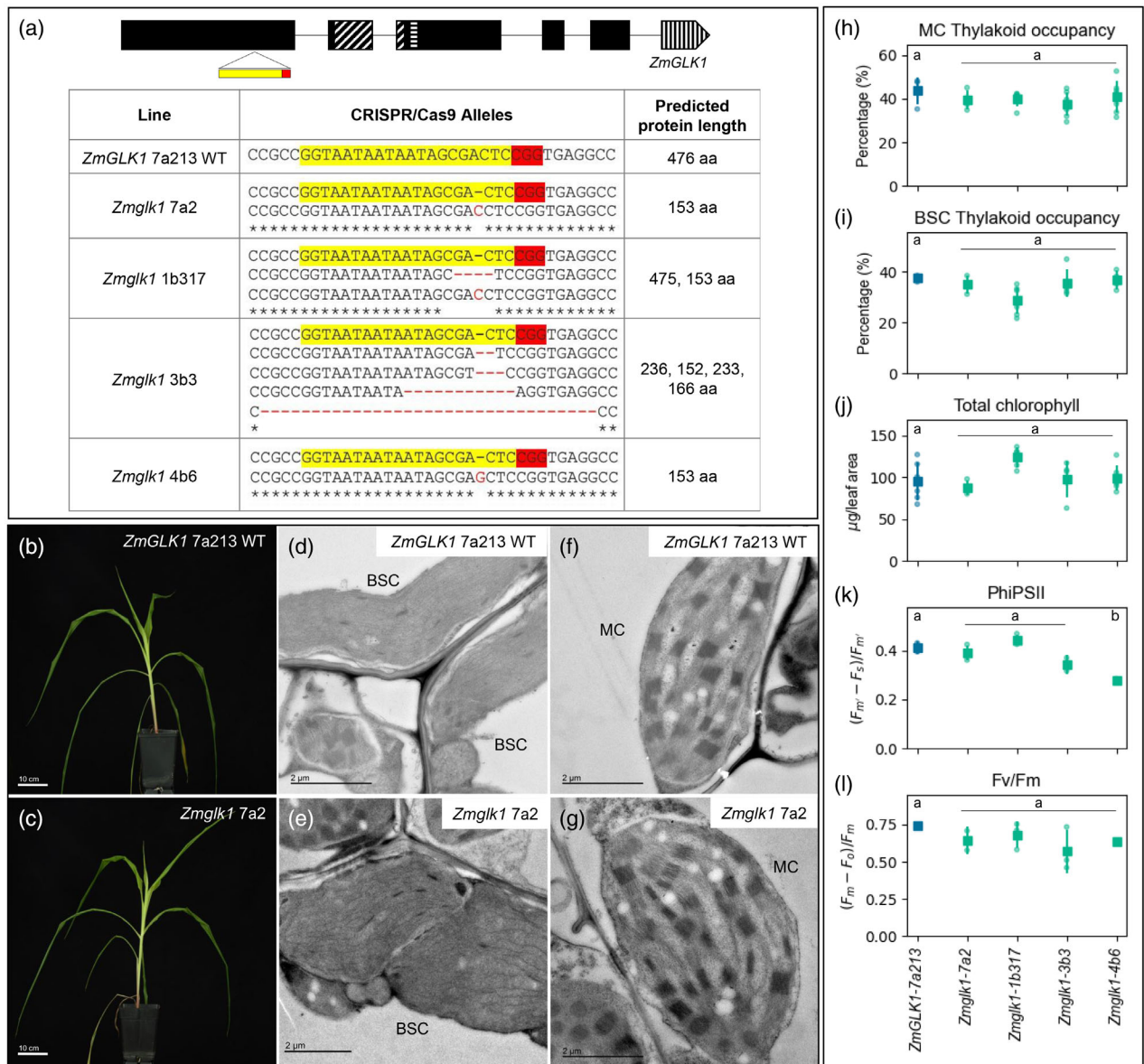
### ***ZmGLK1* is not necessary for chloroplast biogenesis in mesophyll cells of maize**

To determine whether *ZmGLK1* regulates chloroplast biogenesis in maize, loss-of-function mutants were generated using CRISPR/Cas9 (Figure 1a). Four independent mutant lines (referred to collectively as *Zmg1k1*) were obtained, all of which exhibited the same overall growth phenotype as segregating wild-type siblings (Figure 1b,c; Figure S2). Ultrastructure analysis revealed that chloroplasts in *Zmg1k1* lines were similar to wild-type, with well-developed granal thylakoids in mesophyll cells and agranal thylakoids in bundle sheath cells (Figure 1d–i; Figure S2). Total chlorophyll levels were also similar in wild-type and *Zmg1k1* mutants (Figure 1j). As such, we concluded that *ZmGLK1* is not necessary for thylakoid assembly or chlorophyll biosynthesis in bundle sheath or mesophyll chloroplasts of maize.

An evaluation of chloroplast function in *Zmg1k1* mutants was made by measuring photosystem II (PSII) parameters. PSII is a large protein complex comprised of the antenna and chlorophyll molecules that harvest light and together initiate the electron transfer reactions of photosynthesis (Nickelsen & Rengstl, 2013). In maize, PSII accumulates in the granal thylakoids of mesophyll cells and is largely absent from the agranal bundle sheath cell chloroplasts (Woo et al., 1970). As such, PSII function is a proxy for mesophyll cell chloroplast function. Fluorescence measurements were therefore taken in light and dark-adapted plants to quantify PSII operating efficiency (PhiPSII) and maximum quantum efficiency of PSII photochemistry (Fv/Fm), in both wild-type and *Zmg1k1* mutant plants. Figures 1k,l show that PSII is operating normally in *Zmg1k1* mutants, suggesting that *ZmGLK1* is not necessary for chloroplast biogenesis or function in maize mesophyll cells.

### ***SvGLK1* is not necessary for chloroplast biogenesis in mesophyll cells of setaria**

Maize and setaria have different origins of  $C_4$  photosynthesis within the PACMAD clade (Li & Brutnell, 2011) but like the orthologous *ZmGLK1*, *SvGLK1* transcripts accumulate preferentially in mesophyll cells (John et al., 2014). Phylogenetic analysis and amino acid sequence comparisons revealed that *SvGLK1* shares many structural features with *ZmGLK1* (Figure S3). To assess whether *SvGLK1* plays a role in chloroplast biogenesis, two independent loss-of-function mutants were generated in setaria using CRISPR/Cas9 (Figure 2a). As in maize, *Svglk1* mutants exhibited overall plant growth (Figure 2b–d) and chloroplast ultrastructure (Figure 2e–j) phenotypes that were similar to segregating wild-type siblings, with thylakoid membranes developing normally in both mesophyll and bundle sheath cells (Figure 2e–l). Furthermore, there was no significant difference in chlorophyll levels between *Svglk1* mutants



**Figure 1.** Chloroplast development is not perturbed in *Zmglk1* mutants of maize.

(a) Gene model of *ZmGLK1* showing predicted exons in black boxes and the target site of CRISPR/Cas9 mediated mutagenesis in yellow and red below. The HLH associated with MYB or MYB-related proteins is depicted by diagonal lines; the AREA/EAA conserved site is highlighted with horizontal lines and the last exon that contains the GCT-box domain is depicted by vertical lines. The table shows the sequence of the wild-type (WT) and mutated alleles with the predicted protein length. Each line represents a single transformation event. In the case of line 3b3 the T-DNA could not be segregated away and thus mutations continued to be generated in each generation. The phenotyped plants all lacked a wild-type allele but contained different combinations of the four mutant alleles indicated. The gRNA sequence is highlighted in yellow and the PAM sequence in red.

(b,c) Whole plant phenotype of wild-type (b) and a representative *Zmglk1* mutant 30 days after sowing (c). Scale bars = 10 cm.

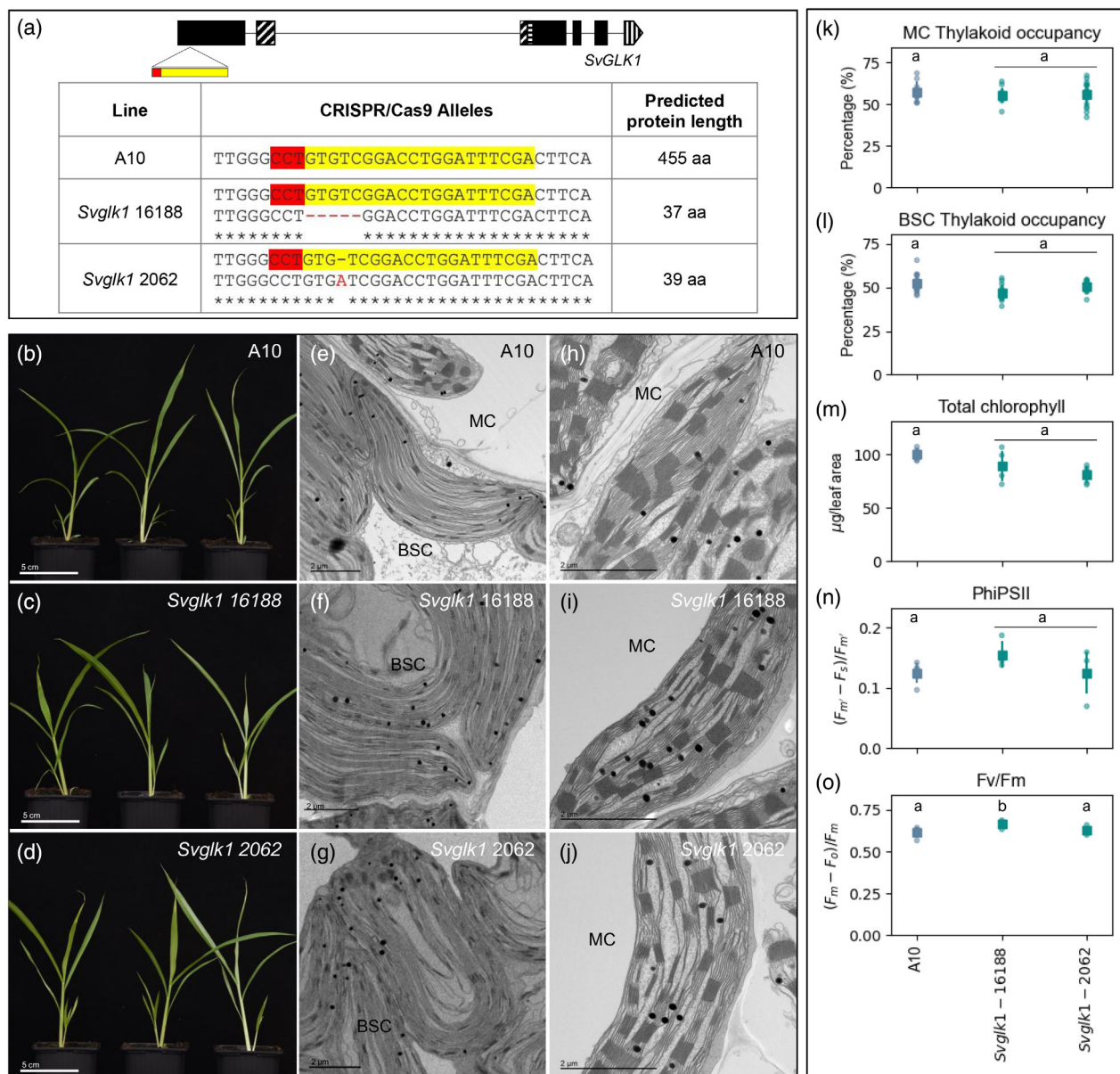
(d-g) Electron micrographs of chloroplast ultrastructure in bundle sheath (BSC; d,e) and mesophyll (MC; f,g) cells of WT (d,f) and a representative *Zmglk1* mutant (e,g). Scale bars = 2  $\mu$ m.

