GOLDEN2-LIKE1 IS SUFFICIENT BUT NOT NECESSARY FOR CHLOROPLAST BIOGENESIS IN MESOPHYLL CELLS OF C_4 GRASSES

Julia Lambret-Frotte[1,2], Georgia Smith[1] & Jane A. Langdale[1*]

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

[2]Current address: NIAB, Park Farm, Villa Road, Impington, CB24 9NZ, Cambridge, UK

[*]For correspondence: jane.langdale@biology.ox.ac.uk

ABSTRACT

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_4 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_4 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 *Setaria viridis* mutants. Collectively our results suggest an evolutionary trajectory in C_4 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.

KEY WORDS: C_4, chloroplasts, development, maize, green foxtail

46 **INTRODUCTION**

47 Chloroplasts are endosymbiotic organelles and are the site of photosynthetic reactions in a range of
48 organisms from algae to land plants (Sagan 1967). Given the importance of chloroplast function for
49 autotrophic growth, development of the organelle is a tightly controlled process that involves many
50 levels of regulation (Waters and Langdale 2009). First described in the mid-1990s, the *golden2* (*g2*)
51 mutant in *Zea mays* (maize) exhibits pale green leaves and impaired chloroplast development
52 specifically in bundle sheath cells (Langdale and Kidner 1994, Hall *et al.* 1998). Subsequent
53 characterization revealed that *ZmG2* encodes a transcription factor (Rossini *et al.* 2001), a founding
54 member of the GARP transcription factor family (Riechmann *et al.* 2000). In most plant species
55 examined, two *G2*-like (*GLK*) genes are present in the genome (Wang *et al.* 2013). *GLK* genes are
56 defined by a characteristic helix-turn-helix DNA-binding domain related to MYB or MYB-like
57 transcriptional factors, a conserved AREAEAA motif located immediately after the last helix and a GCT-
58 box in the C-terminal region that is responsible for either homo- or heterodimerization (Rossini *et al.*
59 2001, Fitter *et al.* 2002, Safi *et al.* 2017). Functional characterization of *GLK* genes in maize and
60 *Arabidopsis thaliana* (Arabidopsis) has shown that they are involved in the transcriptional activation of
61 multiple photosynthesis-related genes including those encoding proteins of the light-harvesting
62 complexes and enzymes of the chlorophyll biosynthesis pathway (Cribb *et al.* 2001, Waters *et al.* 2009).
63 This pivotal role in chloroplast biogenesis has been associated with a myriad of physiological processes
64 including stomatal movement, hormone signalling, defence responses and fruit ripening (Kobayashi *et*
65 *al.* 2012, Nguyen *et al.* 2014, Li *et al.* 2020). The emerging picture is one where GLK proteins act as
66 linchpins in any metabolic or developmental pathway that requires chloroplast function to be
67 modulated.

68

69 During the evolution of land plants, the *GLK* gene family underwent multiple independent duplication
70 events (Wang *et al.* 2013). In cases such as Arabidopsis where duplications likely occurred within
71 species, the resulting paralogs are functionally redundant (Waters *et al.* 2009). However, duplication
72 events in the monocots are more intriguing. The common ancestor of the Poales underwent a whole
73 genome duplication that subdivided the *GLK* gene family into two sister clades: *GLK1* and *GLK2* (Wang
74 *et al.* 2013). Notably, this event occurred before the many independent speciation events that led to
75 the evolution of C_4 photosynthesis within the Poales group (Christin and Osborne 2013). In monocot
76 species such as rice that perform C_3 photosynthesis, the *GLK* paralogs are functionally redundant as
77 in Arabidopsis (Waters *et al.* 2009, Wang *et al.* 2013). Unlike C_3 photosynthesis, where carbon reactions
78 occur predominantly in mesophyll cell chloroplasts, in C_4 plants carbon assimilation and fixation
79 reactions are compartmentalized in the chloroplasts of mesophyll and bundle sheath cells, respectively
80 (Langdale 2011). This biochemical compartmentalization is often accompanied by anatomical
81 differences in chloroplast ultrastructure (Woo *et al.* 1970, Sage *et al.* 2011). In the C_4 monocot species
82 maize, *Setaria viridis* (setaria) and *Sorghum bicolor* (sorghum), *GLK1* is expressed preferentially in

83 mesophyll cells whereas *GLK2* transcripts accumulate preferentially in bundle sheath cells, with the
84 level of specificity differing between species (Li *et al.* 2010, Chang *et al.* 2012, Wang *et al.* 2013, John
85 *et al.* 2014, Tausta *et al.* 2014). Notwithstanding that protein levels dictate functional capacity, defects
86 in bundle sheath but not mesophyll chloroplast development in *g2* mutants of maize (Langdale and
87 Kidner 1994, Cribb *et al.* 2001, Rossini *et al.* 2001) demonstrated that the preferential localization of
88 *ZmG2* transcripts in bundle sheath cells is functionally relevant and that *GLK* paralogs in C_4 grasses
89 can control chloroplast development in a cell-specific manner.
90
91 *GLK* paralogs have been extensively characterized in Arabidopsis, however, little is known about gene
92 function in monocots, and in particular the function of *GLK1* genes has not been investigated. Because
93 *GLK1* transcripts accumulate at much higher levels in mesophyll cells than bundle sheath cells of both
94 maize (10-fold) and setaria (30-fold) (Chang et al. 2012, John et al. 2014; Figure S1), and *Zmg2* mutants
95 of maize show bundle sheath specific chloroplast defects, it has been hypothesized that GLK1 controls
96 chloroplast biogenesis specifically in mesophyll cells of C_4 grasses (Rossini *et al.* 2001). Here we test
97 this hypothesis by generating and characterizing loss-of-function mutants in maize and setaria, two
98 species that represent independent origins of C_4. Our findings reveal that contrary to expectations,
99 neither *ZmGLK1* nor *SvGLK1* are necessary for chloroplast development in mesophyll cells whereas
100 loss of *SvGLK2* function perturbs chloroplast development in both bundle sheath and mesophyll cells.
101
102 **RESULTS**
103 ***ZmGLK1* is not necessary for chloroplast biogenesis in mesophyll cells of maize**
104 To determine whether *ZmGLK1* regulates chloroplast biogenesis in maize, loss-of-function mutants
105 were generated using CRISPR/Cas9 (Figure 1A). Four independent mutant lines (referred to
106 collectively as *Zmglk1*) were obtained, all of which exhibited the same overall growth phenotype as
107 segregating wild-type siblings (Figure 1B, C; Figure S2). Ultrastructure analysis revealed that
108 chloroplasts in *Zmglk1* lines were similar to wild-type, with well-developed granal thylakoids in
109 mesophyll cells and agranal thylakoids in bundle sheath cells (Figure 1D-I; Figure S2). Total chlorophyll
110 levels were also similar in wild-type and *Zmglk1* mutants (Figure 1J). As such, we concluded that
111 *ZmGLK1* is not necessary for thylakoid assembly or chlorophyll biosynthesis in bundle sheath or
112 mesophyll chloroplasts of maize.
113
114 An evaluation of chloroplast function in *Zmglk1* mutants was made by measuring photosystem II (PSII)
115 parameters. PSII is a large protein complex comprised of the antenna and chlorophyll molecules that
116 harvest light and together initiate the electron transfer reactions of photosynthesis (Nickelsen and
117 Rengstl 2013). In maize, PSII accumulates in the granal thylakoids of mesophyll cells and is largely
118 absent from the agranal bundle sheath cell chloroplasts (Woo *et al.* 1970). As such, PSII function is a
119 proxy for mesophyll cell chloroplast function. Fluorescence measurements were therefore taken in light

120 and dark-adapted plants to quantify PSII operating efficiency (PhiPSII) and maximum quantum
121 efficiency of PSII photochemistry (Fv/Fm), in both wild-type and *Zmglk1* mutant plants. Figures 1K-L
122 show that PSII is operating normally in *Zmglk1* mutants, suggesting that *ZmGLK1* is not necessary for
123 chloroplast biogenesis or function in maize mesophyll cells.
124
125 ***SvGLK1* is not necessary for chloroplast biogenesis in mesophyll cells of setaria**
126 Maize and setaria have different origins of C_4 photosynthesis within the PACMAD clade (Li and Brutnell
127 2011) but like the orthologous *ZmGLK1*, *SvGLK1* transcripts accumulate preferentially in mesophyll
128 cells (John *et al.* 2014). Phylogenetic analysis and amino acid sequence comparisons revealed that
129 *SvGLK1* shares many structural features with *ZmGLK1* (Figure S3). To assess whether *SvGLK1* plays
130 a role in chloroplast biogenesis, two independent loss-of-function mutants were generated in setaria
131 using CRISPR/Cas9 (Figure 2A). As in maize, *Svglk1* mutants exhibited overall plant growth (Figure 2B-
132 D) and chloroplast ultrastructure (Figure 2E-J) phenotypes that were similar to segregating wild-type
133 siblings, with thylakoid membranes developing normally in both mesophyll and bundle sheath cells
134 (Figure 2E-L). Furthermore, there was no significant difference in chlorophyll levels between *Svglk1*
135 mutants and wild-type (Figure 2M), and electron flow through PSII was unperturbed (Figure 2N-O).
136 Collectively, these results suggest that the preferential expression of *GLK1* genes in mesophyll cells of
137 C_4 grass species does not reflect a requirement for gene function during chloroplast development
138 and/or maintenance.
139
140 ***SvGLK2* regulates chloroplast development in both mesophyll and bundle sheath cells of setaria**
141 Given that dimorphic chloroplast development is observed in bundle sheath and mesophyll cells of
142 both maize and setaria, we next sought to determine whether *SvGLK2,* like its ortholog *ZmG2* (Figure
143 S3), regulates bundle sheath chloroplast development. To this end, CRISPR/Cas9 was used to generate
144 loss-of-function *Svglk2* mutants (Figure 3A). In contrast to maize where *g2* mutants are weaker than
145 wild-type but are viable and fertile, any *Svglk2* mutant plants that regenerated after transformation did
146 not progress beyond the seedling stage before dying. The only viable line recovered from tissue culture
147 was heterozygous for *SvGLK2* and this was used to isolate a null homozygous line via self-pollination.
148 *Svglk2* 8G5 mutants, henceforth referred to as *Svglk2*, exhibited severe growth defects, had pale yellow
149 leaves and failed to complete the life cycle (Figure 3B). Leaf cross sections showed misshapen and
150 smaller chloroplasts in the bundle sheath cells of *Svglk2* mutants compared to wild-type (Figure 3C-D),
151 and observations of chloroplast ultrastructure revealed impaired chloroplast anatomy with
152 discontinuous thylakoids that occupied less of the organelle than in wild-type (Figure 3E-F), albeit a
153 reduction that was not statistically significant (Figure 3I). More striking, given the bundle sheath specific
154 perturbations observed in maize *g2* mutants, was the aberrant phenotype of mesophyll chloroplasts in
155 *Svglk2* mutants. Compared to wild-type there was a significant reduction in thylakoid membrane
156 coverage (15% less than wild-type) and a noticeable reduction in the size and frequency of grana