(h-l) Measurements of thylakoid occupancy in mesophyll cells (h), thylakoid occupancy in bundle sheath cells (i), total chlorophyll (j), PSII operating efficiency (k), and maximum quantum efficiency of PSII photochemistry (l). The WT segregant is depicted in blue and the four independent *Zmglk1* lines are depicted in green. Each dot represents a biological replicate, the square is the average value for that line and the bars are the associated standard deviations. A minimum of five biological replicates were evaluated for thylakoid occupancy and chlorophyll quantification, and three biological replicates were assessed for PSII fluorescence. Different letters indicate statistically different groups ( $P$  value  $\leq 0.05$ , one-way ANOVA and Tukey's HSD; see Table S2 for raw data).

and wild-type (Figure 2m), and electron flow through PSII was unperturbed (Figure 2n,o). Collectively, these results suggest that the preferential expression of *GLK1* genes in

mesophyll cells of *C<sub>4</sub>* grass species does not reflect a requirement for gene function during chloroplast development and/or maintenance.





**Figure 2.** Chloroplast development is not perturbed in *Svglk1* mutants of setaria.

(a) Gene model of *SvGLK1* showing predicted exons in black boxes and the target site of CRISPR/Cas9 mediated mutagenesis in yellow and red below. The HLH associated with MYB or MYB-related proteins is depicted by diagonal lines; the AREAEAA conserved site is highlighted with horizontal lines and the last exon that contains the GCT-box domain is depicted by vertical lines. The table shows the sequence of the wild-type (WT) and mutated alleles with the predicted protein length. The gRNA sequence is highlighted in yellow and the PAM sequence in red.

(b–d) Whole plant phenotype of wild-type (b) and *Svglk1* mutants 20 days after sowing (c,d). Scale bars = 5 cm.

(e–j) Electron micrographs of chloroplast ultrastructure in bundle sheath (BSC; e–g) and mesophyll (MC; h–j) cells of WT (e,h) and *Svglk1* mutants (f,g,i,j). Scale bars = 2  $\mu$ m.

(k–o) Measurements of thylakoid occupancy in mesophyll cells (k), thylakoid occupancy in bundle sheath cells (l), total chlorophyll (m), PSII operating efficiency (n) and maximum quantum efficiency of PSII photochemistry (o). The WT segregant is depicted in blue and the two independent *Svglk1* lines are depicted in green. Each dot represents a biological replicate, the square is the average value for that line and the bars are the associated standard deviations. Twelve chloroplasts were assessed for thylakoid occupancy quantification, and five biological replicates were assessed for chlorophyll quantification and PSII fluorescence for each sample. Different letters indicate statistically different groups ( $P$  value  $\leq 0.05$ , one-way ANOVA and Tukey's HSD; see Table S2 for raw data).

### ***SvGLK2* regulates chloroplast development in both mesophyll and bundle sheath cells of setaria.**

Given that dimorphic chloroplast development is observed in bundle sheath and mesophyll cells of both maize and

setaria, we next sought to determine whether *SvGLK2*, like its ortholog *ZmG2* (Figure S3), regulates bundle sheath chloroplast development. To this end, CRISPR/Cas9 was used to generate loss-of-function *Svglk2* mutants

(Figure 3a). In contrast to maize where *g2* mutants are weaker than wild-type but are viable and fertile, any *Svglk2* mutant plants that regenerated after transformation did not progress beyond the seedling stage before dying. The only viable line recovered from tissue culture was heterozygous for *SvGLK2* and this was used to isolate a null homozygous line via self-pollination. *Svglk2* 8G5 mutants, henceforth referred to as *Svglk2*, exhibited severe growth defects, had pale yellow leaves, and failed to complete the life cycle (Figure 3b). Leaf cross sections showed misshapen and smaller chloroplasts in the bundle sheath cells of *Svglk2* mutants compared to wild-type (Figure 3c,d), and observations of chloroplast ultrastructure revealed impaired chloroplast anatomy with discontinuous thylakoids (Figure 3e,f), but thylakoid occupancy within the organelle was not significantly different from wild-type (Figure 3i). More striking, given the bundle sheath-specific perturbations observed in maize *g2* mutants, was the aberrant phenotype of mesophyll chloroplasts in *Svglk2* mutants. Compared to the wild-type there was a significant reduction in thylakoid membrane coverage (15% less than the wild-type) and a noticeable reduction in the size and frequency of grana (Figure 3g,h, j). As in bundle sheath cells, thylakoid membranes in mutant mesophyll chloroplasts appeared fragmented (Figure 3h). The observed perturbations in bundle sheath and mesophyll cells were accompanied by a reduction in leaf chlorophyll content of more than 80% (Figure 3k). Together, the *Svglk2* and *Svglk1* mutant phenotypes suggest that *SvGLK1* cannot compensate for loss of *SvGLK2* function whereas *SvGLK2* can compensate for loss of *SvGLK1* function. As such, despite *SvGLK2* transcripts comprising only 43% of the total *GLK* transcript pool in mesophyll cells (as compared to 98% in bundle sheath cells) (John et al., 2014), *SvGLK2* is the primary regulator of chloroplast biogenesis in both cell-types.

#### ***GLK2* transcript levels are not upregulated in *glk1* mutants of maize or setaria**

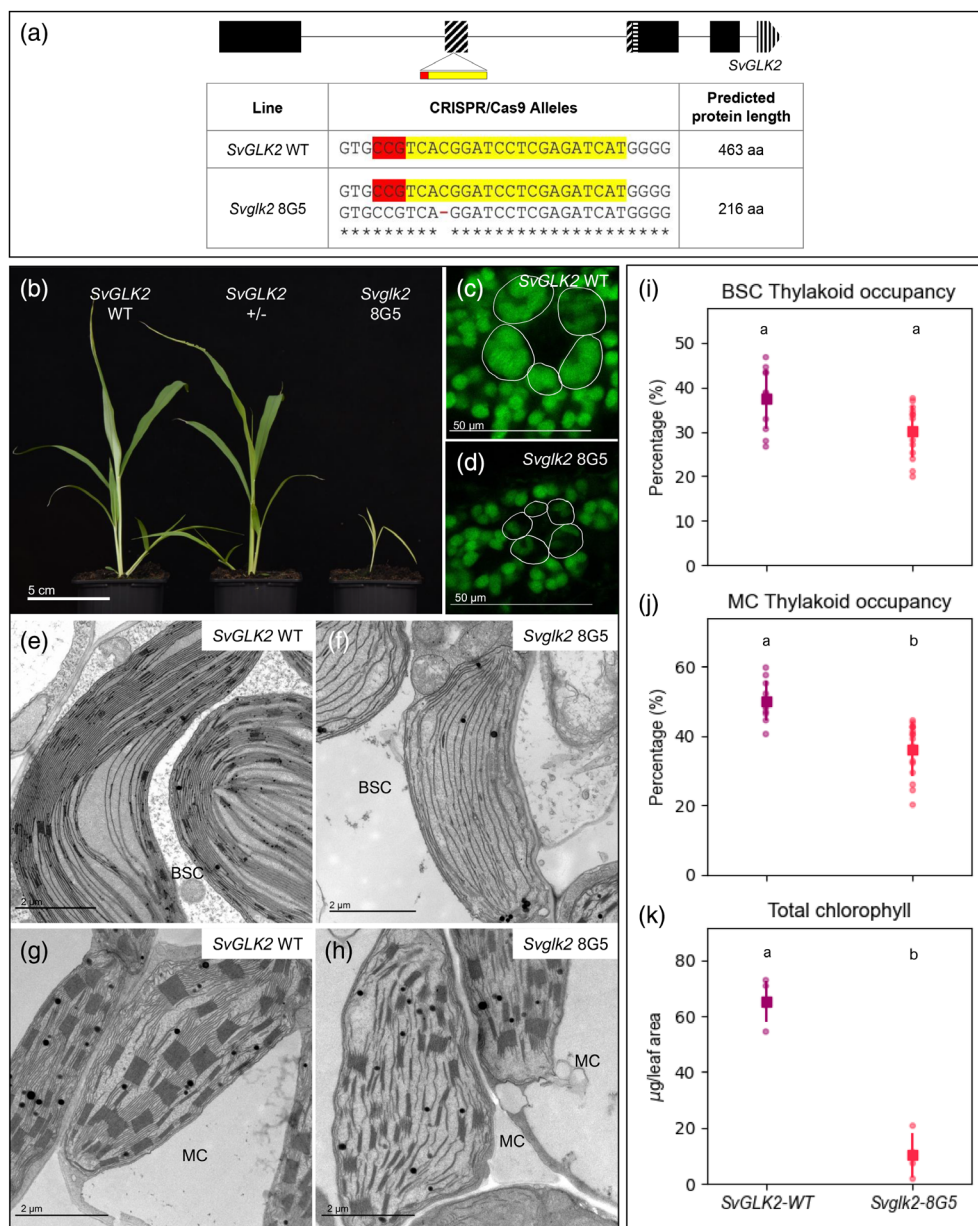
Given that *ZmGLK1* and *SvGLK1* transcripts represent 95% and 57% of the total *GLK* transcript pool in mesophyll cells of the respective species (Chang et al., 2012; John et al., 2014; Figure S1), the absence of chloroplast defects in this cell-type in loss of function mutants is intriguing. It is particularly notable because all *GLK* genes characterized to date have a role in chloroplast biogenesis, albeit a redundant role alongside paralogs in many cases (Fitter et al., 2002; Wang et al., 2013). Even a redundant role is unlikely in setaria given the *Svglk2* mutant phenotype. However, redundancy could explain the lack of a *Zmglk1* mutant phenotype if *ZmG2* transcripts are translated very efficiently in mesophyll cells of maize. Alternatively, loss of *ZmGLK1* function could lead to altered regulation of *ZmG2* expression and/or to the upregulation of genes encoding

the GATA transcription factors *GNC* and *CGA1* which also regulate chloroplast development (Chiang et al., 2012; Hudson et al., 2013) and have been shown to partially rescue loss of *GLK* function in Arabidopsis (Zubo et al., 2018). To test this hypothesis, transcript levels of *ZmGLK1*, *ZmG2*, *ZmGNC*, and *ZmCGA1* were quantified in maize using the *Zmglk1* gene-edited mutants generated above alongside the previously reported transposon-induced *Zmg2-bsd1-s1* mutant (Cribb et al., 2001). In each case the mutant allele produces transcripts that encode non-functional protein(s). Figure 4 reveals that loss of *Zmglk1* function does not lead to consistently elevated levels of *ZmGLK1* (Figure 4a), *ZmG2* (Figure 4b), *ZmGNC* (Figure 4c) or *ZmCGA1* (Figure 4d) transcripts. A small but significant increase in *ZmG2*, *ZmGNC* and *ZmCGA1* transcript levels was seen in one of the four *Zmglk1* mutant lines but that cannot explain how chloroplasts develop normally in all four lines. We thus conclude that transcription of *ZmG2*, *ZmGNC* and *ZmCGA1* is not upregulated in mesophyll cells when *Zmglk1* function is lost.