157 (Figure 3G-H, J). As in bundle sheath cells, thylakoid membranes in mutant mesophyll chloroplasts
158 appeared fragmented (Figure 3H). The observed perturbations in bundle sheath and mesophyll cells
159 were accompanied by a reduction in leaf chlorophyll content of more than 80% (Figure 3K). Together,
160 the *Svglk2* and *Svglk1* mutant phenotypes suggest that SvGLK1 cannot compensate for loss of SvGLK2
161 function whereas SvGLK2 can compensate for loss of SvGLK1 function. As such, despite *SvGLK2*
162 transcripts comprising only 43% of the total *GLK* transcript pool in mesophyll cells (as compared to
163 98% in bundle sheath cells) (John *et al.* 2014), SvGLK2 is the primary regulator of chloroplast
164 biogenesis in both cell-types.
165
166 ***GLK2* transcript levels are not upregulated in *glk1* mutants of maize or setaria**
167 Given that *ZmGLK1* and *SvGLK1* transcripts represent 95% and 57% of the total *GLK* transcript pool in
168 mesophyll cells of the respective species (Chang et al. 2012, John et al. 2014; Figure S1), the absence
169 of chloroplast defects in this cell-type in loss of function mutants is intriguing. It is particularly notable
170 because all *GLK* genes characterized to date have a role in chloroplast biogenesis, albeit a redundant
171 role alongside paralogs in many cases (Fitter *et al.* 2002, Wang *et al.* 2013). Even a redundant role is
172 unlikely in setaria given the *Svglk2* mutant phenotype. However, redundancy could explain the lack of
173 a *Zmglk1* mutant phenotype if *ZmG2* transcripts are translated very efficiently in mesophyll cells of
174 maize. Alternatively, loss of *ZmGLK1* function could lead to altered regulation of *ZmG2* expression
175 and/or to the upregulation of genes encoding the GATA transcription factors GNC and CGA1 which
176 also regulate chloroplast development (Chiang *et al.* 2012, Hudson *et al.* 2013) and have been shown
177 to partially rescue loss of GLK function in Arabidopsis (Zubo *et al.* 2018). To test this hypothesis,
178 transcript levels of *ZmGLK1*, *ZmG2*, *ZmGNC* and *ZmCGA1* were quantified in maize using the *Zmglk1*
179 gene edited mutants generated above alongside the previously reported transposon-induced *Zmg2-*
180 *bsd1-s1* mutant (Cribb *et al.* 2001). In each case the mutant allele produces transcripts that encode
181 non-functional protein(s). Figure 4 reveals that loss of *Zmglk1* function does not lead to consistently
182 elevated levels of *ZmGLK1* (Figure 4A), *ZmG2* (Figure 4B), *ZmGNC* (Figure 4C) or *ZmCGA1* (Figure
183 4D) transcripts. A small but significant increase in *ZmG2*, *ZmGNC* and *ZmCGA1* transcript levels was
184 seen in one of the four *Zmglk1* mutant lines but that cannot explain how chloroplasts develop normally
185 in all four lines. We thus conclude that transcription of *ZmG2, ZmGNC* and *ZmCGA1* is not upregulated
186 in mesophyll cells when *Zmglk1* function is lost.
187
188 Whereas loss of ZmGLK1 function in maize does not impact on *ZmGLK1* or *ZmG2* transcript levels,
189 loss of ZmG2 function in bundle sheath cells leads to a statistically significant reduction in the
190 accumulation of both *ZmGLK1* and *ZmG2* transcripts (Figure 4A, B), and a similar situation is seen in
191 setaria where both *SvGLK1* and *SvGLK2* transcript levels are dramatically reduced in *Svglk2* mutants
192 (Figure 4E, F). This observation suggests that the ZmG2/SvGLK2 orthologs, either directly or indirectly,
193 promote the accumulation of both *GLK1* and *GLK2* transcripts in the respective species. Consistent

194 with this suggestion, multiple G-box elements which are bound by Arabidopsis GLK proteins (Waters
195 *et al.* 2009) are present in *ZmGLK1*, *ZmG2* and *SvGLK2* promoter sequences (Figure S4), and a GLK
196 binding site (Tu *et al.* 2022) is present 1840 bp upstream of the ATG in the *SvGLK1* gene. Reduced
197 levels of *SvGLK1* transcripts in one of the two *Svglk1* mutant lines (Figure 4E) suggest that *SvGLK1*
198 might regulate its own transcript levels via this binding site but the absence of such an effect in the
199 other mutant line makes it equally likely that the specific edits in line #2062 lead to transcript turnover,
200 possibly via the nonsense-mediated decay pathway. Collectively these data suggest that chloroplast
201 development in C_4 grasses is predominantly regulated by genes of the *GLK2* clade.
202
203 ***ZmGLK1* and *ZmG2* are both sufficient to restore dimorphic chloroplast development in**
204 **mesophyll and bundle sheath cells of *Svglk2* mutants**
205 Our findings thus far show that *SvGLK2* and *ZmG2* play a more central role in controlling chloroplast
206 biogenesis in C_4 grasses than their respective *GLK1* paralogs. To evaluate whether this difference in
207 necessity also applies to sufficiency, *ZmGLK1* and *ZmG2* were expressed under the control of the
208 constitutive maize ubiquitin gene promoter ($ZmUBI_{pro}$) in the chloroplast-defective *Svglk2* mutant
209 background. Three independent lines were obtained for each construct and quantitative analysis of
210 transgene transcript levels revealed differences between lines that correlated directly with plant
211 phenotype (Figure 5A-D). For example, the weaker plants in line #17808-15254 accumulated around
212 half the level of *ZmGLK1* transcripts seen in healthier plants of the other two lines (Figure 5A, 5C).
213 Notably plants of all six lines were viable, healthy and fertile, suggesting complementation of the *Svglk2*
214 mutant phenotype (compare *Svglk2* plants in Figure 3B with Figure 5A, B).
215
216 Given that loss of *Svglk2* function leads to reduced levels of both *SvGLK1* and *SvGLK2* transcripts
217 (Figure 4E, F), we next sought to evaluate whether ZmGLK1 or ZmG2 could influence endogenous
218 *SvGLK1* or *SvGLK2* expression. Figure 5E, F show that *SvGLK1* and *SvGLK2* transcripts were restored
219 to wild-type levels in *Svglk2* lines expressing either *ZmGLK1* or *ZmG2* transgenes. In fact, regardless
220 of the expression level of either *ZmGLK1* or *ZmG2*, all complemented lines exhibit the same wild-type-
221 like level of *SvGLK1* and *SvGLK2* expression (Figure 5C-F). Both ZmGLK1 and ZmG2 can thus activate
222 expression of *SvGLK1* and *SvGLK2*, providing further support for the existence of a regulatory network
223 between *GLK* genes and for the likely relevance of G-box and GLK binding motifs in the promoter
224 regions (Figure S4). The functional significance of this observation must be considered in the context
225 of transgene expression driven by a constitutive promoter and gene-edited *Svglk2* transcripts encoding
226 non-functional proteins (and hence any effect on *SvGLK1* expression is a direct effect of the maize
227 transgene and not via *SvGLK2*). In this regard, we must deduce that the transgenic lines expressing
228 *ZmGLK1* accumulate transcripts encoding functional ZmGLK1 and SvGLK1 proteins in both mesophyll
229 and bundle sheath cells, and those expressing *ZmG2* accumulate transcripts encoding functional
230 ZmG2 and SvGLK1 proteins, similarly in both cell types.

231 To determine whether *ZmGLK1* and/or *ZmG2* can rescue chloroplast ultrastructure defects in
232 mesophyll and/or bundle sheath cells of *Svglk2* mutants, confocal images and transmission electron
233 micrographs of transverse leaf sections from all six transgenic lines were examined. In all cases,
234 chloroplasts were restored to normal size (Figure 6A-C, Figure S5), with predominantly agranal
235 thylakoids in bundle sheath cell chloroplasts (Figure 6D-F, Figure S5) and granal thylakoids in
236 mesophyll chloroplasts (Figure 6G-I, Figure S5). Quantification revealed that with the exception of the
237 one line that had low levels of transgene expression (#17808-15254), thylakoid occupancy levels in
238 mesophyll (Figure 6J) and bundle sheath (Figure 6K) chloroplasts of transgenic lines were statistically
239 equivalent (or in one case higher) than in wild-type plants. Consistent with ectopic expression of *GLK*
240 genes, mature chloroplasts were also observed in the leaf vasculature of all complemented lines, more
241 with *ZmG2* than *ZmGLK1* (Figure 6A-C, Figure S5). Irrespective of which maize ortholog was
242 overexpressed, the total chlorophyll content in leaves of transgenic plants was restored to wild-type
243 levels (Figure 6L). Notwithstanding the fact that SvGLK1 may be functional in bundle sheath and
244 mesophyll cells of all transgenic plants, collectively these data demonstrate that ZmGLK1 and ZmG2
245 proteins are functionally equivalent in the context of the C_4 setaria leaf and that the dimorphic identity
246 of bundle sheath and mesophyll cell chloroplasts develops regardless of which maize ortholog is
247 overexpressed.
248
249 **DISCUSSION**
250 Prior to this study it was known that *GLK* genes are duplicated in many land plant species and that in
251 species that carry out C_3 photosynthesis the duplicate genes act redundantly to regulate chloroplast
252 biogenesis in all photosynthetic cells (Waters *et al.* 2008, Bravo-Garcia *et al.* 2009, Wang *et al.* 2013).
253 It had also been hypothesized that *GLK1* and *GLK2* genes regulate dimorphic chloroplast development
254 in the two photosynthetic cell-types of C_4 species, with each gene functioning in one of the two cell-
255 types (Rossini *et al.* 2001). This hypothesis was based on the observation that *GLK1* genes are
256 preferentially expressed in mesophyll cells of maize, sorghum and setaria whereas *GLK2/G2* genes are
257 preferentially expressed in bundle sheath cells (Li *et al.* 2010, Chang *et al.* 2012, John *et al.* 2014, Wang
258 *et al.* 2014), and loss of *Zmg2* function in maize perturbs bundle sheath but not mesophyll chloroplast
259 development (Cribb *et al.* 2001). Through the phenotypic characterization of *glk1* loss-of-function
260 mutants in maize and setaria (Figures 1, 2 & 4), *glk2* loss-of-function mutants in setaria (Figures 3 & 4),
261 and *glk2* setaria mutants complemented with maize *GLK1* and *GLK2* genes (Figure 5 & 6), we have
262 revealed a more complex situation both with respect to the developmental processes operating within
263 individual C_4 species and to likely evolutionary trajectories.
264
265 **GLK proteins have equivalent functional capacity**
266 *GLK* genes duplicated in a common ancestor of the Poales, before the speciation of maize, sorghum,
267 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* (Yasumura *et al.* 2005, Bravo-Garcia *et al.* 2009), which last shared a common ancestor with flowering plants over 450 million years ago (Kenrick and 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 85-95 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 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 & 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 of *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.

305 Whereas normal mesophyll chloroplast development in the absence of GLK1 function can be explained
306 in setaria because SvGLK2 likely compensates for loss of GLK1 function, the mutant phenotype in
307 maize was unexpected. In every land plant species examined to date, a role for *GLK* genes in
308 chloroplast development has been revealed, albeit often a redundant role with paralogs operating in
309 the same cell-type. Because ZmGLK1 is highly expressed in maize mesophyll cells and no other GLK-
310 like transcripts appear as possible candidates for a redundant paralog, a previously unknown
311 mechanism must be operating in this cell-type. Although GLK gene function has previously been shown
312 to be cell-autonomous in Arabidopsis, it is possible that a non-cell-autonomous mechanism could
313 operate in maize. Indeed, there is some evidence of this because *ZmGLK1* transcript levels are reduced
314 when G2 function is lost, even though *G2* transcript levels in wild-type mesophyll cells are only 5% of
315 the GLK pool (Figure 4). In such a scenario, loss of GLK1 function in mesophyll cells would be
316 compensated for by direct or indirect activity of G2 proteins in the bundle sheath cells. Any such effect
317 would only be revealed in double *Zmglk1;g2* mutants, which would have defective chloroplasts in both
318 cell-types. Regardless of whether ZmG2 compensates for loss of ZmGLK1 function, the GLK2 paralog
319 plays the dominant role in chloroplast biogenesis in maize, as in setaria.
320
321 ***cis*-regulatory sequences are more variable in *GLK2* than *GLK1* genes**
322 Given that *GLK* gene paralogs encode functionally equivalent proteins both in C_3 species where they
323 act redundantly and in the C_4 species maize, the question arises of why the paralogs are expressed in
324 a cell-type preferential manner in C_4 grass species (at least in the ones that have been examined).
325 Estimates based on the topology of BEP (C_3 species only) and PACMAD (C_3 & C_4 species) clades in
326 the Poales suggest that maize diverged from rice ~80 million years ago (Gallaher *et al.* 2022) and that
327 within the PACMAD clade, maize and sorghum share a common history of C_4 evolution whereas
328 independent events occurred in maize and setaria around ~45 million years apart (Gallaher *et al.* 2022).
329 Both *GLK* gene paralogs are predominantly expressed in mesophyll cells of C_3 rice leaves, with *GLK1*
330 transcripts accumulating at higher levels than *GLK2* (Hua *et al.* 2021), and gene function being
331 redundant (Wang *et al.* 2013). Notably, both transcripts can also be detected in bundle sheath cells, at
332 levels roughly equivalent to those of *GLK2* in mesophyll cells, suggesting that the few chloroplasts that
333 develop in rice bundle sheath cells are also regulated by the redundant action of both genes. By
334 contrast, an element of sub-functionalization is apparent in the C_4 species maize and setaria, with the
335 *GLK2* paralog expressed preferentially in bundle sheath cells. However, data from a number of different
336 C_3, C_4 and intermediate species (Washburn *et al.* 2021) (Figure S1) invoke a more complex scenario
337 because preferential expression of *GLK2* paralogs in bundle sheath cells is evident in species where
338 decarboxylation reactions occur in the bundle sheath chloroplasts (NADP-ME type) but not in species
339 where decarboxylation occurs in the mitochondria (NAD-ME type) or in the cytoplasm (PEP-CK type).
340 This suggests that preferential localization of GLK2 in bundle sheath cells is associated with specific
341 aspects of chloroplast function as opposed to general activation of photosynthesis.