Whereas loss of *ZmGLK1* function in maize does not impact on *ZmGLK1* or *ZmG2* transcript levels, loss of *ZmG2* function in bundle sheath cells leads to a statistically significant reduction in the accumulation of both *ZmGLK1* and *ZmG2* transcripts (Figure 4a,b), and a similar situation is seen in setaria where both *SvGLK1* and *SvGLK2* transcript levels are dramatically reduced in *Svglk2* mutants (Figure 4e,f). This observation suggests that the *ZmG2*/*SvGLK2* orthologs, either directly or indirectly, promote the accumulation of both *GLK1* and *GLK2* transcripts in the respective species. Consistent with this suggestion, multiple G-box elements which are bound by Arabidopsis *GLK* proteins (Waters et al., 2009) are present in *ZmGLK1*, *ZmG2* and *SvGLK2* promoter sequences (Figure S4), and a *GLK* binding site (Tu et al., 2022) is present 1840 bp upstream of the ATG in the *SvGLK1* gene. Reduced levels of *SvGLK1* transcripts in one of the two *Svglk1* mutant lines (Figure 4e) suggest that *SvGLK1* might regulate its own transcript levels via this binding site but the absence of such an effect in the other mutant line makes it equally likely that the specific edits in line #2062 lead to transcript turnover, possibly via the nonsense-mediated decay pathway. Collectively these data suggest that chloroplast development in *C<sub>4</sub>* grasses is predominantly regulated by genes of the *GLK2* clade.

#### ***ZmGLK1* and *ZmG2* are both sufficient to restore dimorphic chloroplast development in mesophyll and bundle sheath cells of *Svglk2* mutants**

Our findings thus far show that *SvGLK2* and *ZmG2* play a more central role in controlling chloroplast biogenesis in *C<sub>4</sub>* grasses than their respective *GLK1* paralogs. To evaluate whether this difference in necessity also applies to sufficiency, *ZmGLK1* and *ZmG2* were expressed under the



**Figure 3.** Chloroplast development is perturbed in both mesophyll and bundle sheath cells of *Svglk2* mutants of setaria.

(a) Gene model of *SvGLK2* showing predicted exons in black boxes and the target site of CRISPR/Cas9 mediated mutagenesis in yellow and red below. The HLH associated with MYB or MYB-related proteins is depicted by diagonal lines; the AREAEAA conserved site is highlighted with horizontal lines and the last exon that contains the GCT-box domain is depicted by vertical lines. The table shows the sequence of the wild-type (WT) and mutated alleles with the predicted protein length. The gRNA sequence is highlighted in yellow and the PAM sequence in red.

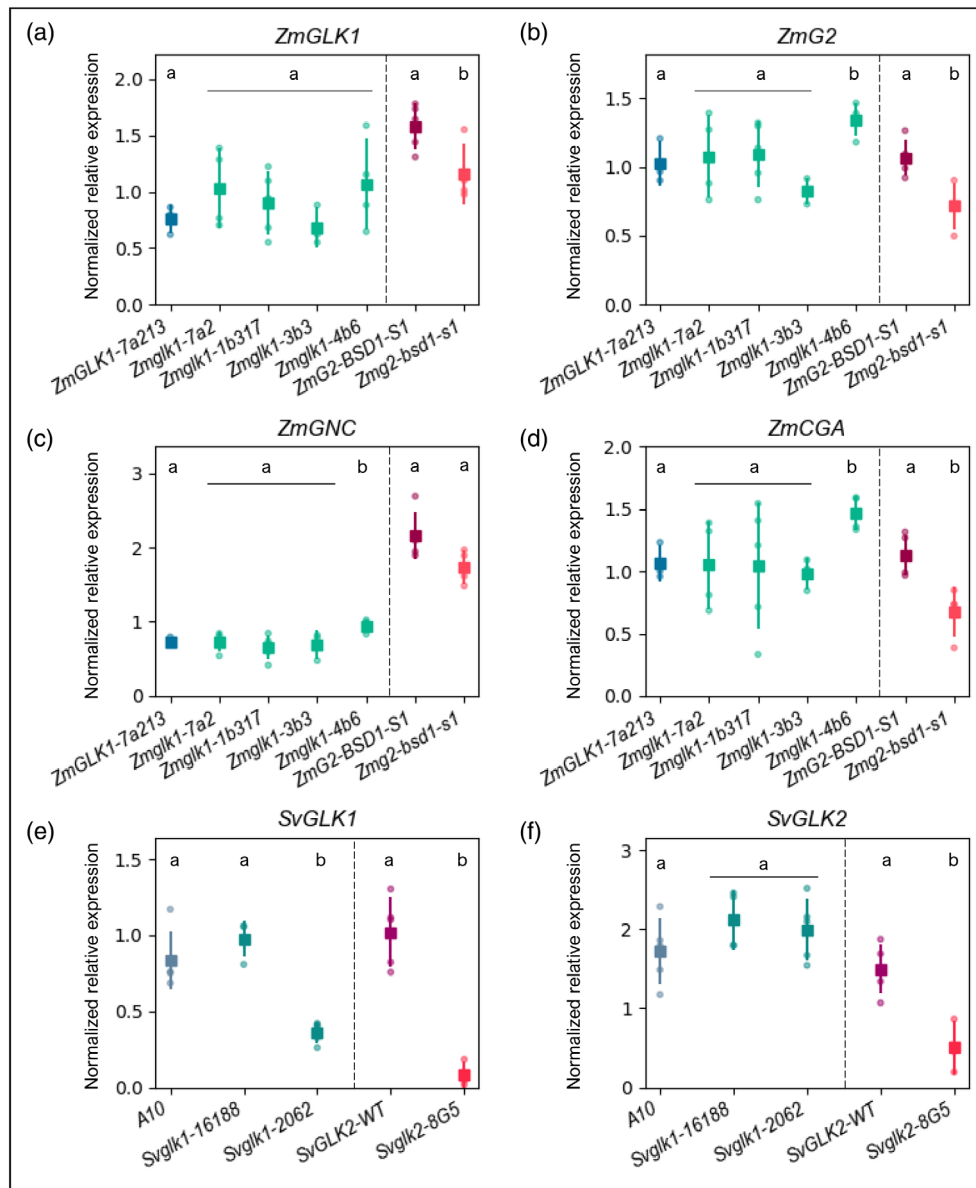
(b) Whole plant phenotype of wild-type, a heterozygous and a homozygous *Svglk2* mutant 20 days after sowing. Scale bar = 5 cm.

(c,d) Confocal images of leaf cross sections showing overall chloroplast structure in WT (c) and the *Svglk2* mutant (d). Bundle sheath cells are outlined in white. Scale bars = 50  $\mu$ m.

(e-h) Electron micrographs of chloroplast ultrastructure in bundle sheath (BSC; e,f) and mesophyll (MC; g,h) cells of WT (e-g) and the *Svglk2* mutant (f,h). Yellow arrows point to fragmented membranes. Scale bars = 2  $\mu$ m.

(i-k) Measurements of thylakoid occupancy in bundle sheath cells (i), thylakoid occupancy in mesophyll cells (j) and total chlorophyll (k). The WT segregant is depicted in purple and the *Svglk2* line is depicted in pink. Each dot represents a biological replicate, the square is the average value for that line and the bars are the associated standard deviations. Twelve chloroplasts were assessed for thylakoid occupancy quantification, and five biological replicates were assessed for chlorophyll quantification and PSII fluorescence for each sample. Different letters indicate statistically different groups ( $P$  value  $\leq 0.05$ , one-way ANOVA and Tukey's HSD; see Table S2 for raw data).





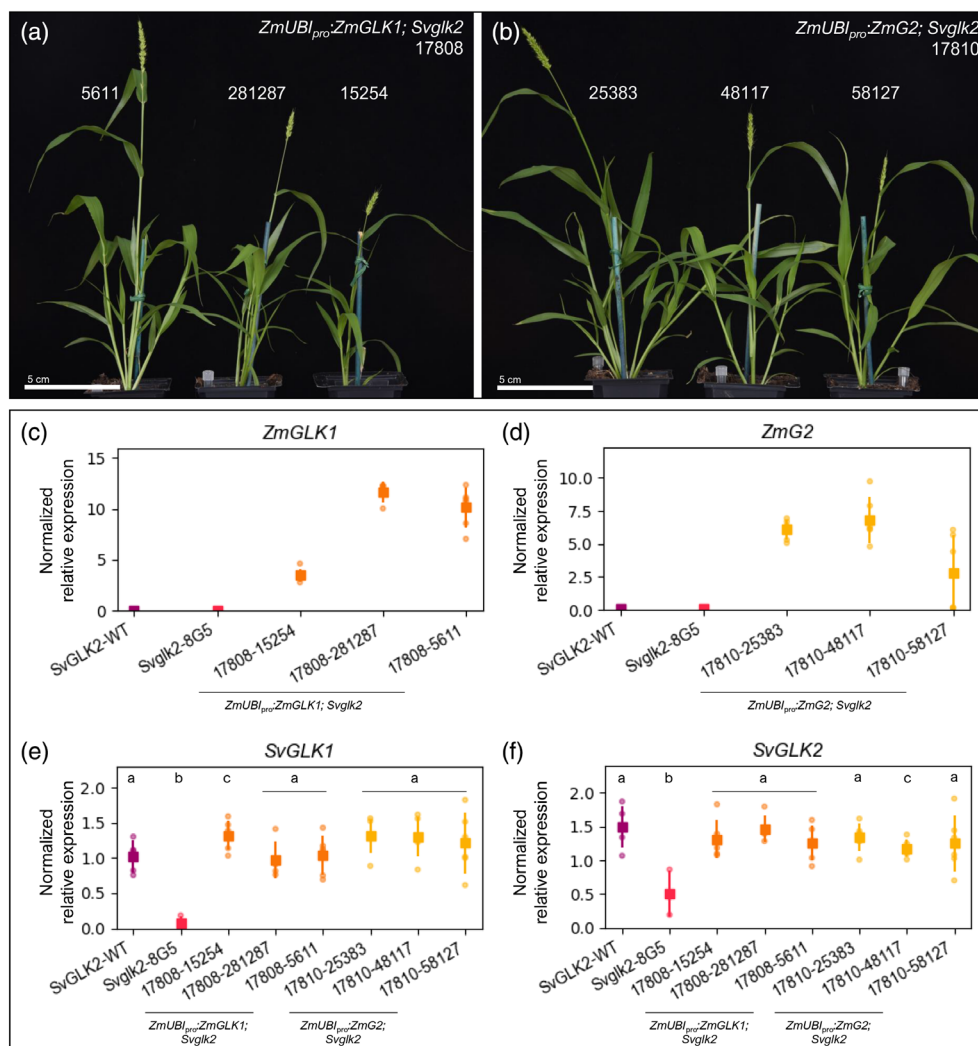
**Figure 4.** *GLK1* and *GLK2* transcript levels are reduced in *Svglk2* and *Zmg2* mutants of setaria and maize.