342 The evolution of C_4 required activation of photosynthesis and a significant increase in chloroplast
343 volume in bundle sheath cells and as such, selection on regulatory regions of *GLK2* genes to enable
344 high levels of activity in that cell type may have been a critical step. Notably, there is more within-
345 species sequence variation in *GLK2* promoters of different rice and maize varieties than in the
346 corresponding *GLK1* promoters (Figure S6), providing more opportunity for change through drift or
347 selection. *GLK2* promoter sequences are also less conserved between species than corresponding
348 *GLK1* sequences (Figure S7). ATAC-seq data from leaves shows that the proximal promoter and 5'UTR
349 regions are the main sites of protein binding in *GLK* genes from maize, setaria and rice (Figure S4),
350 with noticeably more DNA binding motifs related to light responsive elements in the 5' UTR sequences
351 of *GLK2* genes than in equivalent regions of *GLK1* genes. In maize, an enrichment of BPC motifs is
352 seen in the *ZmG2* 5'UTR and the sequences are almost completely conserved across 1200 maize
353 varieties. This level of sequence conservation in the 5'UTR is in stark contrast to the rest of the promoter
354 region which is highly variable (Figure S6). Intriguingly, very few DNA binding motifs are found in either
355 the *SvGLK1* or *SvGLK2* promoters (Figure S4) and no variation relevant to the domestication process
356 was found when compared to *Setaria italica* (He *et al.* 2023). As in *ZmG2*, however, BPC motifs are
357 present in both the *SvGLK2* and *OsGLK2* 5'UTR sequences but they are not found in any of the *GLK1*
358 gene sequences. This observation suggests that the *GLK1* and *GLK2* 5' UTR sequences diverged in
359 the last common ancestor of maize, setaria and rice and thus prior to the evolution of C_4. Because BPC
360 binding proteins act to recruit polycomb complex proteins and thus mediate chromatin changes
361 (Theune *et al.* 2019), this divergence may have enabled novel post-transcriptional regulation of *GLK*2
362 orthologs during the evolution of C_4. Differences in the presence and position of upstream open reading
363 frames (uORFs) within the 5' UTR may have further facilitated de-repression of GLK2 translation (Zhang
364 *et al.* 2020) in that putative uORFs in ZmGLK1 and SvGLK1 are situated less than 10 base pairs after
365 the transcription start site (TSS) and thus if functional are predicted to be highly repressive, whereas
366 the uORF in ZmGLK2 is over 200 bp downstream of the TSS and no uORFs are predicted in SvGLK2
367 (Figure S7). In rice, predicted uORFs are positioned 100 bp downstream of the TSS in both genes and
368 may thus have no differential effect. Although requiring experimental verification, collectively these
369 observations support the hypothesis that altered *GLK2* gene regulation was associated with the
370 evolution of functional chloroplasts in bundle sheath cells of C_4 grasses.
371

372 **METHODS**

373 **Plant growth conditions**

374 Maize and *S. viridis* seeds were germinated in hydrated vermiculite or on paper, respectively, and
375 placed in an Aralab Fitoclima D1200 PLH incubator under 12 h light/12 h dark cycle at an intensity of
376 320 µmol photon m^{-2} s^{-1}, with temperatures ranging from 31 °C during the day to 22 °C at night and
377 humidity between 50-60 %. About 7-10 days after sowing (DAS), seedlings were transferred to pots

378 containing John Innes no. 2 soil mixture and irrigated with water to keep the soil moist. Plants were
379 kept under these conditions until sampled for experimental analysis.
380
381 For genetic crosses, maize plants were transferred to larger pots with the same soil mixture
382 supplemented with Osmocote slow-release granules. Plants were grown in a greenhouse at the
383 Department of Biology at the University of Oxford, UK with 16 h light/8 h dark cycle, daytime
384 temperature 28 °C and night-time temperature 20 °C, with supplemental light provided when natural
385 light levels were below 120 µmol photon m^{-2} s^{-1}.
386
387 **Construct design and assembly**
388 *ZmG2* and *ZmGLK1* orthologs were identified in *S. viridis* via phylogenetic reconstruction of the GARP
389 gene family (see Figure S2). Guide RNAs (gRNA) to drive CRISPR/Cas9 site-directed mutagenesis
390 were designed using the CRISPOR (Concordet and Haeussler 2018) online tool
391 (http://crispor.tefor.net/) to achieve high editing efficiency for the target genes *ZmGLK1*
392 (Zm00001d044785), *SvGLK1* (Sevir.4G123200) and *SvGLK2* (Sevir.5G003100), and to minimize off-
393 targets. For maize, the designed gRNAs were sent to the Wisconsin Crop Innovation Centre (University
394 of Wisconsin) to be cloned and transformed using their standard pipeline for maize transformation. For
395 *S. viridis*, the gRNA sequences were synthesized as oligonucleotides with appropriate flanking sites for
396 Golden Gate cloning (Table S1; JLF_Ox43 and JLF_Ox44 for *SvGLK1*; JLF_Ox51 and JLF_Ox52 for
397 *SvGLK2*). Prior to cloning, the oligonucleotides were annealed and phosphorylated by incubating 10
398 µM of each with 1x Ligase Buffer (New England Biolabs) and 10 U of T4 Polynucleotide Kinase (New
399 England Biolabs) in a final reaction volume of 15 µl at 37°C for 1.5 hours. After that period, 2 µl of 0.5
400 M NaCl was added, and the reaction was placed in a water bath heated to 95 °C for 2 min. The reaction
401 was cooled to room temperature before being diluted 100-fold and cloned downstream of the *TaU6*
402 promoter (module pJG310) using the Golden Gate one-step one-pot protocol (Engler *et al.* 2009) with
403 an appropriate restriction enzyme (*Esp3I*). This module was then combined with another level 1 module
404 containing *TaCas9* downstream of the maize ubiquitin promoter (pJG471) into the final acceptor and
405 binary vector pTC278. Both modules and the vector were provided by Dr Dan Voytas (U. of Minnesota).
406
407 To overexpress *ZmGLK1* and *ZmG2*, the coding sequences of the maize genes were domesticated to
408 avoid recognition sites for the type II restriction enzymes used in Golden Gate assembly: *BsaI*, *BpiI*,
409 *Esp3I* and *DraIII*. The final sequence was synthesized as a level 0 module with the appropriate SC
410 flanking sites (Engler and Marillonnet 2013). *ZmGLK1* and *ZmG2* level 0 modules (EC17054 and
411 EC17056) were cloned downstream of the maize ubiquitin promoter ($ZmUBI_{pro}$, EC15455) and
412 upstream of a *nos* terminator (tNOS, EC41421) in level 1 backbone position forward 2 using the Golden
413 Gate one-step one-pot protocol (Engler *et al.* 2009). The previously described *hygromycin*
414 *phosphotransferase* (HygR) coding sequence cloned downstream of the rice actin promoter ($OsACT_{pro}$)

415 was used as the selectable marker (Vlad *et al.* 2019). Level 1 modules were assembled into the binary
416 vector pAGM4723 to obtain the constructs EC17808 and EC17810 (hereafter referred to as 17808 and
417 17810, respectively) depicted in Figure S5B and S5C.
418
419 **Plant transformation**
420 *S. viridis* accessions A10 and ME034V were used to obtain CRISPR/Cas9-driven mutations for *SvGLK1*
421 and *SvGLK2*, respectively. *Agrobacterium tumefaciens* strain AGL1 harbouring the construct of interest
422 was co-cultivated with callus generated from seeds of *S. viridis*. Callus induction, transformation and
423 seedling regeneration were performed as described by VanEck and collaborators (Van Eck and
424 Startwood 2015). Complementation lines were obtained by transforming constructs 17808 and 17810
425 into callus generated from a population of seeds that segregated for the gene-edited *Svglk2* allele but
426 no longer segregated the original t-DNA which had been recombined out. The presence of t-DNA in
427 regenerated seedlings was confirmed via duplex PCR (Table S1; JLF_Ox12 and JLF_Ox13 amplify the
428 HygR gene; JLF_Ox14 and JLF_Ox15 amplify an endogenous gene used as a DNA quality control).
429 The occurrence of gene editing by CRISPR/Cas9 was verified by PCR amplification of the region
430 flanking the expected mutation site (Table S1; JLF_Ox45 and JLF_Ox46 for *SvGLK1*; JLF_Ox53 and
431 JLF_Ox54 for *SvGLK2*) followed by endonuclease digestion with *MmeI* or *Tsp45I* for *SvGLK1* and
432 *SvGLK2*, respectively, following the manufacturers' guidelines.
433
434 *Z. mays* Hi-II was transformed by the Wisconsin Crop Innovation Centre (University of Wisconsin)
435 according to their transformation pipeline. The presence of the t-DNA was assessed in T1 lines by PCR
436 (Table S1; JLF_Ox178 and JLF_Ox179) and *Zmglk1* mutants were identified by PCR amplification of
437 the region flanking the expected mutation site (Table S1; JLF_Ox39 and JLF_Ox175) followed by
438 endonuclease digestion with *HinfI* according to the manufacturer's instructions.
439
440 For all lines, mutated sequences were verified by Sanger sequencing (Source BioScience, UK).
441 Experimental analyses were performed on homozygous plants in the T2 or subsequent generations.
442
443 **Confocal imaging**
444 The mid-portion of fully expanded leaf 7 from 21 DAS *S. viridis* plants was fixed in 4 %
445 paraformaldehyde/0.2 % glutaraldehyde in 25 mM phosphate buffer pH 7.2. After fixation, leaves were
446 stored in 0.2 M Na-EDTA pH 9.0. Free-hand cross sections were imaged using a Leica SP9 confocal
447 microscope with an Argon laser set to excite chlorophyll at 663 nm and to emit from 650-800 nm.
448
449 **Ultrastructure analysis via transmission electron microscopy**
450 Sample fixation and preparation were performed by the Electron Microscopy Facility team at the Sir
451 Willian Dunn School of Pathology (University of Oxford). Leaf discs of 2 mm diameter were collected

452 from leaf 5 of 40 DAS maize or 25 DAS *S. viridis* plants and placed into 3 % glutaraldehyde in 0.06 M
453 Sorenson's phosphate buffer pH 7.2. Samples were transferred to a Leica AMW microwave tissue
454 processor for fixation, osmification (with 1 % OsO_4), dehydration with an acetone series, and resin
455 infiltration (LRW Hard). The samples were then removed from the microwave and transferred to fresh
456 resin for four days, renewing the resin twice a day. The embedded samples were transferred into gelatin
457 capsules and filled with resin for polymerization at 60 °C for about 65 h. Sections of 150 nm were cut
458 using a Leica UC7 ultramicrotome, collected onto formvar-coated copper slot grids and post-stained
459 with lead citrate. Images were acquired on a Thermo Fischer Tecnai T12 TEM with a Gatan OneView
460 CMOS camera.
461
462 Thylakoid occupancy of chloroplasts was quantified using ImageJ. The thylakoid area was measured
463 by adjusting the image threshold to create a binary mask on which only the thylakoid membrane was
464 apparent. This area value was divided by the chloroplast planar area to give the thylakoid occupancy
465 level.
466
467 **Chlorophyll extraction and quantification**
468 Leaf discs were harvested from fully expanded leaf 5 at 30 DAS for maize and 21 DAS for *S. viridis*,
469 and immediately frozen in liquid nitrogen. For total chlorophyll extraction, finely powdered tissue was
470 resuspended in 80 % acetone buffered in 100 mM Tris-HCl pH 8.0 (Chazaux *et al.* 2022). Following a
471 centrifugation step, the absorbance of each sample was read in technical triplicates in a FLUOstar
472 Omega spectrometer (BMG Labtech) at 646 and 663 nm. Background absorbance was measured at
473 750 nm. Absorbance values recorded at 750 nm were subtracted from values recorded at 646 and 663
474 nm for normalization purposes. Technical replicates were averaged for each sample. Concentration of
475 chlorophylls a + b were calculated in micrograms (µg) per leaf area using the following the equation:
476 $(19.54 \times Abs_{646} + 8.29 \times Abs_{663}) \times$ dilution factor (Porra *et al.* 1989).
477
478 **Fluorescence quantification of PSII**
479 Photosystem II (PSII) fluorescence was measured using a LiCOR Li-6800 portable photosynthesis
480 system. LiCOR chamber conditions were set to a flow intensity of 500 umol s^{-1}, delta pressure of 0.2
481 kPa, relative air humidity of 50 %, reference concentration of CO_2 at 400 umol mol^{-1} and temperature
482 exchange at 25 °C. For quantification of PSII operating efficiency (PhiPSII), plants were adapted to light
483 under the conditions described above for at least 1.5 h and actinic light intensity on the LiCOR chamber
484 was set to 1000 µmol photon m^{-2} s^{-1}. After clamping, each leaf was adapted for a few minutes until
485 stabilization was reached before logging the fluorescence value. Quantification of maximum quantum
486 efficiency of PSII photochemistry (Fv/Fm) was performed on plants that were dark-adapted for at least
487 16 h. LiCOR actinic lights were off. In both cases, the fluorescence action log was set to FoFm (dark),