(a–f) Quantitative analysis of gene expression via qPCR in segregating wild-type (WT) siblings and *glk* mutants of maize (a–d) and setaria (e, f). Relative normalized expression of *ZmGLK1* (a), *ZmG2* (b), *ZmGNC* (c), *ZmCGA1* (d), *SvGLK1* (e) and *SvGLK2* (f) was calculated based on the expression of two reference genes for each sample. Each dot represents a biological replicate, the square is the mean value and the bars are the associated standard deviations. A minimum of three and five biological replicates were evaluated for maize and setaria samples, respectively. Statistical significance was calculated using a Student's *t*-test and pairwise comparison with the respective WT sample. For each species comparison, different letters indicate statistically different groups ( $P$  value  $\leq 0.05$ ; see Table S2 for raw data).

control of the constitutive maize ubiquitin gene promoter (*ZmUBI<sub>pro</sub>*) in the chloroplast-defective *Svglk2* mutant background. Three independent lines were obtained for each construct and quantitative analysis of transgene transcript levels revealed differences between lines that correlated directly with plant phenotype (Figure 5a–d). For example, the weaker plants in line #17808–15 254 accumulated around half the level of *ZmGLK1* transcripts seen in healthier plants of the other two lines (Figure 5a, c).

Notably, plants of all six lines were viable, healthy, and fertile, suggesting complementation of the *Svglk2* mutant phenotype (compare *Svglk2* plants in Figure 3b with Figure 5a, b).

Given that loss of *Svglk2* function leads to reduced levels of both *SvGLK1* and *SvGLK2* transcripts (Figure 4e, f), we next sought to evaluate whether *ZmGLK1* or *ZmG2* could influence endogenous *SvGLK1* or *SvGLK2* expression. Figure 5e, f show that *SvGLK1* and *SvGLK2* transcripts



**Figure 5.** *ZmGLK1* and *ZmG2* complement the *Svglk2* phenotype and activate transcription of the setaria homologs.

(a,b) Whole plant phenotype of *Svglk2* mutant overexpressing *ZmGLK1* (a) or *ZmG2* (b) 30 days after sowing. Scale bars = 5 cm.

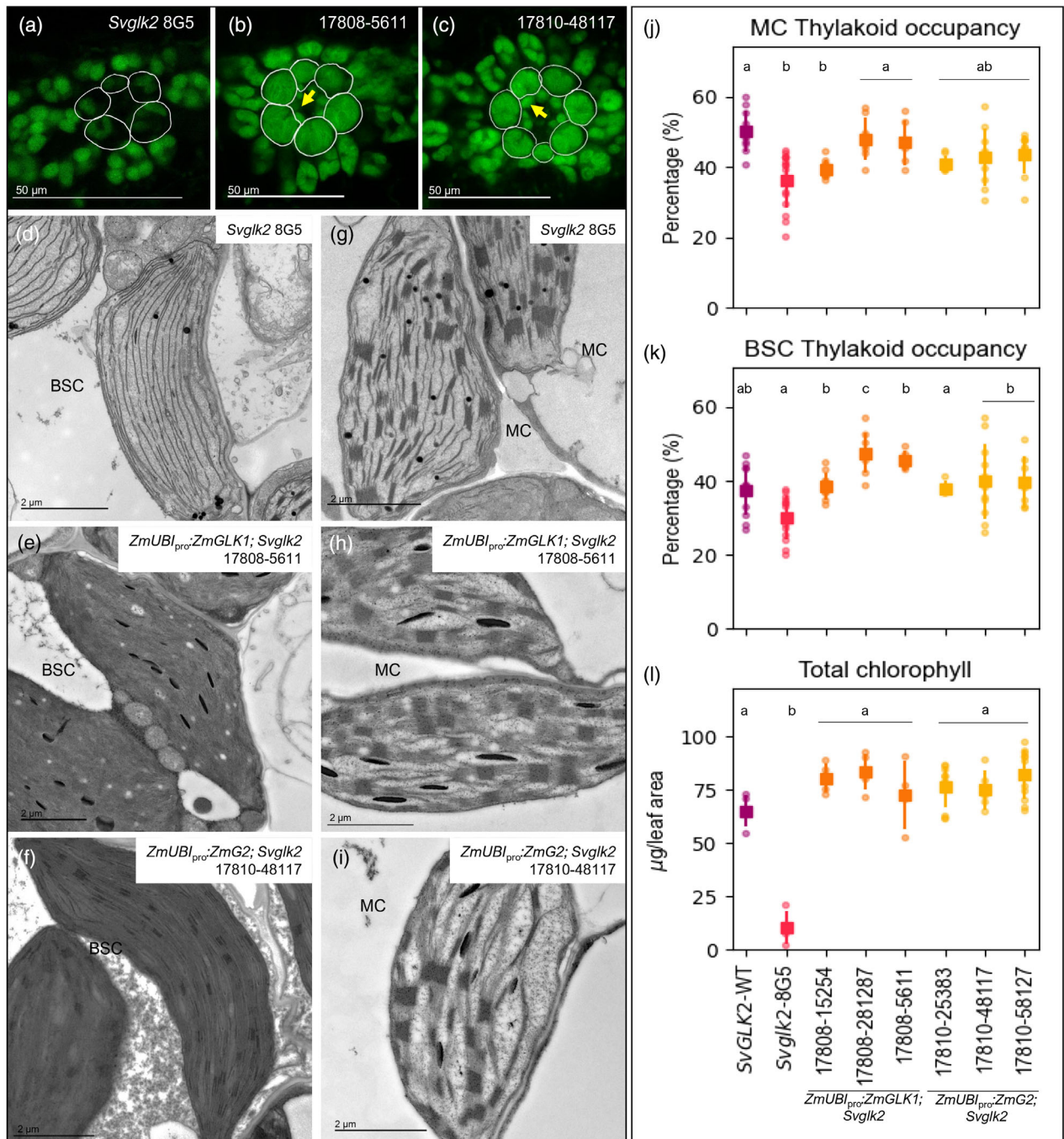
(c–f) Quantitative analysis of gene expression via qPCR of *ZmGLK1* (c), *ZmG2* (d), *SvGLK1* (e) and *SvGLK2* (f). Relative gene expression was calculated based on the expression of two reference genes for each sample. Each dot represents one of the five biological replicates evaluated, the square is the mean value and the bars are the associated standard deviations. Statistical significance was calculated using a Student's *t*-test and pairwise comparison with the respective wild-type (WT) sample. Different letters indicate statistically different groups ( $P$  value  $\leq 0.05$ ; see Table S2 for raw data).

were restored to wild-type levels in *Svglk2* lines expressing either *ZmGLK1* or *ZmG2* transgenes. In fact, regardless of the expression level of either *ZmGLK1* or *ZmG2*, all complemented lines exhibit the same wild-type-like level of *SvGLK1* and *SvGLK2* expression (Figure 5c–f). Both *ZmGLK1* and *ZmG2* can thus activate the expression of *SvGLK1* and *SvGLK2*, providing further support for the existence of a regulatory network between *GLK* genes and for the likely relevance of G-box and *GLK* binding motifs in the promoter regions (Figure S4). The functional significance of this observation must be considered in the context of transgene expression driven by a constitutive promoter and gene-edited *Svglk2* transcripts encoding non-functional proteins (and hence any effect on *SvGLK1*

expression is a direct effect of the maize transgene and not via *SvGLK2*). In this regard, we must deduce that the transgenic lines expressing *ZmGLK1* accumulate transcripts encoding functional *ZmGLK1* and *SvGLK1* proteins in both mesophyll and bundle sheath cells, and those expressing *ZmG2* accumulate transcripts encoding functional *ZmG2* and *SvGLK1* proteins, similarly in both cell types.

To determine whether *ZmGLK1* and/or *ZmG2* can rescue chloroplast ultrastructure defects in mesophyll and/or bundle sheath cells of *Svglk2* mutants, confocal images and transmission electron micrographs of transverse leaf sections from all six transgenic lines were examined. In all cases, chloroplasts were restored to normal size (Figure 6a–c; Figure S5), with predominantly agranal





**Figure 6.** *ZmGLK1* and *ZmG2* are sufficient to induce dimorphic chloroplast development in both mesophyll and bundle sheath cells of setaria. (a–c) Confocal images of leaf cross sections showing overall chloroplast morphology in the *Svglk2* mutant (a, same as Figure 3d), and representative lines expressing *ZmGLK1* (b) or *ZmG2* (c). Bundle sheath cells are outlined in white. Yellow arrows indicate ectopic chloroplasts developing in vascular cells. Scale bars = 50  $\mu$ m. (d–i) Electron micrographs of chloroplast ultrastructure in bundle sheath (BSC; d–f) and mesophyll (MC; g–i) cells of the *Svglk2* mutant (d–g; same as Figure 3g,h) and one representative line overexpressing *ZmGLK1* (e–h) or *ZmG2* (f–i). Scale bars = 2  $\mu$ m. (j–l) Measurements of thylakoid occupancy in mesophyll cells (j), thylakoid occupancy in bundle sheath cells (k) and total chlorophyll (l). The wild-type (WT) segregant is depicted in purple, the *Svglk2* line in pink, the three independent lines overexpressing *ZmGLK1* in orange and those overexpressing *ZmG2* in yellow. Each dot represents a biological replicate, the square is the average value for that line and the bars are the associated standard deviations. A minimum of 10 chloroplasts were evaluated for thylakoid occupancy quantification and five to ten biological replicates were used for chlorophyll quantification. Different letters indicate statistically different groups in ( $P$  value  $\leq 0.05$ , one-way ANOVA and Tukey's HSD; see Table S2 for raw data).

thylakoids in bundle sheath cell chloroplasts (Figure 6d–f; Figure S5) and granal thylakoids in mesophyll chloroplasts (Figure 6g–i; Figure S5). Quantification revealed that with the exception of the one line that had low levels of transgene expression (#17808–15 254), thylakoid occupancy levels in the mesophyll (Figure 6j) and bundle sheath (Figure 6k) chloroplasts of transgenic lines were statistically equivalent (or in one case higher) than in wild-type plants. Consistent with the ectopic expression of *GLK* genes, mature chloroplasts were also observed in the leaf vasculature of all complemented lines, more with *ZmG2* than *ZmGLK1* (Figure 6a–c; Figure S5). Irrespective of which maize ortholog was overexpressed, the total chlorophyll content in leaves of transgenic plants was restored to wild-type levels (Figure 6l). Notwithstanding the fact that SvGLK1 may be functional in bundle sheath and mesophyll cells of all transgenic plants, collectively these data demonstrate that ZmGLK1 and ZmG2 proteins are functionally equivalent in the context of the  $C_4$  setaria leaf and that the dimorphic identity of bundle sheath and mesophyll cell chloroplasts develops regardless of which maize ortholog is overexpressed.

## DISCUSSION

Prior to this study, it was known that *GLK* genes are duplicated in many land plant species and that in species that carry out  $C_3$  photosynthesis the duplicate genes act redundantly to regulate chloroplast *biogenesis* in all photosynthetic cells (Bravo-Garcia et al., 2009; Wang et al., 2013; Waters et al., 2008). It had also been hypothesized that *GLK1* and *GLK2* genes regulate dimorphic chloroplast development in the two photosynthetic cell types of  $C_4$  species, with each gene functioning in one of the two cell types (Rossini et al., 2001). This hypothesis was based on the observation that *GLK1* genes are preferentially expressed in mesophyll cells of maize, sorghum and setaria whereas *GLK2/G2* genes are preferentially expressed in bundle sheath cells (Chang et al., 2012; John et al., 2014; Li et al., 2010; Wang et al., 2014), and loss of *Zmg2* function in maize perturbs bundle sheath but not mesophyll chloroplast development (Cribb et al., 2001). Through the phenotypic characterization of *glk1* loss-of-function mutants in maize and setaria (Figures 1, 2 and 4), *glk2* loss-of-function mutants in setaria (Figures 3 and 4), and *glk2* setaria mutants complemented with maize *GLK1* and *GLK2* genes (Figures 5 and 6), we have revealed a more complex situation both with respect to the developmental processes operating within individual  $C_4$  species and to likely evolutionary trajectories. Specifically, we reveal that spatially restricted *GLK* gene expression patterns do not predict sites of gene function.

### GLK proteins have equivalent functional capacity

*GLK* genes duplicated in a common ancestor of the Poales, before the speciation of maize, sorghum, setaria, and rice,

and before the evolution of  $C_4$  photosynthesis (Wang et al., 2013). Given that *GLK* genes regulate chloroplast biogenesis in the moss *Physcomitrium patens* (Bravo-Garcia et al., 2009; Yasumura et al., 2005), which last shared a common ancestor with flowering plants over 450 million years ago (Kenrick & Crane, 1997; Wickett et al., 2014), it is reasonable to assume that prior to this duplication the ancestral *GLK* gene played a role in chloroplast development. The duplication event, which must have occurred between 85 and 95 million years ago (Gallaher et al., 2022), provided an opportunity for neo- and/or sub-functionalization of gene activity and thus for divergent roles to evolve in different grass species. It has previously been suggested that the compartmentalized expression of *GLK1* and *GLK2* genes in mesophyll and bundle sheath cells of the  $C_4$  species reflects distinct roles in the biogenesis of granal versus agranal chloroplasts. This hypothesis was supported by the observation that *GLK* homologs are not differentially expressed in *Gynandropsis gynandra* (formerly known as *Cleome gynandra*), a  $C_4$  dicot that exhibits granal thylakoids in chloroplasts of both mesophyll and bundle sheath cells (Marshall et al., 2007; Wang et al., 2013). However, data presented here show that appropriate cell-type specific thylakoid stacking is restored in mesophyll and bundle sheath cells of *Svglk2* mutants when either *ZmGLK1* or *ZmG2* is constitutively expressed (Figures 5 and 6). These results demonstrate that the two maize proteins have equivalent functional capacity. Furthermore, the data are consistent with a recent report which showed that the identity of genes activated by *GLK* transcription factors is determined by *cis* regulatory factors, such that a heterologous *GLK* protein will activate downstream targets that are normally activated by endogenous *GLK* proteins in the host species as opposed to targets that are activated by the heterologous protein in its species of origin (Tu et al., 2022). Together these observations suggest that the evolution of dimorphic chloroplasts in  $C_4$  species was unlikely to have been mediated by sub-functionalization of *GLK* protein activity.

### Despite equivalent functional capacity, *GLK2* plays a dominant role in both maize and setaria

The difference in the proportion of *GLK2:GLK1* transcripts in mesophyll cells of maize (5:95) versus setaria (43:57) can partially explain the different phenotypes observed in loss-of-function mutants. In maize, where relative levels of *ZmG2* are only 5% in mesophyll cells, *Zmg2* mutants exhibit perturbed chloroplast development only in bundle sheath cells (where *ZmG2* transcripts comprise 65% of the total *GLK* pool). By contrast, relative *GLK2* levels of 43% in mesophyll and 98% in bundle sheath cells of setaria are associated with defective chloroplast development in both cell types (Figure 3). This observation was initially counter-intuitive because it would be expected that, as in  $C_3$  plants,

the high levels of *SvGLK1* transcripts in mesophyll cells would compensate for loss of *SvGLK2* function. However, we observed that *SvGLK1* transcript accumulation was virtually abolished in *Svglk2* mutants, and levels of the non-functional *Svglk2* transcripts were also reduced (Figure 4). As such, we conclude that *SvGLK2* activates expression (or transcript stabilization) both of itself and *SvGLK1*, and that the *Svglk2* mutant phenotype reflects loss of function of both genes. Collectively, these data suggest that *SvGLK2* is the primary regulator of chloroplast development in both bundle sheath and mesophyll cells of setaria.

Whereas normal mesophyll chloroplast development in the absence of *GLK1* function can be explained in setaria because *SvGLK2* likely compensates for loss of *GLK1* function, the mutant phenotype in maize was unexpected. In every land plant species examined to date, a role for *GLK* genes in chloroplast development has been revealed, albeit often a redundant role with paralogs operating in the same cell-type. Because *ZmGLK1* is highly expressed in maize mesophyll cells and no other *GLK*-like transcripts appear as possible candidates for a redundant paralog, a previously unknown mechanism must be operating in this cell-type. Although *GLK* gene function has previously been shown to be cell-autonomous in *Arabidopsis*, it is possible that a non-cell-autonomous mechanism could operate in maize. Indeed, there is some evidence of this because *ZmGLK1* transcript levels are reduced when *G2* function is lost, even though *G2* transcript levels in wild-type mesophyll cells are only 5% of the *GLK* pool (Figure 4). In such a scenario, loss of *GLK1* function in mesophyll cells would be compensated for by direct or indirect activity of *G2* proteins in the bundle sheath cells. Any such effect would only be revealed in double *Zmglk1;g2* mutants, which would have defective chloroplasts in both cell-types. Regardless of whether *ZmG2* compensates for loss of *ZmGLK1* function, the *GLK2* paralog plays the dominant role in chloroplast biogenesis in maize, as in setaria.

#### **Cis-regulatory sequences are more variable in *GLK2* than *GLK1* genes**

Given that *GLK* gene paralogs encode functionally equivalent proteins both in *C<sub>3</sub>* species where they act redundantly and in the *C<sub>4</sub>* species maize, the question arises of why the paralogs are expressed in a cell-type preferential manner in *C<sub>4</sub>* grass species (at least in the ones that have been examined). Estimates based on the topology of BEP (*C<sub>3</sub>* species only) and PACMAD (*C<sub>3</sub>* & *C<sub>4</sub>* species) clades in the Poales suggest that maize diverged from rice ~80 million years ago (Gallagher et al., 2022) and that within the PACMAD clade, maize and sorghum share a common history of *C<sub>4</sub>* evolution whereas independent events occurred in maize and setaria around ~45 million years apart (Gallagher et al., 2022). Both *GLK* gene paralogs are predominantly expressed in mesophyll cells of *C<sub>3</sub>* rice leaves, with *GLK1*

transcripts accumulating at higher levels than *GLK2* (Hua et al., 2021), and gene function being redundant (Wang et al., 2013). Notably, both transcripts can also be detected in bundle sheath cells, at levels roughly equivalent to those of *GLK2* in mesophyll cells, suggesting that the few chloroplasts that develop in rice bundle sheath cells are also regulated by the redundant action of both genes. By contrast, an element of sub-functionalization is apparent in the *C<sub>4</sub>* species maize and setaria, with the *GLK2* paralog expressed preferentially in bundle sheath cells. However, data from a number of different *C<sub>3</sub>*, *C<sub>4</sub>*, and intermediate species (Washburn et al., 2021) (Figure S1) invoke a more complex scenario because preferential expression of *GLK2* paralogs in bundle sheath cells is evident in species where decarboxylation reactions occur in the bundle sheath chloroplasts (NADP-ME type) but not in species where decarboxylation occurs in the mitochondria (NAD-ME type) or in the cytoplasm (PEP-CK type). This suggests that preferential localization of *GLK2* in bundle sheath cells is associated with specific aspects of chloroplast function as opposed to general activation of photosynthesis.

The evolution of *C<sub>4</sub>* required activation of photosynthesis and a significant increase in chloroplast volume in bundle sheath cells and as such, selection on regulatory regions of *GLK2* genes to enable high levels of activity in that cell type may have been a critical step. Notably, there is more within-species sequence variation in *GLK2* promoters of different rice and maize varieties than in the corresponding *GLK1* promoters (Figure S6), providing more opportunity for change through drift or selection. *GLK2* promoter sequences are also less conserved between species than corresponding *GLK1* sequences (Figure S7). ATAC-seq data from leaves shows that the proximal promoter and 5'-UTR regions are the main sites of protein binding in *GLK* genes from maize, setaria, and rice (Figure S4), with noticeably more DNA binding motifs related to light-responsive elements in the 5'-UTR sequences of *GLK2* genes than in equivalent regions of *GLK1* genes. In maize, an enrichment of BPC motifs is seen in the *ZmG2* 5'-UTR, and the sequences are almost completely conserved across 1200 maize varieties. This level of sequence conservation in the 5'-UTR is in stark contrast to the rest of the promoter region which is highly variable (Figure S6). Intriguingly, very few DNA binding motifs are found in either the *SvGLK1* or *SvGLK2* promoters (Figure S4), and no variation relevant to the domestication process was found when compared to *Setaria italica* (He et al., 2023). As in *ZmG2*, however, BPC motifs are present in both the *SvGLK2* and *OsGLK2* 5'-UTR sequences but they are not found in any of the *GLK1* gene sequences. This observation suggests that the *GLK1* and *GLK2* 5'-UTR sequences diverged in the last common ancestor of maize, setaria, and rice and thus prior to the evolution of *C<sub>4</sub>*. Because BPC binding proteins act to recruit polycomb complex proteins and thus mediate chromatin



changes (Theune et al., 2019), this divergence may have enabled novel post-transcriptional regulation of *GLK2* orthologs during the evolution of *C<sub>4</sub>*. Differences in the presence and position of upstream open reading frames (uORFs) within the 5'-UTR may have further facilitated de-repression of *GLK2* translation (Zhang et al., 2020) in that putative uORFs in *ZmGLK1* and *SvGLK1* are situated less than 10 base pairs after the transcription start site (TSS) and thus if functional are predicted to be highly repressive, whereas the uORF in *ZmG2* is over 200 bp downstream of the TSS and no uORFs are predicted in *SvGLK2* (Figure S7). In rice, predicted uORFs are positioned 100 bp downstream of the TSS in both genes and may thus have no differential effect. Although requiring experimental verification, collectively these observations support the hypothesis that altered *GLK2* gene regulation was associated with the evolution of functional chloroplasts in bundle sheath cells of *C<sub>4</sub>* grasses.

## EXPERIMENTAL PROCEDURES

### Plant growth conditions

Maize and *S. viridis* seeds were germinated in hydrated vermiculite or on paper, respectively, and placed in an Aralab Fitoclima D1200 PLH incubator under 12 h light/12 h dark cycle at an intensity of 320  $\mu\text{mol photon m}^{-2} \text{sec}^{-1}$ , with temperatures ranging from 31°C during the day to 22°C at night and humidity between 50–60%. About 7–10 days after sowing (DAS), seedlings were transferred to pots containing John Innes no. 2 soil mixture and irrigated with water to keep the soil moist. Plants were kept under these conditions until sampled for experimental analysis.