488 FsFm' (light), rectangular flash type with red target at 8000 µmol photon m^{-2} s^{-1}, duration of 1000 ms,
489 output rate of 100 Hz and margin of 5 points.
490
491 **Gene expression analysis via qPCR**
492 The mid-portion of fully expanded leaf 5 from at least three biological replicates of 30 DAS maize or 21
493 DAS *S. viridis* were harvested and immediately frozen in liquid nitrogen. Tissue was ground prior to
494 total RNA extraction using a RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's
495 instructions. To avoid any DNA contamination, samples were treated with TURBO DNA-free DNase
496 (Thermo Fischer Scientific) according to the manufacturer's guidelines. For first strand cDNA synthesis,
497 100 ng of total RNA was added to 1 µl of Maxima Enzyme Mix (Thermo Fischer Scientific) and 1x
498 Reaction Mix to a final volume of 10 µL. The mixture was incubated at 25 °C for 10 min, followed by 50
499 °C for 15 min. Enzyme inactivation was carried out at 85 °C for 5 min.
500
501 Amplification reactions were performed in a StepOnePlus Real-Time PCR System (Thermo Fischer
502 Scientific) using SYBRGreen to monitor dsDNA synthesis. The reaction mixtures contained 5 µL of
503 diluted cDNA (1:25), 0.2 µM of each primer and 10 µL of SYBR™ Green PCR Master Mix (Thermo
504 Fischer Scientific) in a total volume of 10 µL. The reaction cycles began with a 5 min denaturation step
505 at 94 °C, followed by 40 amplification cycles of 15 s at 94 °C, 10 s at 60 °C, 15 s at 72 °C. After each
506 cycle, the fluorescence was measured at 60 °C for 35 s. The melting curve was produced by a cycle of
507 95 °C for 15 s, 60 °C for 1 min, 95 °C for 30 s and finally 60 °C for 15 s. Every qPCR reaction was
508 repeated three times to make technical replicates.
509
510 Transcript levels for multiple target genes were measured with primers designed to specifically amplify
511 the gene of interest (Table S1). Two reference genes were used to normalize gene expression for
512 maize (Lin *et al.* 2014) (Table S1; ZmCYP_F and ZmCYP_R, ZmEF1α_F and ZmEF1α_R) and *S. viridis*
513 (Lambret-Frotté *et al.* 2015) (Table S1; JLF_Ox26 and JLF_Ox27, JLF_Ox28 and JLF_Ox29). Exported
514 *Rn* values were used to calculate amplification efficiency for each primer pair and the Cq values for
515 each qPCR reaction using the '*qpcR*' package from R (Ritz and Spiess 2008). Normalized relative
516 expression of the target genes was calculated using the R package '*EasyqpcR'*
517 (https://www.bioconductor.org/packages//2.12/bioc/html/EasyqpcR.html), according to the ΔΔCq
518 model (Pfaffl 2001) based on the previously calculated Cq and amplification efficiencies.
519
520 **Statistical analysis**
521 Datasets were tested for Normal distributions using the Shapiro-Wilk test (Shapiro and Wilk 1965). For
522 thylakoid coverage, chlorophyll concentration and PSII fluorescence, variance was calculated using a
523 one-way ANOVA followed by a pairwise comparison analysis with Tukey's HSD test (Tukey 1949). For
524 relative gene expression, statistical significance was calculated based on a Student's t-test, comparing

525 the mean variance of each mutant line with their respective controls. All statistical analyses were
526 performed using *scipy* v.1.5.0 library and plots were prepared using *matplotlib* v3.5.1 library; both from
527 Python v.3.7.6, using Visual Studio Code as a code editor.
528
529 **LIST OF SUPPLEMENTARY MATERIAL**
530 Table S1. Oligonucleotides used for cloning, genotyping and qPCR.
531 Table S2. Data and statistical analyses supporting Figures 1-6.
532 Figure S1. *GLK* transcript accumulation in bundle sheath and mesophyll cells.
533 Figure S2. Phenotypic characterization of three independent *Zmglk1* mutant lines.
534 Figure S3. Comparison of *GLK* sequences.
535 Figure S4. DNA binding motifs in regulatory regions of *GLK* genes.
536 Figure S5. Phenotypic characterization of *Svglk2* mutant lines complemented with *ZmGLK1* or *ZmG2*.
537 Figure S6. Sequence variability in regulatory regions of *GLK* genes.
538 Figure S7. Alignment of *GLK* gene promoter and 5' UTR regions.
539
540 **AUTHOR CONTRIBUTIONS**
541 J.L-F and J.A.L conceived and designed the experiments; G.S carried out setaria transformation, plant
542 cultivation and genotyping; J.L-F carried out all of the remaining experiments and analysed the data.
543 J.L-F and J.A.L wrote the manuscript.
544
545 **ACKNOWLEDGEMENTS**
546 The authors thank Dan Voytas for making available the CRISPR/Cas9 cloning vectors for gene editing
547 in setaria; Raman Dhaliwal, Charlotte Melia and Errin Jonson for processing and imaging the TEM
548 samples; Ray Collier, Mike Pettersen and the Wisconsin Crop Innovation Centre team for maize
549 transformation; John Baker for plant photography; Roxaana Clayton, Julie Bull and Lizzie Jamison for
550 technical support; Emma Raven for assisting with initial characterization of mutant lines; Jacob
551 Washburn for providing TPM data for Figure S1B; Sophie Johnson, Chiara Perico, Daniela Vlad,
552 Sovanna Tan, Thomas Hughes and Maricris Zaidem for discussion throughout the experimental work
553 and during manuscript preparation.
554
555 **COMPETING INTERESTS**
556 No competing interests are declared.
557
558 **FUNDING**
559 This research was funded by a BBSRC sLoLa grant (BB/P003117/1), a Newton International Fellowship
560 from The Royal Society to J.L-F (NF171598, 2018 – 2020) and by the Bill and Melinda Gates Foundation
561 C_4 Rice grant awarded to the University of Oxford (2015-2019; OPP1129902).

562 **REFERENCES**

563 Bravo-Garcia, A., Yasumura, Y., and Langdale, J.A. (2009) Specialization of the *Golden2-like* regulatory
564 pathway during land plant evolution. *New Phytologist* 183, 133–141.

565 Chang, Y.-M., Liu, W.-Y., Shih, A.C.-C., Shen, M.-N., Lu, C.-H., Lu, M.-Y.J., Yang, H.-W., Wang, T.-Y.,
566 Chen, S.C.-C., Chen, S.M., Li, W.-H., and Ku, M.S.B. (2012) Characterizing Regulatory and
567 Functional Differentiation between Maize Mesophyll and Bundle Sheath Cells by Transcriptomic
568 Analysis. *Plant Physiology* 160, 165–177.

569 Chazaux, M., Schiphorst, C., Lazzari, G., and Caffarri, S. (2022) Precise estimation of chlorophyll a, b
570 and carotenoid content by deconvolution of the absorption spectrum and new simultaneous
571 equations for Chl determination. *Plant Journal* 109, 1630–1648.

572 Chiang, Y.-H., Zubo, Y.O., Tapken, W., Kim, H.J., Lavanway, A.M., Howard, L., Pilon, M., Kieber, J.J.,
573 and Schaller, G.E. (2012). Functional Characterization of the GATA Transcription Factors GNC
574 and CGA1 Reveals Their Key Role in Chloroplast Development, Growth, and Division in
575 Arabidopsis. *Plant Physiology* 160, 332–348.

576 Christin, P.-A. and Osborne, C.P. (2013) The recurrent assembly of C_4 photosynthesis, an evolutionary
577 tale. *Photosynthesis Research* 117, 163–75.

578 Concordet, J.-P. and Haeussler, M. (2018) CRISPOR: intuitive guide selection for CRISPR/Cas9
579 genome editing experiments and screens. *Nucleic Acids Research* 46, W242–W245.

580 Cribb, L., Hall, L., and Langdale, J.A. (2001) Four Mutant Alleles Elucidate the Role of the G2 Protein in
581 the Development of C_4 and C_3 Photosynthesizing Maize Tissues. *Genetics* 159 787–797.

582 Van Eck, J. and Startwood, K. (2015) *Setaria viridis*. *In*: K. Wang (Ed.). *Agrobacterium Protocols*.
583 Springer, 417.

584 Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009) Golden gate shuffling: a one-pot DNA
585 shuffling method based on type IIs restriction enzymes. *PloS ONE* 4, e5553.

586 Engler, C. and Marillonnet, S. (2013) Combinatorial DNA assembly using golden gate cloning. *Methods*
587 *in Molecular Biology* 1073, 141–156.

588 Fitter, D.W., Martin, D.J., Copley, M.J., Scotland, R.W., and Langdale, J.A. (2002) *GLK* gene pairs
589 regulate chloroplast development in diverse plant species. *Plant Journal* 31, 713–727.

590 Gallaher, T.J., Peterson, P.M., Soreng, R.J., Zuloaga, F.O., Li, D.-Z., Clark, L.G., Tyrrell, C.D., Welker,
591 C.A.D., Kellogg, E.A., and Teisher, J.K. (2022) Grasses through space and time: An overview of
592 the biogeographical and macroevolutionary history of Poaceae. *Journal of Systematics and*
593 *Evolution* 60, 522–569.

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

596 He, Q., Tang, S., Zhi, H., Chen, J., Zhang, J., Liang, H., Alam, O., Li, H., Zhang, H., Xing, L., Li, X., Zhang,
597 W., Wang, H., Shi, J., Du, H., Wu, H., Wang, L., Yang, P., Xing, L., Yan, H., Song, Z., Liu, J., Wang,
598 H., Tian, X., Qiao, Z., Feng, G., Guo, R., Zhu, W., Ren, Y., Hao, H., Li, M., Zhang, A., Guo, E., Yan,

599 F., Li, Q., Liu, Y., Tian, B., Zhao, X., Jia, R., Feng, B., Zhang, J., Wei, J., Lai, J., Jia, G., Purugganan,
600 M., and Diao, X. (2023) A graph-based genome and pan-genome variation of the model plant
601 *Setaria*. *Nature Genetics* 55, 1232-1242
602 Hua, L., Stevenson, S.R., Reyna-Llorens, I., Xiong, H., Kopriva, S., and Hibberd, J.M. (2021) The bundle
603 sheath of rice is conditioned to play an active role in water transport as well as sulfur assimilation
604 and jasmonic acid synthesis. *Plant Journal* 107, 268–286.
605 Hudson, D., Guevara, D.R., Hand, A.J., Xu, Z., Hao, L., Chen, X., Zhu, T., Bi, Y.-M., and Rothstein, S.J.
606 (2013) Rice Cytokinin GATA Transcription Factor1 Regulates Chloroplast Development and Plant
607 Architecture. *Plant Physiology* 162, 132–144.
608 John, C.R., Smith-Unna, R.D., Woodfield, H., Covshoff, S., and Hibberd, J.M. (2014) Evolutionary
609 convergence of cell-specific gene expression in independent lineages of C_4 grasses. *Plant*
610 *Physiology* 165, 62–75.
611 Katoh, K., Rozewicki, J., and Yamada, K.D. (2019) MAFFT online service: multiple sequence alignment,
612 interactive sequence choice and visualization. *Briefings in Bioinformatics* 20, 1160–1166.
613 Kenrick, P. and Crane, P.R. (1997) The origin and early evolution of plants on land. *Nature* 389, 33–39.
614 Kobayashi, K., Baba, S., Obayashi, T., Sato, M., Toyooka, K., Keränen, M., Aro, E.M., Fukaki, H., Ohta,
615 H., Sugimoto, K., and Masuda, T. (2012) Regulation of root greening by light and auxin/cytokinin
616 signaling in Arabidopsis. *Plant Cell*, 24, 1081–1095.
617 Lambret-Frotté, J., Almeida, L.C.S. De, Moura, S.M. De, and Souza, F.L.F. (2015) Validating Internal
618 Control Genes for the Accurate Normalization of qPCR Expression Analysis of the Novel Model
619 Plant *Setaria viridis*. *PloS One* 10, e0135006.
620 Langdale, J.A. (2011) C_4 cycles: past, present, and future research on C_4 photosynthesis. *The Plant*
621 *Cell* 23, 3879–92.
622 Langdale, J.A. and Kidner, C.A. (1994) *bundle sheath defective*, a mutation that disrupts cellular
623 differentiation in maize leaves. *Development* 120, 673–681.
624 Li, P. and Brutnell, T.P. (2011) *Setaria viridis* and *Setaria italica*, model genetic systems for the Panicoid
625 grasses. *Journal of Experimental Botany* 62, 3031–7.
626 Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S.L., Kebrom, T.H., Provart, N., Patel, R., Myers,
627 C.R., Reidel, E.J., Turgeon, R., Liu, P., Sun, Q., Nelson, T., and Brutnell, T.P. (2010) The
628 developmental dynamics of the maize leaf transcriptome. *Nature Genetics* 42, 1060–7.
629 Li, X., Wang, P., Li, J., Wei, S., Yan, Y., Yang, J., Zhao, M., Langdale, J.A., and Zhou, W. (2020) Maize
630 *GOLDEN2-LIKE* genes enhance biomass and grain yields in rice by improving photosynthesis and
631 reducing photoinhibition. *Communications Biology* 3, 151.
632 Lin, Y., Zhang, C., Lan, H., Gao, S., Liu, H., Liu, J., Cao, M., Pan, G., Rong, T., and Zhang, S. (2014)
633 Validation of potential reference genes for qPCR in maize across abiotic stresses, hormone
634 treatments, and tissue types. *PloS One* 9, e95445.

635 Marshall, D.M., Muhaidat, R., Brown, N.J., Liu, Z., Stanley, S., Griffiths, H., Sage, R.F., and Hibberd, J.M.
636 (2007) Cleome, a genus closely related to Arabidopsis, contains species spanning a
637 developmental progression from C_3 to C_4 photosynthesis. *Plant Journal* 51, 886–896.
638 Murray, A., Mendieta, J.P., Vollmers, C., and Schmitz, R.J. (2022) Simple and accurate transcriptional
639 start site identification using Smar2C2 and examination of conserved promoter features. *The Plant*
640 *Journal* 112, 583–596.
641 Nguyen, C. V., Vrebalov, J.T., Gapper, N.E., Zheng, Y., Zhong, S., Fei, Z., and Giovannoni, J.J. (2014)
642 Tomato *GOLDEN2-LIKE* transcription factors reveal molecular gradients that function during fruit
643 development and ripening. *Plant Cell* 26, 585–601.
644 Nickelsen, J. and Rengstl, B. (2013) Photosystem II Assembly: From Cyanobacteria to Plants. *Annual*
645 *Review of Plant Biology* 64, 609–635.
646 Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic*
647 *Acids Research* 29, e45.
648 Porra, R.J., Thompson, W.A., and Kriedemann, P.E. (1989) Determination of accurate extinction
649 coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four
650 different solvents: verification of the concentration of chlorophyll standards by atomic absorption
651 spectroscopy. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 975, 384–394.
652 Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe,
653 O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman,
654 B.K., and Yu, G. (2000) Arabidopsis transcription factors: genome-wide comparative analysis
655 among eukaryotes. *Science* 290, 2105–2110.
656 Ritz, C. and Spiess, A.-N. (2008) qpcR: an R package for sigmoidal model selection in quantitative real-
657 time polymerase chain reaction analysis. *Bioinformatics* 24, 1549–1551.
658 Rossini, L., Cribb, L., Martin, D.J., and Langdale, J.A. (2001) The maize *Golden2* gene defines a novel
659 class of transcriptional regulators in plants. *Plant Cell* 13, 1231–1244.
660 Safi, A., Medici, A., Szponarski, W., Ruffel, S., Lacombe, B., and Krouk, G. (2017) The world according
661 to GARP transcription factors. *Current Opinion in Plant Biology* 39, 159–167.
662 Sagan, L. (1967) On the origin of mitosing cells. *Journal of Theoretical Biology* 14, 225-IN6.
663 Sage, R.F., Christin, P.-A., and Edwards, E.J. (2011) The C_4 plant lineages of planet Earth. *Journal of*
664 *Experimental Botany* 62, 3155–69.
665 Shapiro, S.S. and Wilk, M.B. (1965) An analysis of variance test for normality (complete samples).
666 *Biometrika* 52, 591–611.
667 Tausta, L.S., Li, P., Si, Y., Gandotra, N., Liu, P., Sun, Q., Brutnell, T.P., and Nelson, T. (2014)
668 Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset
669 of C_4-related processes. *Journal of Experimental Botany* 65, 3543–3555.

670 Theune, M.L., Bloss, U., Brand, L.H., Ladwig, F., and Wanke, D. (2019) Phylogenetic analyses and
671 GAGA-motif binding studies of BBR/BPC proteins lend to clues in GAGA-MOTIF recognition and
672 a regulatory role in brassinosteroid signaling. *Frontiers in Plant Science* 10, e00466.
673 Tu, X., Ren, S., Shen, W., Li, J., Li, Y., Li, C., Li, Y., Zong, Z., Xie, W., Grierson, D., Fei, Z., Giovannoni,
674 J., Li, P., and Zhong, S. (2022) Limited conservation in cross-species comparison of GLK
675 transcription factor binding suggested wide-spread cistrome divergence. *Nature Communications*
676 13, 7632.
677 Tukey, J.W. (1949) Comparing Individual Means in the Analysis of Variance. *Biometrics* 5, 99–114.
678 Vlad, D., Abu-Jamous, B., Wang, P., and Langdale, J.A. (2019) A modular steroid-inducible gene
679 expression system for use in rice. *BMC Plant Biology* 19, 426.
680 Wang, L., Czedik-Eysenberg, A., Mertz, R. a, Si, Y., Tohge, T., Nunes-Nesi, A., Arrivault, S., Dedow, L.K.,
681 Bryant, D.W., Zhou, W., Xu, J., Weissmann, S., Studer, A., Li, P., Zhang, C., LaRue, T., Shao, Y.,
682 Ding, Z., Sun, Q., Patel, R. V, Turgeon, R., Zhu, X., Provart, N.J., Mockler, T.C., Fernie, A.R., Stitt,
683 M., Liu, P., and Brutnell, T.P. (2014) Comparative analyses of C_4 and C_3 photosynthesis in
684 developing leaves of maize and rice. *Nature Biotechnology*, 32, 1158-65.
685 Wang, P., Fouracre, J., Kelly, S., Karki, S., Gowik, U., Aubry, S., Shaw, M.K., Westhoff, P., Slamet-Loedin,
686 I.H., Quick, W.P., Hibberd, J.M., and Langdale, J.A. (2013) Evolution of *GOLDEN2-LIKE* gene
687 function in C_3 and C_4 plants. *Planta* 237, 481–495.
688 Washburn, J.D., Strable, J., Dickinson, P., Kothapalli, S.S., Brose, J.M., Covshoff, S., Conant, G.C.,
689 Hibberd, J.M., and Pires, J.C. (2021) Distinct C_4 sub-types and C_3 bundle sheath isolation in the
690 Paniceae grasses. *Plant Direct* 5, 1–14.
691 Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009) Jalview Version 2—
692 a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191.
693 Waters, M.T. and Langdale, J.A. (2009) The making of a chloroplast. *EMBO J.* 28, 2861–2873.
694 Waters, M.T., Moylan, E.C., and Langdale, J.A. (2008) GLK transcription factors regulate chloroplast
695 development in a cell-autonomous manner. *Plant Journal* 56, 432–444.
696 Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J., and Langdale, J.A. (2009) GLK
697 Transcription Factors Coordinate Expression of the Photosynthetic Apparatus in Arabidopsis. *The*
698 *Plant Cell* 21, 1109–1128.
699 Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S.,
700 Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., Ruhfel, B.R., Wafula, E., Der, J.P., Graham, S.W.,
701 Mathews, S., Melkonian, M., Soltis, D.E., Soltis, P.S., Miles, N.W., Rothfels, C.J., Pokorny, L., Shaw,
702 A.J., DeGironimo, L., Stevenson, D.W., Surek, B., Villarreal, J.C., Roure, B., Philippe, H.,
703 dePamphilis, C.W., Chen, T., Deyholos, M.K., Baucom, R.S., Kutchan, T.M., Augustin, M.M., Wang,
704 J., Zhang, Y., Tian, Z., Yan, Z., Wu, X., Sun, X., Wong, G.K.-S., and Leebens-Mack, J. (2014)
705 Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of*
706 *the National Academy of Sciences* 111, E4859–E4868.

707 Woo, K.C., Anderson, J.M., Boardman, N.K., Downton, W.J.S., Osmond, C.B., and Thorne, S.W. (1970)
708 Deficient Photosystem II in Agranal Bundle Sheath Chloroplasts of C_4 Plants. *Proceedings of the*
709 *National Academy of Sciences* 67, 18–25.
710 Yasumura, Y., Moylan, E.C., and Langdale, J.A. (2005) A Conserved Transcription Factor Mediates
711 Nuclear Control of Organelle Biogenesis in Anciently Diverged Land Plants. *Plant Cell* 17, 1894–
712 1907.
713 Zhang, T., Wu, A., Yue, Y., and Zhao, Y. (2020) uORFs: Important cis-regulatory elements in plants.
714 *International Journal of Molecular Sciences* 21, 1-4.
715 Zubo, Y.O., Blakley, I.C., Franco-Zorrilla, J.M., Yamburenko, M. V, Solano, R., Kieber, J.J., Loraine, A.E.,
716 and Schaller, G.E. (2018) Coordination of Chloroplast Development through the Action of the
717 GNC and GLK Transcription Factor Families. *Plant Physiology* 178, 130–147.
718
719

720 **FIGURE LEGENDS**

721 **Figure 1. Chloroplast development is not perturbed in *Zmglk1* mutants of maize**. **A)** Gene model
722 of *ZmGLK1* showing predicted exons in black boxes and the target site of CRISPR/Cas9 mediated
723 mutagenesis in yellow and red below. The HLH associated with MYB or MYB-related proteins is
724 depicted by diagonal lines; the AREAEAA conserved site is highlighted with horizontal lines and the last
725 exon that contains the GCT-box domain is depicted by vertical lines. The table shows the sequence of
726 the wild-type (WT) and mutated alleles with the predicted protein length. The gRNA sequence is
727 highlighted in yellow and the PAM sequence in red. **B-C)** Whole plant phenotype of wild-type (B) and
728 a representative *Zmglk1* mutant 30 days after sowing (C). Scale bars = 10 cm. **D-G)** Electron
729 micrographs of chloroplast ultrastructure in bundle sheath (BSC; D-E) and mesophyll (MC; F-G) cells
730 of WT (D, F) and a representative *Zmglk1* mutant (E, G). Scale bars = 2 µm. **H-L)** Measurements of
731 thylakoid occupancy in mesophyll cells (H), thylakoid occupancy in bundle sheath cells (I), total
732 chlorophyll (J), PSII operating efficiency (K) and maximum quantum efficiency of PSII photochemistry
733 (L). The WT segregant is depicted in blue and the four independent *Zmglk1* lines are depicted in green.
734 Each dot represents a biological replicate, the square is the average value for that line and the bars are
735 the associated standard deviations. A minimum of 5 biological replicates were evaluated for thylakoid
736 occupancy and chlorophyll quantification, and 3 biological replicates were assessed for PSII
737 fluorescence. Different letters indicate statistically different groups (p value ≤0.05, one-way ANOVA
738 and Tukey's HSD; see Table S2 for raw data).

739

740 **Figure 2. Chloroplast development is not perturbed in *Svglk1* mutants of setaria. A)** Gene model
741 of *SvGLK1* showing predicted exons in black boxes and the target site of CRISPR/Cas9 mediated
742 mutagenesis in yellow and red below. The HLH associated with MYB or MYB-related proteins is
743 depicted by diagonal lines; the AREAEAA conserved site is highlighted with horizontal lines and the last
744 exon that contains the GCT-box domain is depicted by vertical lines. The table shows the sequence of
745 the wild-type (WT) and mutated alleles with the predicted protein length. The gRNA sequence is
746 highlighted in yellow and the PAM sequence in red. **B-D)** Whole plant phenotype of wild-type (B) and
747 *Svglk1* mutants 20 days after sowing (C-D). Scale bars = 5 cm. **E-J)** Electron micrographs of chloroplast
748 ultrastructure in bundle sheath (BSC; E-G) and mesophyll (MC; H-J) cells of WT (E, H) and *Svglk1*
749 mutants (F, G, I, J). Scale bars = 2 µm. **K-O)** Measurements of thylakoid occupancy in mesophyll cells
750 (K), thylakoid occupancy in bundle sheath cells (L), total chlorophyll (M), PSII operating efficiency (N)
751 and maximum quantum efficiency of PSII photochemistry (O). The WT segregant is depicted in blue
752 and the two independent *Svglk1* lines are depicted in green. Each dot represents a biological replicate,
753 the square is the average value for that line and the bars are the associated standard deviations. Twelve
754 chloroplasts were assessed for thylakoid occupancy quantification, and 5 biological replicates were
755 assessed for chlorophyll quantification and PSII fluorescence for each sample. Different letters indicate