For genetic crosses, maize plants were transferred to larger pots with the same soil mixture supplemented with Osmocote slow-release granules. Plants were grown in a greenhouse at the Department of Biology at the University of Oxford, UK with 16 h light/8 h dark cycle, daytime temperature 28°C and night-time temperature 20°C, with supplemental light provided when natural light levels were below 120  $\mu\text{mol photon m}^{-2} \text{sec}^{-1}$ .

### Construct design and assembly

*ZmG2* and *ZmGLK1* orthologs were identified in *S. viridis* via phylogenetic reconstruction of the GARP gene family (see Figure S2). Guide RNAs (gRNA) to drive CRISPR/Cas9 site-directed mutagenesis were designed using the CRISPOR (Concordet & Haeussler, 2018) online tool (<http://crispor.tefor.net/>) to achieve high editing efficiency for the target genes *ZmGLK1* (Zm00001d044785), *SvGLK1* (Sevir.4G123200), and *SvGLK2* (Sevir.5G003100), and to minimize off-targets. For maize, the designed gRNAs were sent to the Wisconsin Crop Innovation Centre (University of Wisconsin) to be cloned and transformed using their standard pipeline for maize transformation. For *S. viridis*, the gRNA sequences were synthesized as oligonucleotides with appropriate flanking sites for Golden Gate cloning (Table S1; JLF\_Ox43 and JLF\_Ox44 for *SvGLK1*; JLF\_Ox51 and JLF\_Ox52 for *SvGLK2*). Prior to cloning, the oligonucleotides were annealed and phosphorylated by incubating 10  $\mu\text{M}$  of each with 1x Ligase Buffer (New England Biolabs, Hitchin, UK) and 10 U of T4 Polynucleotide Kinase (New England Biolabs, Hitchin, UK) in a final reaction volume of 15  $\mu\text{l}$  at 37°C for 1.5 h. After that period, 2  $\mu\text{l}$  of 0.5 M NaCl was added, and the

reaction was placed in a water bath heated to 95°C for 2 min. The reaction was cooled to room temperature before being diluted 100-fold and cloned downstream of the *TaU6* promoter (module pJG310) using the Golden Gate one-step one-pot protocol (Engler et al., 2009) with an appropriate restriction enzyme (*Esp3I*). This module was then combined with another level 1 module containing *TaCas9* downstream of the maize ubiquitin promoter (pJG471) into the final acceptor and binary vector pTC278. Both modules and the vector were provided by Dr Dan Voytas (U. of Minnesota).

To overexpress *ZmGLK1* and *ZmG2*, the coding sequences of the maize genes were domesticated to avoid recognition sites for the type II restriction enzymes used in Golden Gate assembly: *BsaI*, *BpiI*, *Esp3I*, and *DrallI*. The final sequence was synthesized as a level 0 module with the appropriate SC flanking sites (Engler & Marillonnet, 2013). *ZmGLK1* and *ZmG2* level 0 modules (EC17054 and EC17056) were cloned downstream of the maize ubiquitin promoter (*ZmUBI<sub>pro</sub>* and EC15455) and upstream of a *nos* terminator (tNOS and EC41421) in level 1 backbone position forward 2 using the Golden Gate one-step one-pot protocol (Engler et al., 2009). The previously described *hygromycin phosphotransferase* (HygR) coding sequence cloned downstream of the rice actin promoter (*OsACT<sub>pro</sub>*) was used as the selectable marker (Vlad et al., 2019). Level 1 modules were assembled into the binary vector pAGM4723 to obtain the constructs EC17808 and EC17810 (hereafter referred to as 17 808 and 17 810, respectively) depicted in Figure S5b,c.

### Plant transformation

*S. viridis* accessions A10 and ME034V were used to obtain CRISPR/Cas9-driven mutations for *SvGLK1* and *SvGLK2*, respectively. *Agrobacterium tumefaciens* strain AGL1 harboring the construct of interest was co-cultivated with callus generated from seeds of *S. viridis*. Callus induction, transformation and seedling regeneration were performed as described by VanEck and collaborators (Van Eck & Startwood, 2015). Complementation lines were obtained by transforming constructs 17 808 and 17 810 into callus generated from a population of seeds that segregated for the gene-edited *Svglk2* allele but no longer segregated the original t-DNA which had been recombined out. The presence of t-DNA in regenerated seedlings was confirmed via duplex PCR (Table S1; JLF\_Ox12 and JLF\_Ox13 amplify the HygR gene; JLF\_Ox14 and JLF\_Ox15 amplify an endogenous gene used as a DNA quality control). The occurrence of gene editing by CRISPR/Cas9 was verified by PCR amplification of the region flanking the expected mutation site (Table S1; JLF\_Ox45 and JLF\_Ox46 for *SvGLK1*; JLF\_Ox53 and JLF\_Ox54 for *SvGLK2*) followed by endonuclease digestion with *MmeI* or *Tsp45I* for *SvGLK1* and *SvGLK2*, respectively, following the manufacturers' guidelines.

*Z. mays* Hi-II was transformed by the Wisconsin Crop Innovation Centre (University of Wisconsin) according to their transformation pipeline. The presence of the t-DNA was assessed in T1 lines by PCR (Table S1; JLF\_Ox178 and JLF\_Ox179) and *Zmglk1* mutants were identified by PCR amplification of the region flanking the expected mutation site (Table S1; JLF\_Ox39 and JLF\_Ox175) followed by endonuclease digestion with *HinfI* according to the manufacturer's instructions.

For all lines, mutated sequences were verified by Sanger sequencing (Source BioScience, UK). Experimental analyses were performed on homozygous plants in the T2 or subsequent generations.

### Confocal imaging

The mid-portion of fully expanded leaf 7 from 21 DAS *S. viridis* plants was fixed in 4% paraformaldehyde/0.2% glutaraldehyde in

25 mM phosphate buffer pH 7.2. After fixation, leaves were stored in 0.2 M Na-EDTA pH 9.0. Free-hand cross sections were imaged using a Leica SP9 confocal microscope with an Argon laser set to excite chlorophyll at 663 nm and to emit from 650–800 nm.

### Ultrastructure analysis via transmission electron microscopy

Sample fixation and preparation were performed by the Electron Microscopy Facility team at the Sir William Dunn School of Pathology (University of Oxford). Leaf discs of 2 mm diameter were collected from leaf 5 of 40 DAS maize or 25 DAS *S. viridis* plants and placed into 3% glutaraldehyde in 0.06 M Sorenson's phosphate buffer pH 7.2. Samples were transferred to a Leica AMW microwave tissue processor for fixation, osmification (with 1% OsO<sub>4</sub>), dehydration with an acetone series, and resin infiltration (LRW Hard). The samples were then removed from the microwave and transferred to fresh resin for 4 days, renewing the resin twice a day. The embedded samples were transferred into gelatin capsules and filled with resin for polymerization at 60°C for about 65 h. Sections of 150 nm were cut using a Leica UC7 ultramicrotome, collected onto formvar-coated copper slot grids and post-stained with lead citrate. Images were acquired on a Thermo Fischer Tecnai T12 TEM with a Gatan OneView CMOS camera.

Thylakoid occupancy of chloroplasts was quantified using ImageJ. The thylakoid area was measured by adjusting the image threshold to create a binary mask on which only the thylakoid membrane was apparent. This area value was divided by the chloroplast planar area to give the thylakoid occupancy level.

### Chlorophyll extraction and quantification

Leaf discs were harvested from fully expanded leaf 5 at 30 DAS for maize and 21 DAS for *S. viridis*, and immediately frozen in liquid nitrogen. For total chlorophyll extraction, finely powdered tissue was resuspended in 80% acetone buffered in 100 mM Tris-HCl pH 8.0 (Chazaux et al., 2022). Following a centrifugation step, the absorbance of each sample was read in technical triplicates in a FLUOstar Omega spectrometer (BMG Labtech) at 646 and 663 nm. Background absorbance was measured at 750 nm. Absorbance values recorded at 750 nm were subtracted from values recorded at 646 and 663 nm for normalization purposes. Technical replicates were averaged for each sample. Concentration of chlorophylls *a* + *b* were calculated in micrograms (μg) per leaf area using the following equation:  $(19.54 \times \text{Abs}_{646} + 8.29 \times \text{Abs}_{663}) \times \text{dilution factor}$  (Porra et al., 1989).

### Fluorescence quantification of PSII

Photosystem II (PSII) fluorescence was measured using a LiCOR Li-6800 portable photosynthesis system. LiCOR chamber conditions were set to a flow intensity of 500 μmol sec<sup>-1</sup>, delta pressure of 0.2 kPa, relative air humidity of 50%, reference concentration of CO<sub>2</sub> at 400 μmol mol<sup>-1</sup> and temperature exchange at 25°C. For quantification of PSII operating efficiency (PhiPSII), plants were adapted to light under the conditions described above for at least 1.5 h and actinic light intensity on the LiCOR chamber was set to 1000 μmol photon m<sup>-2</sup> sec<sup>-1</sup>. After clamping, each leaf was adapted for a few minutes until stabilization was reached before logging the fluorescence value. Quantification of maximum quantum efficiency of PSII photochemistry (Fv/Fm) was performed on plants that were dark-adapted for at least 16 h. LiCOR actinic lights were off. In both cases, the fluorescence action log was set to FoFm (dark), FsFm' (light), rectangular flash type with red target at

8000 μmol photon m<sup>-2</sup> sec<sup>-1</sup>, duration of 1000 ms, output rate of 100 Hz and margin of 5 points.

### Gene expression analysis via qPCR

The mid-portion of fully expanded leaf 5 from at least three biological replicates of 30 DAS maize or 21 DAS *S. viridis* were harvested and immediately frozen in liquid nitrogen. Tissue was ground prior to total RNA extraction using a RNeasy Plant Mini Kit (Qiagen, Manchester, UK) according to the manufacturer's instructions. To avoid any DNA contamination, samples were treated with TURBO DNA-free DNase (Thermo Fischer Scientific, Waltham, CA, USA) according to the manufacturer's guidelines. For first strand cDNA synthesis, 100 ng of total RNA was added to 1 μl of Maxima Enzyme Mix (Thermo Fischer Scientific) and 1x Reaction Mix to a final volume of 10 μl. The mixture was incubated at 25°C for 10 min, followed by 50°C for 15 min. Enzyme inactivation was carried out at 85°C for 5 min.

Amplification reactions were performed in a StepOnePlus Real-Time PCR System (Thermo Fischer Scientific) using SYBR-Green to monitor dsDNA synthesis. The reaction mixtures contained 5 μl of diluted cDNA (1:25), 0.2 μM of each primer and 10 μl of SYBR™ Green PCR Master Mix (Thermo Fischer Scientific) in a total volume of 10 μl. The reaction cycles began with a 5 min denaturation step at 94°C, followed by 40 amplification cycles of 15 sec at 94°C, 10 sec at 60°C, 15 sec at 72°C. After each cycle, the fluorescence was measured at 60°C for 35 sec. The melting curve was produced by a cycle of 95°C for 15 sec, 60°C for 1 min, 95°C for 30 sec and finally 60°C for 15 sec. Every qPCR reaction was repeated three times to make technical replicates.