756 statistically different groups (p value ≤0.05, one-way ANOVA and Tukey's HSD; see Table S2 for raw
757 data).
758
759 **Figure 3. Chloroplast development is perturbed in both mesophyll and bundle sheath cells of**
760 ***Svglk2* mutants of setaria. A)** Gene model of *SvGLK2* showing predicted exons in black boxes and
761 the target site of CRISPR/Cas9 mediated mutagenesis in yellow and red below. The HLH associated
762 with MYB or MYB-related proteins is depicted by diagonal lines; the AREAEAA conserved site is
763 highlighted with horizontal lines and the last exon that contains the GCT-box domain is depicted by
764 vertical lines. The table shows the sequence of the wild-type (WT) and mutated alleles with the
765 predicted protein length. The gRNA sequence is highlighted in yellow and the PAM sequence in red.
766 **B)** Whole plant phenotype of wild-type, a heterozygous and a homozygous *Svglk2* mutant 20 days after
767 sowing. Scale bar = 5 cm. **C-D)** Confocal images of leaf cross sections showing overall chloroplast
768 structure in WT (C) and the *Svglk2* mutant (D). Bundle sheath cells are outlined in white. Scale bars =
769 50 µm. **E-H)** Electron micrographs of chloroplast ultrastructure in bundle sheath (BSC; E-F) and
770 mesophyll (MC; G-H) cells of WT (E, G) and the *Svglk2* mutant (F, H). Yellow arrows point to fragmented
771 membranes. Scale bars = 2 µm. **I-K)** Measurements of thylakoid occupancy in bundle sheath cells (I),
772 thylakoid occupancy in mesophyll cells (J) and total chlorophyll (K). The WT segregant is depicted in
773 purple and the *Svglk2* line is depicted in pink. Each dot represents a biological replicate, the square is
774 the average value for that line and the bars are the associated standard deviations. Twelve chloroplasts
775 were assessed for thylakoid occupancy quantification, and 5 biological replicates were assessed for
776 chlorophyll quantification and PSII fluorescence for each sample. Different letters indicate statistically
777 different groups (p value ≤0.05, one-way ANOVA and Tukey's HSD; see Table S2 for raw data).
778
779 **Figure 4. *GLK1* and *GLK2* transcript levels are reduced in *Svglk2* and *Zmg2* mutants of setaria**
780 **and maize. A-H)** Quantitative analysis of gene expression via qPCR in segregating wild-type (WT)
781 siblings and *glk* mutants of maize (A-D) and setaria (E-F). Relative normalized expression of *ZmGLK1*
782 (A), *ZmG2* (B), *ZmGNC* (C), *ZmCGA1* (D), *SvGLK1* (E) and *SvGLK2* (F) was calculated based on the
783 expression of two reference genes for each sample. Each dot represents a biological replicate, the
784 square is the mean value and the bars are the associated standard deviations. A minimum of 3 and 5
785 biological replicates were evaluated for maize and setaria samples, respectively. Statistical significance
786 was calculated using a Student's t-test and pairwise comparison with the respective WT sample. For
787 each species comparison, different letters indicate statistically different groups (p value ≤0.05; see
788 Table S2 for raw data).
789
790 **Figure 5. *ZmGLK1* and *ZmG2* complement the *Svglk2* phenotype and activate transcription of**
791 **the setaria homologs**. **A-B)** Whole plant phenotype of *Svglk2* mutant overexpressing *ZmGLK1* (A) or
792 *ZmG2* (B) 30 days after sowing. Scale bars = 5 cm. **C-F)** Quantitative analysis of gene expression via

793 qPCR of *ZmGLK1* (C), *ZmG2* (D), *SvGLK1* (E) and *SvGLK2* (F). Relative gene expression was calculated
794 based on the expression of two reference genes for each sample. Each dot represents one of the 5
795 biological replicates evaluated, the square is the mean value and the bars are the associated standard
796 deviations. Statistical significance was calculated using a Student's t-test and pairwise comparison with
797 the respective wild-type (WT) sample. Different letters indicate statistically different groups (p value
798 ≤0.05; see Table S2 for raw data).
799
800 **Figure 6. *ZmGLK1* and *ZmG2* are sufficient to induce dimorphic chloroplast development in both**
801 **mesophyll and bundle sheath cells of setaria**. **A-C)** Confocal images of leaf cross sections showing
802 overall chloroplast morphology in the *Svglk2* mutant (A, same as Figure 3D), and representative lines
803 expressing *ZmGLK1* (B) or *ZmG2* (C). Bundle sheath cells are outlined in white. Yellow arrows indicate
804 ectopic chloroplasts developing in vascular cells. Scale bars = 50 µm. **D-I)** Electron micrographs of
805 chloroplast ultrastructure in bundle sheath (BSC; D-F) and mesophyll (MC; G-I) cells of the *Svglk2*
806 mutant (D-G; same as Figure 3G, H) and one representative line overexpressing *ZmGLK1* (E-H) or
807 ZmG2 (F-I). Scale bars = 2 µm. **J-L)** Measurements of thylakoid occupancy in mesophyll cells (J),
808 thylakoid occupancy in bundle sheath cells (K) and total chlorophyll (L). The wild-type (WT) segregant
809 is depicted in purple, the *Svglk2* line in pink, the three independent lines overexpressing *ZmGLK1* in
810 orange and those overexpressing *ZmG2* in yellow. Each dot represents a biological replicate, the
811 square is the average value for that line and the bars are the associated standard deviations. A minimum
812 of 10 chloroplasts were evaluated for thylakoid occupancy quantification and 5-10 biological replicates
813 were used for chlorophyll quantification. Different letters indicate statistically different groups in (p value
814 ≤0.05, one-way ANOVA and Tukey's HSD; see Table S2 for raw data).
815
816

817 **SUPPLEMENTARY FIGURE LEGENDS**

818 **Figure S1. *GLK* transcript accumulation in bundle sheath and mesophyll cells. A)** Gene expression
819 in reads per kilobase million (RPKM) in mesophyll (MC) and bundle sheath cells (BSC) of setaria
820 (*SvPEPC*, *SvNADP-ME*, *SvGLK1*, *SvGLK2*) and maize (*ZmPEPC*, *ZmNADP-ME*, *ZmGLK1*, *ZmG2*).
821 Maize data were published by Li et al. (2010) and Chang et al. (2012), and setaria data were published
822 by John et al. (2014). *PEPC* and *NADP-ME* transcripts accumulate specifically in mesophyll and bundle
823 sheath cells respectively, and thus act as markers for cross contamination of RNA samples between
824 cell-types. As such, mesophyll contamination of bundle sheath transcriptomes was 22% (Li et al), 4.5%
825 (Chang et al) and 3.5% (John et al), whereas bundle sheath contamination of mesophyll transcriptomes
826 was 10.7% (Li et al), 7% (Chang et al) and 6.6% (John et al). Disregarding the Li et al (2010) data
827 because of the high level of cross contamination, *GLK1* transcript levels were higher in mesophyll than
828 bundle sheath cells by 8-fold in maize and 21-fold in setaria, whereas *GLK2* transcript levels were
829 higher in bundle sheath than mesophyll cells by 4.8-fold in maize and 3.2-fold in setaria. **B)** Log2 fold
830 change between bundle sheath and mesophyll transcripts for *S. viridis* (John et al. 2014), *Z. mays* (Li
831 et al. 2010, Chang et al. 2012), and three different species of Paniceae grasses (*Urochloa fusca*,
832 *Panicum hallii* and *Digitaria californica*) (Washburn et al. 2022). All species carry out C_4 photosynthesis,
833 but each uses a different decarboxylation pathway in the bundle sheath cells. *PEPC* was used as a
834 mesophyll cell marker in each case and genes encoding the relevant decarboxylation enzyme (i.e.
835 *NADP-ME, PCK* and *NAD-ME*) were used as bundle sheath cell markers.

836

837 **Figure S2. Phenotypic characterization of three independent *Zmglk1* mutant lines. A-D)** Whole
838 plant phenotype 30 days after sowing. Scale bars = 10 cm. **E-H)** Transmission electron microscopy
839 (TEM) images showing mesophyll chloroplast ultrastructure. Scale = 2 µm, 2 µm, 5 µm and 2 µm,
840 respectively. **I-L)** TEM images showing bundle sheath chloroplast ultrastructure. Scale bars = 2 µm, 5
841 µm, 10 µm and 2 µm, respectively.

842

843 **Figure S3. Comparison of GLK sequences**. **A)** Phylogenetic reconstruction of the *GLK* gene family.
844 Protein sequences of the extensively characterised GLK members ZmGLK1, ZmG2, OsGLK1 and
845 OsGLK2 were used to create a HMMER profile that was used as a query to search for homologs in
846 whole protein databases of 11 species (*Zea mays, Setaria viridis, S. italica, Panicum virgatum, Sorghum*
847 *bicolor, Oryza sativa, Brachypodium distachyon, Vitis vinifera, Arabidopsis thaliana, Marchantia*
848 *polymorpha* and *Physcomitrium patens*). The presence of the MYB domain in those sequences was
849 confirmed using InterproScan, and only the sequences containing the conserved AREAEAA domain
850 characteristic of the GLK subfamily were selected. The resulting 25 amino acid sequences were aligned
851 using Clustal Omega. Phylogenetic reconstruction was performed using Bayesian inference on Mr.
852 Bayes using Poisson with fixed substitution rates as a model. Two independent MCMC runs with
853 1,000,000 generations were performed and no further improvement of log likelihood of prior

854 probabilities was observed. The tree was visualized on FigTree. Clade support on each node was
855 assessed using Bayesian posterior probabilities. Members of the GLK1 subgroup are depicted in green
856 and GLK2 in blue. **B)** Amino acid sequence alignment showing conservation between ZmGLK1,
857 SvGLK1, ZmG2 and SvGLK2. The diagonal linesl show the HLH associated with MYB or MYB-related
858 proteins; the AREAEAA conserved site is depicted in horizontal lines and the last exon that contains
859 the GCT-box domain is depicted in vertical lines.
860
861 **Figure S4. DNA binding motifs in regulatory regions of *GLK* genes**. **A)** Schematics of maize, setaria
862 and rice *GLK* genes showing binding sites revealed from ATAC assays with leaf RNA. Data was
863 retrieved from https://epigenome.genetics.uga.edu/PlantEpigenome/index.html. The yellow shading
864 highlights the 800 bp upstream of the transcription start site plus the 5'UTR regions. **B)** Table showing
865 the number of DNA binding motifs found in regulatory regions of each gene. Plant non-redundant motifs
866 (JASPAR) was used to screen the above-mentioned regions of each gene, using the FIMO tool with a
867 p value of <0.0001. The photosynthesis-related/light responsive motifs (BPC, ABI3-like, GT-1, GT-4,
868 CNA, MYB-B, STZ, bZIP/bHLH(G-box), GATA, MYB-like/I-box, TCP, TGA and WRKY) were as
869 described by Sing et al. (2023) and rice and maize GLK-specific binding sites were as described in Tu
870 et al. (2022). **C, D)** Schematics of 800 bp upstream of the transcription start site plus the 5'UTR regions
871 for each gene, showing the position of all predicted DNA binding motifs (C) and of known light-
872 responsive motifs (D). The 5' UTR is shown in white and promoter sequence in grey.
873
874 **Figure S5. Phenotypic characterization of *Svglk2* mutant lines complemented with *ZmGLK1* or**
875 ***ZmG2***. **A)** Schematic of the gene edited *Svglk2* mutant allele. **B-C)** Schematic of the constructs used
876 to express *ZmGLK1* (B) or *ZmG2* (C) in the *Svglk2* background. *HygR* depicts the hygromycin
877 phosphotransferase gene and $OsACT_{pro}$ and $ZmUBI_{pro}$ represent the constitutive rice actin and maize
878 ubiquitin promoters, respectively. LB and RB refer to left and right borders. **D-M)** Transmission electron
879 microscopy images showing mesophyll (MC) and bundle sheath (BSC) cell chloroplast ultrastructure
880 in the mutant line (D, E) and in lines overexpressing *ZmGLK1* (F-I) or *ZmG2* (J-M) in the mutant
881 background. Scale bar sizes are indicated on each image. D & E are the same images as Figure 3G &
882 3H). **N-R)** Confocal images of leaf cross sections from the mutant (N) and from lines overexpressing
883 *ZmGLK1* (O, P) or *ZmG2* (Q, R) showing morphology of chloroplasts. Bundle sheath cells are outlined
884 in white. Scale bars = 50 µm. Yellow arrows indicate ectopic chloroplast formation in vascular cells
885
886 **Figure S6. Sequence variability in regulatory regions of *GLK* genes.** The SNPs from 3025 variants
887 of rice (SNP seek, IRRI) and 1210 variants of maize (Maize SNPDB) were used to calculate the
888 frequency of each nucleotide for each SNP position. The nucleotide with higher prevalence in each
889 position is plotted on the graphs for *OsGLK1*, *OsGLK2, ZmGLK1* and *ZmG2*. The 2Kb upstream of the
890 transcription start site is shown in grey, the 5'UTR in white and the first exon in black.

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

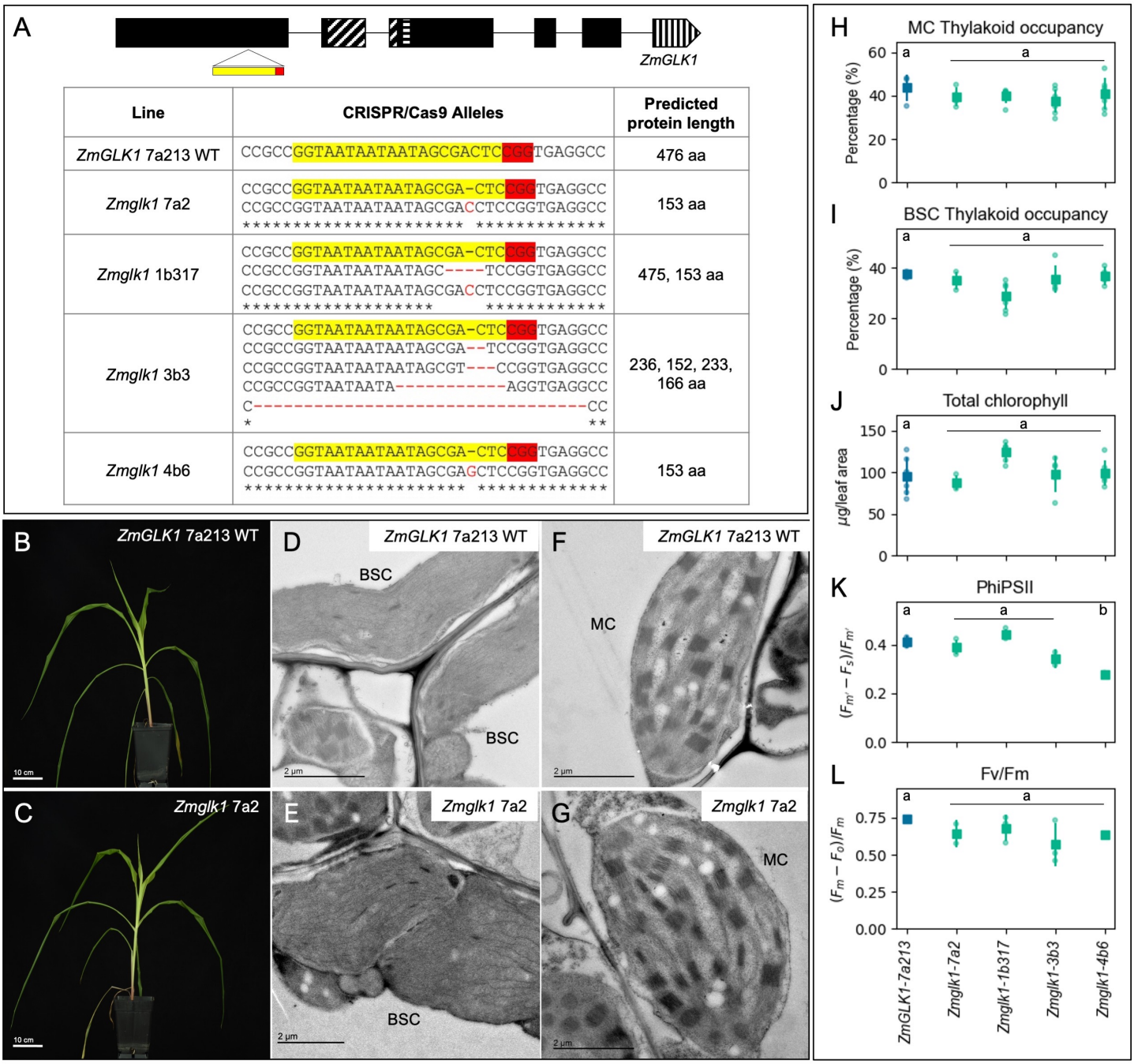

Figure 1. Chloroplast development is not perturbed in *Zmglk1* mutants of maize. **A)** Gene model of *ZmGLK1* showing predicted exons in black boxes and the target site of CRISPR/Cas9 mediated mutagenesis in yellow and red below. The HLH associated with MYB or MYB-related proteins is depicted by diagonal lines; the 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-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 µm. **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 5 biological replicates were evaluated for thylakoid occupancy and chlorophyll quantification, and 3 biological replicates were assessed for PSII fluorescence. Different letters indicate statistically different groups (p value ≤0.05, one-way ANOVA and Tukey's HSD; see Table S2 for raw data).

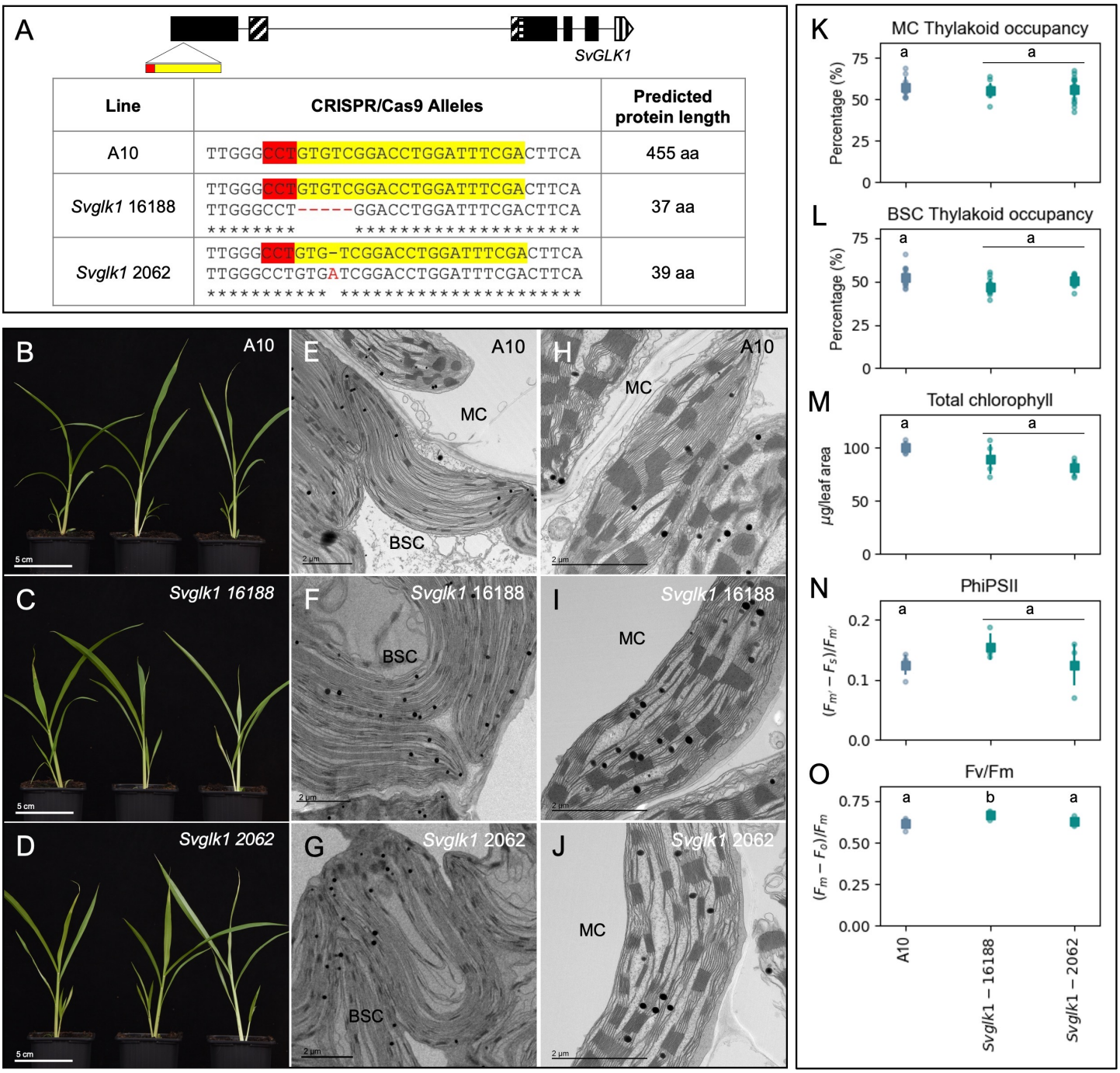

Figure 2. Chloroplast development is not perturbed in *Svglk1* mutants of setaria. A) Gene model of *SvGLK1* showing predicted exons in black boxes and the target site of CRISPR/Cas9 mediated mutagenesis in yellow and red below. The HLH associated with MYB or MYB-related proteins is depicted by diagonal 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 µm. **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 5 biological replicates were assessed for chlorophyll quantification and PSII fluorescence for each sample. Different letters indicate statistically different groups (p value ≤0.05, one-way ANOVA and Tukey's HSD; see Table S2 for raw data).

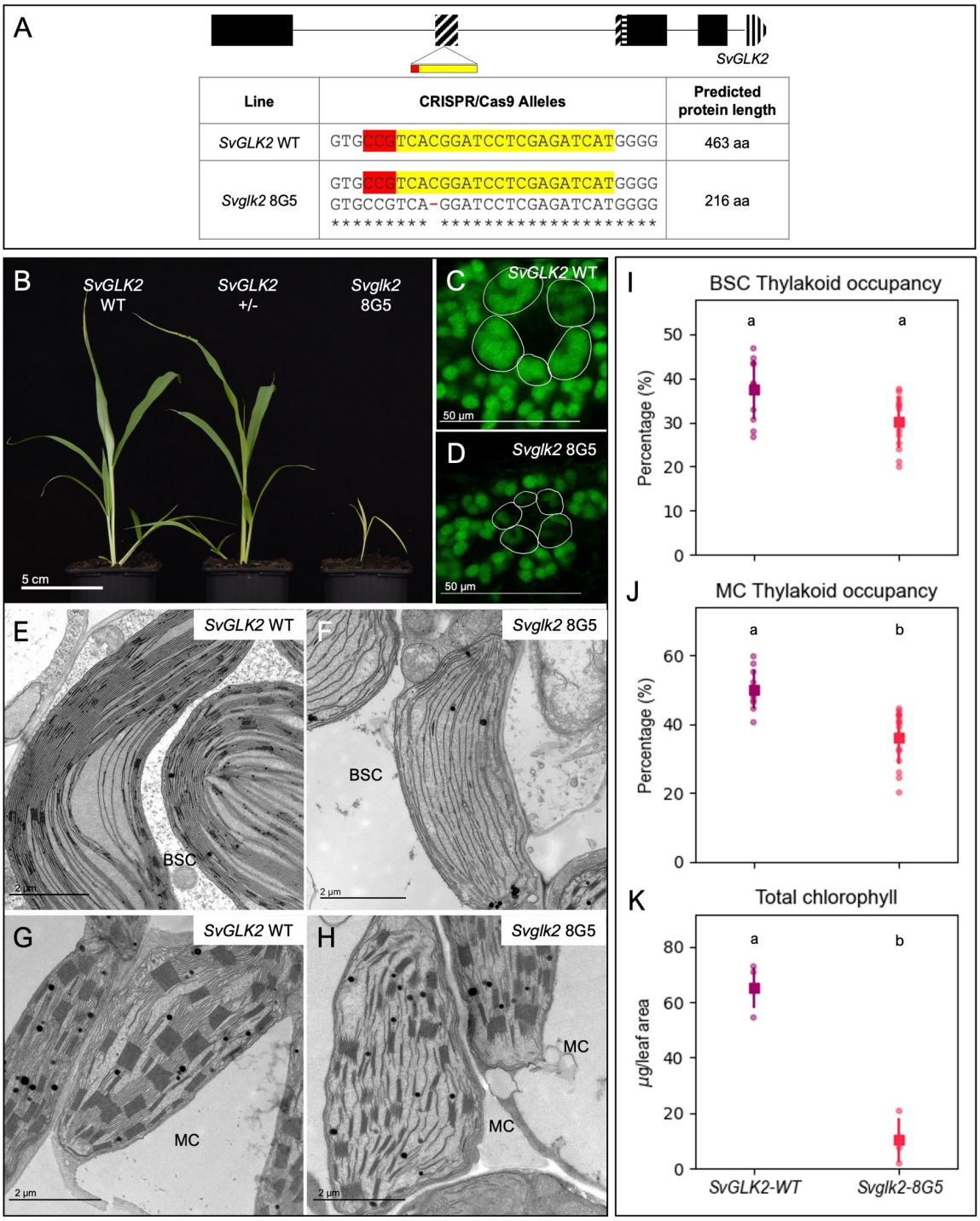

Figure 3. Chloroplast development is perturbed in both mesophyll and bundle sheath cells of *Svglk2* mutants of setaria. A) Gene model of *SvGLK2* showing predicted exons in black boxes and the target site of CRISPR/Cas9 mediated mutagenesis in yellow and red below. The HLH associated with MYB or MYB-related 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 µm. **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 µm. **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 5 biological replicates were assessed for chlorophyll quantification and PSII fluorescence for each sample. Different letters indicate statistically different groups (p value ≤0.05, one-way ANOVA and Tukey's HSD; see Table S2 for raw data).