Transcript levels for multiple target genes were measured with primers designed to specifically amplify the gene of interest (Table S1). Two reference genes were used to normalize gene expression for maize (Lin et al., 2014) (Table S1; ZmCYP\_F and ZmCYP\_R, ZmEF1α\_F and ZmEF1α\_R) and *S. viridis* (Lambret-Frotté et al., 2015) (Table S1; JLF\_Ox26 and JLF\_Ox27, JLF\_Ox28 and JLF\_Ox29). Exported *Rn* values were used to calculate amplification efficiency for each primer pair and the Cq values for each qPCR reaction using the 'qpcR' package from R (Ritz & Spiess, 2008). Normalized relative expression of the target genes was calculated using the R package 'EasyqpcR' (<https://www.bioconductor.org/packages/2.12/bioc/html/EasyqpcR.html>), according to the ΔΔCq model (Pfaffl, 2001) based on the previously calculated Cq and amplification efficiencies.

### Statistical analysis

Datasets were tested for Normal distributions using the Shapiro-Wilk test (Shapiro & Wilk, 1965). For thylakoid coverage, chlorophyll concentration and PSII fluorescence, variance was calculated using a one-way ANOVA followed by a pairwise comparison analysis with Tukey's HSD test (Tukey, 1949). For relative gene expression, statistical significance was calculated based on a Student's *t*-test, comparing the mean variance of each mutant line with their respective controls. All statistical analyses were performed using *scipy* v.1.5.0 library and plots were prepared using *matplotlib* v3.5.1 library; both from Python v.3.7.6, using Visual Studio Code as a code editor.

### AUTHOR CONTRIBUTIONS

JL-F and JAL conceived and designed the experiments; GS carried out setaria transformation, plant cultivation, and genotyping; JL-F carried out all of the remaining

experiments and analyzed the data. JL-F and JAL wrote the manuscript.

## ACKNOWLEDGEMENTS

The authors thank Dan Voytas for making available the CRISPR/Cas9 cloning vectors for gene editing in setaria; Raman Dhaliwal, Charlotte Melia, and Errin Jonson for processing and imaging the TEM samples; Ray Collier, Mike Pettersen, and the Wisconsin Crop Innovation Centre team for maize transformation; John Baker for plant photography; Roxaana Clayton, Julie Bull and Lizzie Jamison for technical support; Emma Raven for assisting with initial characterization of mutant lines; Jacob Washburn for providing TPM data for Figure S1b; Sophie Johnson, Chiara Perico, Daniela Vlad, Sovanna Tan, Thomas Hughes and Maricris Zaidem for discussion throughout the experimental work and during manuscript preparation.

## CONFLICT OF INTEREST

No competing interests are declared.

## FUNDING INFORMATION

This research was funded by a BBSRC sLoLa grant (BB/P003117/1), a Newton International Fellowship from The Royal Society to J.L.-F. (NF171598, 2018–2020), and by the Bill and Melinda Gates Foundation C<sub>4</sub> Rice grant awarded to the University of Oxford (2015–2019; OPP1129902).

## DATA AVAILABILITY STATEMENT

All relevant data can be found within the manuscript and its supporting materials.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Figure S1.** *GLK* transcript accumulation in bundle sheath and mesophyll cells. (a) Gene expression in reads per kilobase million (RPKM) in mesophyll (MC) and bundle sheath cells (BSC) of setaria (*SvPEPC*, *SvNADP-ME*, *SvGLK1*, and *SvGLK2*) and maize (*ZmPEPC*, *ZmNADP-ME*, *ZmGLK1*, and *ZmG2*). Maize data were published by Li et al. (2010) and Chang et al. (2012), and setaria data were published by John et al. (2014). *PEPC* and *NADP-ME* transcripts accumulate specifically in mesophyll and bundle sheath cells, respectively, and thus act as markers for cross-contamination of RNA samples between cell types. As such, mesophyll contamination of bundle sheath transcriptomes was 22% (Li et al), 4.5% (Chang et al), and 3.5% (John et al), whereas bundle sheath contamination of mesophyll transcriptomes was 10.7% (Li et al), 7% (Chang et al), and 6.6% (John et al). Disregarding the Li et al. (2010) data because of the high level of cross-contamination, *GLK1* transcript levels were higher in mesophyll than bundle sheath cells by 8-fold in maize and 21-fold in setaria, whereas *GLK2* transcript levels were higher in bundle sheath than mesophyll cells by 4.8-fold in maize and 3.2-fold in setaria. (b) Log<sub>2</sub> fold change between bundle sheath and mesophyll transcripts for *S. viridis* (John et al., 2014), *Z. mays* (Chang et al., 2012; Li et al., 2010), and three different species of Paniceae grasses (*Urochloa fusca*, *Panicum hallii*, and *Digitaria californica*) (Washburn et al. 2022). All species carry out C<sub>4</sub> photosynthesis, but each uses a different decarboxylation pathway in the bundle sheath cells. *PEPC*

was used as a mesophyll cell marker in each case and genes encoding the relevant decarboxylation enzyme (i.e., *NADP-ME*, *PCK*, and *NAD-ME*) were used as bundle sheath cell markers.

**Figure S2.** Phenotypic characterization of three independent *Zmglk1* mutant lines. (a–d) Whole plant phenotype 30 days after sowing. Scale bars = 10 cm. (e–h) Transmission electron microscopy (TEM) images showing mesophyll chloroplast ultrastructure. Scale = 2 μm, 2 μm, 5 μm, and 2 μm, respectively. (i–l) TEM images showing bundle sheath chloroplast ultrastructure. Scale bars = 2 μm, 5 μm, 10 μm, and 2 μm, respectively.

**Figure S3.** Comparison of *GLK* sequences. (a) Phylogenetic reconstruction of the *GLK* gene family. Protein sequences of the extensively characterized *GLK* members *ZmGLK1*, *ZmG2*, *OsGLK1*, and *OsGLK2* were used to create a HMMER profile that was used as a query to search for homologs in whole protein databases of 11 species (*Zea mays*, *Setaria viridis*, *S. italica*, *Panicum virgatum*, *Sorghum bicolor*, *Oryza sativa*, *Brachypodium distachyon*, *Vitis vinifera*, *Arabidopsis thaliana*, *Marchantia polymorpha*, and *Physcomitrium patens*). The presence of the MYB domain in those sequences was confirmed using InterproScan, and only the sequences containing the conserved AREAEAA domain characteristic of the *GLK* subfamily were selected. The resulting 25 amino acid sequences were aligned using Clustal Omega. Phylogenetic reconstruction was performed using Bayesian inference on Mr. Bayes using Poisson with fixed substitution rates as a model. Two independent MCMC runs with 1 000 000 generations were performed and no further improvement of log-likelihood of prior probabilities was observed. The tree was visualized on FigTree. Clade support on each node was assessed using Bayesian posterior probabilities. Members of the *GLK1* subgroup are depicted in green and *GLK2* in blue. (b) Amino acid sequence alignment showing conservation between *ZmGLK1*, *SvGLK1*, *ZmG2*, and *SvGLK2*. The diagonal lines show the HLH associated with MYB or MYB-related proteins; the AREAEAA conserved site is depicted in horizontal lines and the last exon that contains the GCT-box domain is depicted in vertical lines.

**Figure S4.** DNA binding motifs in regulatory regions of *GLK* genes. (a) Schematics of maize, setaria, and rice *GLK* genes showing binding sites revealed from ATAC assays with leaf RNA. Data was retrieved from <https://epigenome.genetics.uga.edu/PlantEpigenome/index.html>. The yellow shading highlights the 800 bp upstream of the transcription start site plus the 5'-UTR regions. (b) Table showing the number of DNA binding motifs found in regulatory regions of each gene. Plant non-redundant motifs (JASPAR) were used to screen the above-mentioned regions of each gene, using the FIMO tool with a *P* value of <0.0001. The photosynthesis-related/light responsive motifs (BPC, ABI3-like, GT-1, GT-4, CNA, MYB-B, STZ, bZIP/bHLHG-box), GATA, MYB-like/l-box, TCP, TGA, and WRKY) were as described by Sing et al. (2023) and rice and maize *GLK*-specific binding sites were as described in Tu et al. (2022). (c,d) Schematics of 800 bp upstream of the transcription start site plus the 5'-UTR regions for each gene, showing the position of all predicted DNA binding motifs (c) and of known light-responsive motifs (d). The 5'-UTR is shown in white and promoter sequence is in gray.

**Figure S5.** Phenotypic characterization of *Svglk2* mutant lines complemented with *ZmGLK1* or *ZmG2*. (a) Schematic of the gene-edited *Svglk2* mutant allele. (b,c) Schematic of the constructs used to express *ZmGLK1* (b) or *ZmG2* (c) in the *Svglk2* background. *HygR* depicts the hygromycin phosphotransferase gene and *OsACT<sub>pro</sub>* and *ZmUBI<sub>pro</sub>* represent the constitutive rice actin and maize ubiquitin promoters, respectively. LB and RB refer to left and right borders. (d–m) Transmission electron microscopy images showing mesophyll (MC) and bundle sheath (BSC) cell chloroplast



ultrastructure in the mutant line (d,e) and in lines overexpressing *ZmGLK1* (f-i) or *ZmG2* (j-m) in the mutant background. Scale bar sizes are indicated on each image. Panels d & e are the same images as Figure 3g,h). (n-r) Confocal images of leaf cross sections from the mutant (n) and from lines overexpressing *ZmGLK1* (o,p) or *ZmG2* (q,r) showing the morphology of chloroplasts. Bundle sheath cells are outlined in white. Scale bars = 50  $\mu$ m. Yellow arrows indicate ectopic chloroplast formation in vascular cells

**Figure S6.** Sequence variability in regulatory regions of *GLK* genes. The SNPs from 3025 variants of rice (SNP seek, IRRI) and 1210 variants of maize (Maize SNPDB) were used to calculate the frequency of each nucleotide for each SNP position. The nucleotide with a higher prevalence in each position is plotted on the graphs for *OsGLK1*, *OsGLK2*, *ZmGLK1*, and *ZmG2*. The 2Kb upstream of the transcription start site is shown in gray, the 5'-UTR in white, and the first exon in black.

**Figure S7.** Alignment of *GLK* gene promoter and 5'-UTR regions. (a, b) *GLK1* (a) and *GLK2*(a) sequences (800 bp upstream of the transcription start site (TSS) plus the 5'-UTR) aligned using MAFFT (Kato et al., 2019) and visualized using JALview (Waterhouse et al., 2009). Conserved bases are indicated by blue shading, the predicted TATA box by the red squares, actual (rice—Murray et al., 2022), or predicted TSS by orange squares and putative uORFs by yellow squares. (c) Amino acid sequences of the predicted uORFs in (a) & (b).