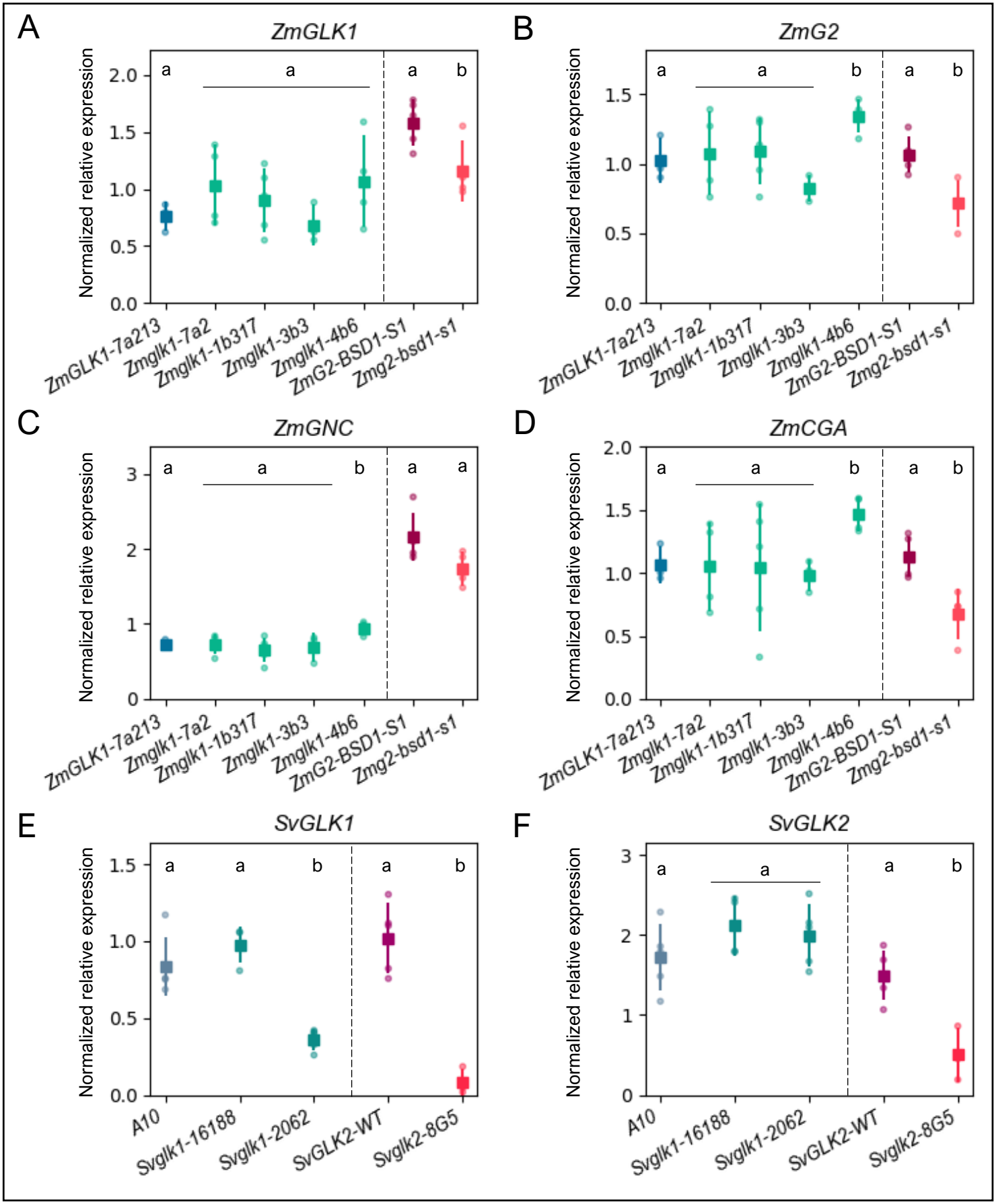

Figure 4. *GLK1* and *GLK2* transcript levels are reduced in *Svglk2* and *Zmg2* mutants of setaria and maize. A-H) Quantitative analysis of gene expression via qPCR in segregating wild-type (WT) siblings and *glk* mutants of maize (A-D) and setaria (E-F). Relative normalized expression of *ZmGLK1* (A), *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 3 and 5 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 ≤0.05; see Table S2 for raw data).

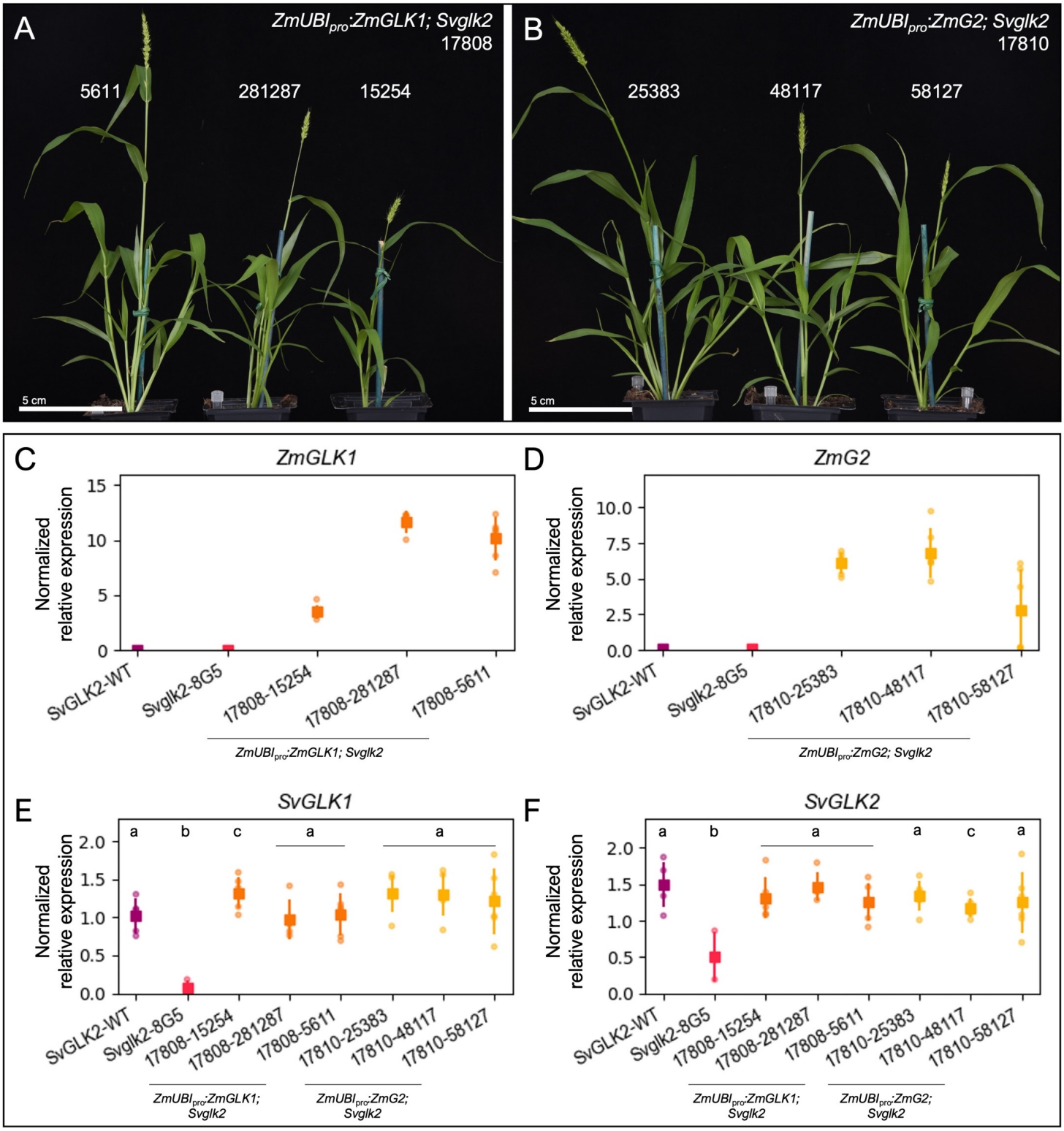

Figure 5. *ZmGLK1* and *ZmG2* complement the *Svglk2* phenotype and activate transcription of the setaria homologs. **A-B)** Whole plant phenotype of *Svglk2* mutant overexpressing *ZmGLK1* (A) or *ZmG2* (B) 30 days after sowing. Scale bars = 5 cm. **C-F)** Quantitative analysis of gene expression via 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 5 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 ≤0.05; see Table S2 for raw data).

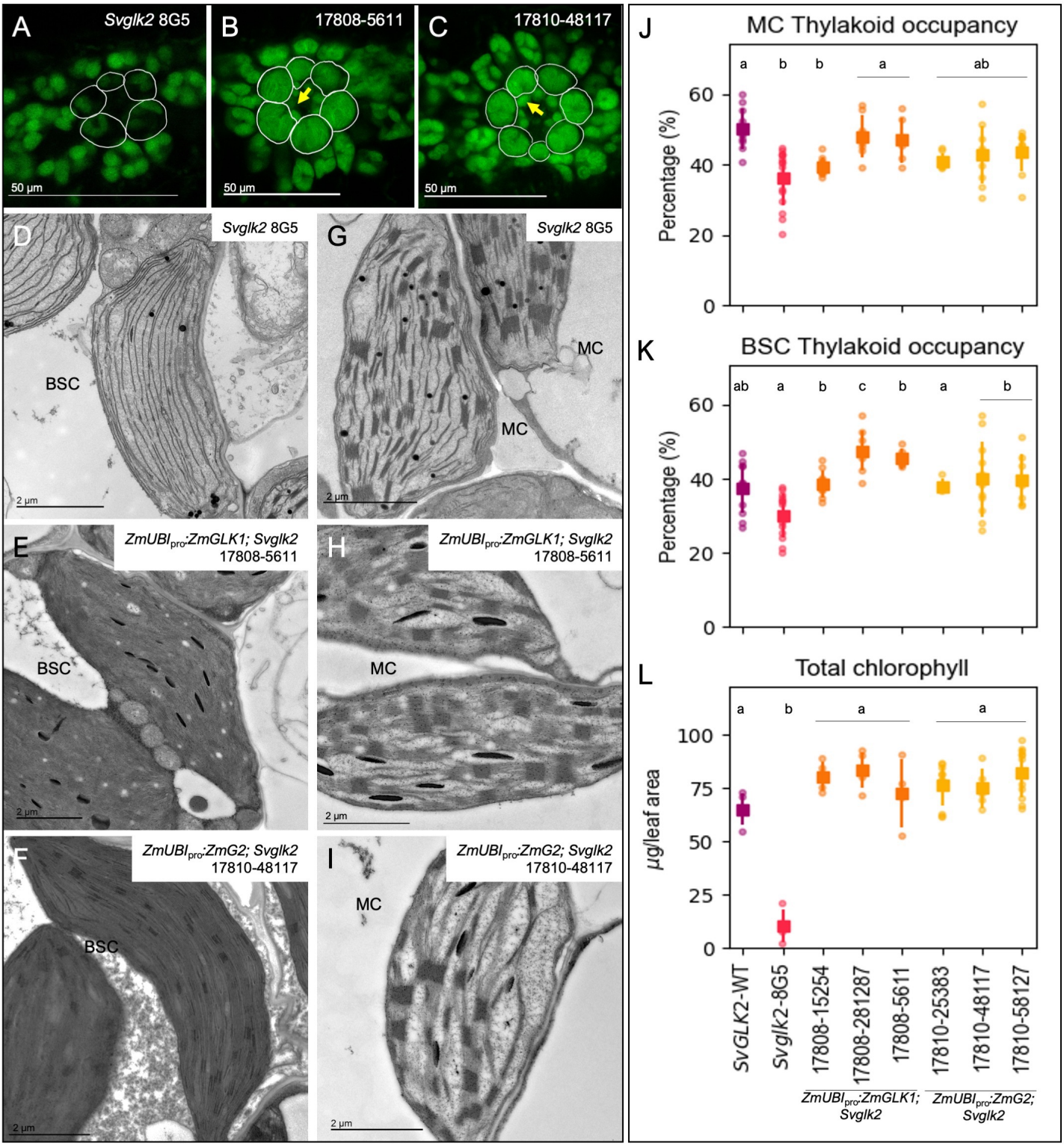

Figure 6. *ZmGLK1* and *ZmG2* are sufficient to induce dimorphic chloroplast development in both mesophyll and bundle sheath cells of setaria. **A-C)** Confocal images of leaf cross sections showing overall chloroplast morphology in the *Svglk2* mutant (A, same as Figure 3D), and representative lines 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 µm. **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 µm. **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 5-10 biological replicates were used for chlorophyll quantification. Different letters indicate statistically different groups in (p value ≤0.05, one-way ANOVA and Tukey's HSD; see Table S2 for raw data).