**Table S1.** Oligonucleotides used for cloning, genotyping, and qPCR.

**Table S2.** Data and statistical analyses supporting Figures 1–6.

## REFERENCES

- Bravo-Garcia, A., Yasumura, Y. & Langdale, J.A. (2009) Specialization of the *Golden2-like* regulatory pathway during land plant evolution. *New Phytologist*, **183**, 133–141.
- Chang, Y.-M., Liu, W.-Y., Shih, A.C.-C., Shen, M.-N., Lu, C.-H., Lu, M.-Y.J. et al. (2012) Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. *Plant Physiology*, **160**, 165–177.
- Chazaux, M., Schiphorst, C., Lazzari, G. & Caffari, S. (2022) Precise estimation of chlorophyll a, b and carotenoid content by deconvolution of the absorption spectrum and new simultaneous equations for Chl determination. *Plant Journal*, **109**, 1630–1648.
- Chiang, Y.-H., Zubo, Y.O., Tapken, W., Kim, H.J., Lavanway, A.M., Howard, L. et al. (2012) Functional characterization of the GATA transcription factors GNC and CGA1 reveals their key role in chloroplast development, growth, and division in Arabidopsis. *Plant Physiology*, **160**, 332–348.
- Christin, P.-A. & Osborne, C.P. (2013) The recurrent assembly of *C<sub>4</sub>* photosynthesis, an evolutionary tale. *Photosynthesis Research*, **117**, 163–175.
- Concordet, J.-P. & Haussler, M. (2018) CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research*, **46**, W242–W245.
- Cribb, L., Hall, L. & Langdale, J.A. (2001) Four mutant alleles elucidate the role of the G2 protein in the development of *C<sub>4</sub>* and *C<sub>3</sub>* photosynthesizing maize tissues. *Genetics*, **159**, 787–797.
- Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS One*, **4**, e5553.
- Engler, C. & Marillonnet, S. (2013) Combinatorial DNA assembly using golden gate cloning. *Methods in Molecular Biology*, **1073**, 141–156.
- 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. *Plant Journal*, **31**, 713–727.
- Gallagher, T.J., Peterson, P.M., Soreng, R.J., Zuloaga, F.O., Li, D.-Z., Clark, L.G. et al. (2022) Grasses through space and time: an overview of the biogeographical and macroevolutionary history of Poaceae. *Journal of Systematics and Evolution*, **60**, 522–569.
- Hall, L.N., Rossini, L., Cribb, L. & Langdale, J.A. (1998) *GOLDEN 2*: a novel transcriptional regulator of cellular differentiation in the maize leaf. *Plant Cell*, **10**, 925–936.
- He, Q., Tang, S., Zhi, H., Chen, J., Zhang, J., Liang, H. et al. (2023) A graph-based genome and pan-genome variation of the model plant *Setaria*. *Nature Genetics*, **55**, 1232–1242.
- Hua, L., Stevenson, S.R., Reyna-Llorens, I., Xiong, H., Kopriva, S. & Hibberd, J.M. (2021) The bundle sheath of rice is conditioned to play an active role in water transport as well as sulfur assimilation and jasmonic acid synthesis. *Plant Journal*, **107**, 268–286.
- Hudson, D., Guevara, D.R., Hand, A.J., Xu, Z., Hao, L., Chen, X. et al. (2013) Rice Cytokinin GATA transcription factor1 regulates chloroplast development and plant architecture. *Plant Physiology*, **162**, 132–144.
- John, C.R., Smith-Unna, R.D., Woodfield, H., Covshoff, S. & Hibberd, J.M. (2014) Evolutionary convergence of cell-specific gene expression in independent lineages of *C<sub>4</sub>* grasses. *Plant Physiology*, **165**, 62–75.
- Katoh, K., Rozewicki, J. & Yamada, K.D. (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, **20**, 1160–1166.
- Kenrick, P. & Crane, P.R. (1997) The origin and early evolution of plants on land. *Nature*, **389**, 33–39.
- Kobayashi, K., Baba, S., Obayashi, T., Sato, M., Toyooka, K., Keränen, M. et al. (2012) Regulation of root greening by light and auxin/cytokinin signaling in Arabidopsis. *Plant Cell*, **24**, 1081–1095.
- Lambret-Frotté, J., De Almeida, L.C.S., De Moura, S.M. & Souza, F.L.F. (2015) Validating internal control genes for the accurate normalization of qPCR expression analysis of the novel model plant *Setaria viridis*. *PLoS One*, **10**, e0135006.
- Langdale, J.A. (2011) *C<sub>4</sub>* cycles: past, present, and future research on *C<sub>4</sub>* photosynthesis. *The Plant Cell*, **23**, 3879–3892.
- Langdale, J.A. & Kidner, C.A. (1994) *Bundle sheath defective*, a mutation that disrupts cellular differentiation in maize leaves. *Development*, **120**, 673–681.
- Li, P. & Brutnell, T.P. (2011) *Setaria viridis* and *Setaria italica*, model genetic systems for the Panicoid grasses. *Journal of Experimental Botany*, **62**, 3031–3037.
- Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S.L. et al. (2010) The developmental dynamics of the maize leaf transcriptome. *Nature Genetics*, **42**, 1060–1067.
- Li, X., Wang, P., Li, J., Wei, S., Yan, Y., Yang, J. et al. (2020) Maize *GOLDEN2-LIKE* genes enhance biomass and grain yields in rice by improving photosynthesis and reducing photoinhibition. *Communications Biology*, **3**, 151.
- Lin, Y., Zhang, C., Lan, H., Gao, S., Liu, H., Liu, J. et al. (2014) Validation of potential reference genes for qPCR in maize across abiotic stresses, hormone treatments, and tissue types. *PLoS One*, **9**, e95445.
- Marshall, D.M., Muhaidat, R., Brown, N.J., Liu, Z., Stanley, S., Griffiths, H. et al. (2007) Cleome, a genus closely related to Arabidopsis, contains species spanning a developmental progression from *C<sub>3</sub>* to *C<sub>4</sub>* photosynthesis. *Plant Journal*, **51**, 886–896.
- Murray, A., Mendieta, J.P., Vollmers, C. & Schmitz, R.J. (2022) Simple and accurate transcriptional start site identification using Smar2C2 and examination of conserved promoter features. *The Plant Journal*, **112**, 583–596.
- Nguyen, C.V., Vrebalov, J.T., Gapper, N.E., Zheng, Y., Zhong, S., Fei, Z. et al. (2014) Tomato *GOLDEN2-LIKE* transcription factors reveal molecular gradients that function during fruit development and ripening. *Plant Cell*, **26**, 585–601.
- Nickelsen, J. & Rengstl, B. (2013) Photosystem II assembly: from cyanobacteria to plants. *Annual Review of Plant Biology*, **64**, 609–635.
- Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, **29**, e45.
- Porra, R.J., Thompson, W.A. & Kriedemann, P.E. (1989) Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **975**, 384–394.
- Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J. et al. (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, **290**, 2105–2110.
- Ritz, C. & Spiess, A.-N. (2008) qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics*, **24**, 1549–1551.

- Rossini, L., Cribb, L., Martin, D.J. & Langdale, J.A. (2001) The maize *Golden2* gene defines a novel class of transcriptional regulators in plants. *Plant Cell*, **13**, 1231–1244.
- Safi, A., Medici, A., Szponarski, W., Ruffel, S., Lacombe, B. & Krouk, G. (2017) The world according to GARP transcription factors. *Current Opinion in Plant Biology*, **39**, 159–167.
- Sagan, L. (1967) On the origin of mitosing cells. *Journal of Theoretical Biology*, **14**, 225–IN6.
- Sage, R.F., Christin, P.-A. & Edwards, E.J. (2011) The C<sub>4</sub> plant lineages of planet earth. *Journal of Experimental Botany*, **62**, 3155–3169.
- Singh, P., Stevenson, S.R., Dickinson, P.J., Reyna-Llorens, I., Tripathi, A., Reeves, G., et al. (2023) C<sub>4</sub> gene induction during de-etiolation evolved through changes in cis to allow integration with ancestral C<sub>3</sub> gene regulatory networks. *Science Advances*, **9**, eade9756.
- Shapiro, S.S. & Wilk, M.B. (1965) An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591–611.
- Tausta, L.S., Li, P., Si, Y., Gandotra, N., Liu, P., Sun, Q. et al. (2014) Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C<sub>4</sub>-related processes. *Journal of Experimental Botany*, **65**, 3543–3555.
- Theune, M.L., Bloss, U., Brand, L.H., Ladwig, F. & Wanke, D. (2019) Phylogenetic analyses and GAGA-motif binding studies of BBR/BPC proteins lend to clues in GAGA-MOTIF recognition and a regulatory role in brassinosteroid signaling. *Frontiers in Plant Science*, **10**, e00466.
- Tu, X., Ren, S., Shen, W., Li, J., Li, Y., Li, C. et al. (2022) Limited conservation in cross-species comparison of GLK transcription factor binding suggested wide-spread cis-trome divergence. *Nature Communications*, **13**, 7632.
- Tukey, J.W. (1949) Comparing individual means in the analysis of variance. *Biometrics*, **5**, 99–114.
- Van Eck, J. & Startwood, K. (2015) *Setaria viridis*. In: Wang, K. (Ed.) *Agrobacterium Protocols*. New York, NY: Springer, p. 417.
- Vlad, D., Abu-Jamous, B., Wang, P. & Langdale, J.A. (2019) A modular steroid-inducible gene expression system for use in rice. *BMC Plant Biology*, **19**, 426.
- Wang, L., Czedik-Eysenberg, A., Mertz, R.a., Si, Y., Tohge, T., Nunes-Nesi, A. et al. (2014) Comparative analyses of C<sub>4</sub> and C<sub>3</sub> photosynthesis in developing leaves of maize and rice. *Nature Biotechnology*, **32**, 1158–1165.
- Wang, P., Fouracre, J., Kelly, S., Karki, S., Gowik, U., Aubry, S. et al. (2013) Evolution of *GOLDEN2-LIKE* gene function in C<sub>3</sub> and C<sub>4</sub> plants. *Planta*, **237**, 481–495.
- Washburn, J.D., Strable, J., Dickinson, P., Kothapalli, S.S., Brose, J.M., Covshoff, S. et al. (2021) Distinct C<sub>4</sub> sub-types and C<sub>3</sub> bundle sheath isolation in the Paniceae grasses. *Plant Direct*, **5**, 1–14.
- Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. & Barton, G.J. (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
- Waters, M.T. & Langdale, J.A. (2009) The making of a chloroplast. *The EMBO Journal*, **28**, 2861–2873.
- Waters, M.T., Moylan, E.C. & Langdale, J.A. (2008) GLK transcription factors regulate chloroplast development in a cell-autonomous manner. *Plant Journal*, **56**, 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**, 1109–1128.
- Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N. et al. (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences*, **111**, E4859–E4868.
- Woo, K.C., Anderson, J.M., Boardman, N.K., Downton, W.J.S., Osmond, C.B. & Thorne, S.W. (1970) Deficient photosystem II in Agranal bundle sheath chloroplasts of C<sub>4</sub> plants. *Proceedings of the National Academy of Sciences*, **67**, 18–25.
- Yasumura, Y., Moylan, E.C. & Langdale, J.A. (2005) A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. *Plant Cell*, **17**, 1894–1907.
- Zhang, T., Wu, A., Yue, Y. & Zhao, Y. (2020) uORFs: important cis-regulatory elements in plants. *International Journal of Molecular Sciences*, **21**, 1–4.
- Zubo, Y.O., Blakley, I.C., Franco-Zorrilla, J.M., Yamburenko, M.V., Solano, R., Kieber, J.J. et al. (2018) Coordination of chloroplast development through the action of the GNC and GLK transcription factor families. *Plant Physiology*, **178**, 130–147.